Assessment Decision Guide

Table of Contents

Introduction

Welcome to the Assessment Decision Guide. We developed this guide as a resource for agencies designing assessment strategies to fill critical vacancies at all levels. The guide includes information on personnel assessment, assessment methods, and assessment strategy design.

Sources of Additional Guidance

The information provided in this guide is one source agencies may consult in developing effective assessment strategies. Agencies are also encouraged to consult the resources described below for additional information and guidance on assessment.

- Testing and Assessment: An Employer's Guide to Good Practices. The U.S. Department of Labor (DOL) developed this document to assist managers and human resource professionals in making assessment-related decisions. This resource is available at http://www.onetcenter.org/guides.html or in hard copy from the Government Printing Office (GPO Order Desk 202/512-1800, Stock Number 029-014-00259-3).

- Delegated Examining Operations Handbook: A Guide for Federal Agency Examining Offices. OPM's Delegated Examining Operations Handbook is a comprehensive document by which OPM delegates competitive examining authority to agencies under the provisions of 5 U.S.C. 1104, as amended by Public Law 104-52 of 1995. Chapter 2, in particular, provides further guidance on assessment methods. This Handbook is located at http://www.opm.gov/deu.

- Uniform Guidelines on Employee Selection Procedures (29 CFR Part 1607). The Guidelines provide a set of principles for determining proper test use and selection procedures, covering topics such as test fairness and adverse impact. You can find the Guidelines (and other Government documents related to the Guidelines) at http://www.uniformguidelines.com/.

- Standards for Educational and Psychological Testing. The Standards provide information on test development, use, and evaluation in employment and other assessment settings. More information can be found on the American Psychological Association's website at http://www.apa.org/science/standards.html.

- Principles for the Validation and Use of Personnel Selection Procedures. The Principles discuss validation research and personnel selection, specifying principles of good practice in the choice, development, evaluation, and use of personnel selection procedures. This resource is available at http://www.siop.org/_Principles/principlesdefault.aspx.

Getting Expert Help

The issues to consider when selecting or developing an assessment strategy or specific assessment tool are complex. Measurement specialists can assist in selecting or developing valid, fair, and effective assessment strategies to meet specific hiring needs. Many vendors offer professionally-developed assessments. Each agency is responsible for ensuring assessments meet applicable regulatory and statutory requirements. Agencies should ensure they develop and maintain the necessary documentation to support their selection processes.

Section I: Introduction to Assessment

What is personnel assessment?
Personnel assessment refers to any method of collecting information on individuals for the purpose of making a selection decision. Selection decisions include, but are not limited to, hiring, placement, promotion, referral, retention, and entry into programs leading to advancement (e.g., apprenticeship, training, career development). Selecting qualified applicants is a critical step in building a talented and committed workforce, supporting an effective organizational culture, and enhancing the overall performance of the agency.

While many applicants may apply for any particular position, quantity does not guarantee quality. Assessment procedures can be a cost-effective tool in narrowing down large applicant pools. Assessment tools can also make the selection decision process more efficient because less time and fewer resources are expended dealing with applicants whose qualifications do not match what is needed by the agency.

Effective personnel assessment involves a systematic approach towards gathering information about applicants' job qualifications. Factors contributing to successful job performance (e.g., oral communication, problem solving) are identified using a process called job analysis. Job analysis identifies the duties performed on the job and the competencies needed for effective job performance. Basing personnel assessment closely on job analysis results makes the connection between job requirements and personnel assessment tools more transparent, thereby improving the perceived fairness of the assessment process.

What are personnel assessment tools?
Generally speaking, an assessment tool is any test or procedure administered to individuals to evaluate their *job-related* competencies, interests, or fitness for employment. The accuracy with which applicant assessment scores can be used to forecast performance on the job is the tool's most important characteristic, referred to as predictive validity (Schmidt & Hunter, 1998).[1]

Not all assessment tools are appropriate for every job and organizational setting. Agencies must consider a number of factors in determining the most appropriate assessment strategy for a particular situation. These considerations include timetables for filling positions, available staff and financial resources, number of positions to be filled, and the nature and complexity of the work performed in the positions to be filled.

Why is effective personnel assessment important?
It is very simple — using effective assessment tools will reduce the degree of error in making hiring decisions. Well-developed assessment tools allow agencies to specifically target the competencies and skills they seek. This helps to ensure the time spent by both applicants and agency personnel adds value to the decision-making process. Selection errors have financial and

---

[1] 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.

practical impacts on organizations. The consequences of even a single selection error can create problems for an entire work unit. For example, managers may have to devote substantial time training and counseling the marginal employee and coworkers must often handle increased workloads as they correct or perform the employee's work. Some selection errors can have agency-wide consequences such as customer service complaints, increases in work-related accidents and injuries, high absenteeism, poor work quality, increased turnover, or damage to the reputation of the agency.

Good assessment will also benefit employees who experience greater organizational commitment and job satisfaction because they are matched to jobs for which they are well suited. In addition, using job-related assessment tools often results in more favorable applicant reactions to the selection process. Such perceptions have lasting consequences for the agency including: promoting a positive image of the organization, increasing the likelihood of the applicant accepting a job offer, increasing the number of job referrals, and reducing the risk of selection system challenges and complaints.

What is a competency and what is competency-based assessment?
OPM defines a competency as "*A measurable pattern of knowledge, skills, abilities, behaviors, and other characteristics that an individual needs to perform work roles or occupational functions successfully.*" Competencies specify the "how" of performing job tasks, or what the person needs to do the job successfully (Shippmann et al., 2000).[2] Competencies represent a whole-person approach to assessing individuals.

Competencies tend to be either general or technical. General competencies reflect the cognitive and social capabilities (e.g., problem solving, interpersonal skills) required for job performance in a variety of occupations. On the other hand, technical competencies are more specific as they are tailored to the particular knowledge and skill requirements necessary for a specific job. OPM has conducted a number of occupational studies to identify competencies for many Federal occupations. These competencies are available in the Delegated Examining Operations Handbook.

How do I determine which competencies are needed for the position?
A job analysis identifies the job tasks, roles, and responsibilities of the incumbent performing the job, as well as the competencies required for performance, the resources used during performance, and the context (or environment) in which performance occurs. As such, a job analysis demonstrates the clear connection between job tasks and the competencies necessary to perform those tasks.

Conducting a job analysis involves collecting information from job experts. The term subject matter expert (SME) is properly applied to anyone who has direct, up-to-date experience of a job and is familiar with all of its tasks. The person might currently hold the job or supervise the job. SMEs must provide accurate information and effectively communicate their ideas. SMEs should rate the job tasks and competencies for importance to successful job performance. Critical

---

[2] Shippman, J. S., Ash, R. A., Carr, L., Hesketh, B., Pearlman, K., Battista, M., Eyde, L. D., Kehoe, J., Prien, E. P., & Sanchez, J. I. (2000). The practice of competency modeling. *Personnel Psychology, 53,* 703-740.

incidents (i.e., examples of particularly effective or ineffective work behaviors) are also developed in some cases to describe essential job functions. Documentation of the job analysis process and the linkages between job tasks, competencies, and selection tool content are essential to ensure an assessment strategy meets legal and professional guidelines. Please refer to the section on conducting a job analysis in OPM's Delegated Examining Operations Handbook for more information.

Section II: Evaluating and Implementing Assessment Tools

In selecting and evaluating an assessment tool, one must consider a number of important factors such as: (1) reliability, (2) validity, (3) technology, (4) the legal context, and (5) face validity/applicant reactions. Each of these issues is discussed below as well as considerations concerning the design and effectiveness of selection systems.

Reliability. The term reliability refers to consistency. Assessment reliability is demonstrated by the consistency of scores obtained when the same applicants are reexamined with the same or equivalent form of an assessment (e.g., a test of keyboarding skills). No assessment procedure is perfectly consistent. If an applicant's keyboarding skills are measured on two separate occasions, the two scores (e.g., net words per minute) are likely to differ.

Reliability reflects the extent to which these individual score differences are due to "true" differences in the competency being assessed and the extent to which they are due to chance, or random, errors. Common sources of such error include variations in:

- Applicant's mental or physical state (e.g., the applicant's level of motivation, alertness, or anxiety at the time of testing)

- Assessment administration (e.g., instructions to applicants, time limits, use of calculators or other resources)

- Measurement conditions (e.g., lighting, temperature, noise level, visual distractions)

- Scoring procedures (e.g., raters who evaluate applicant performance in interviews, assessment center exercises, writing tests)

A goal of good assessment is to minimize random sources of error. As a general rule, the smaller the amount of error, the higher the reliability.

Reliability is expressed as a positive decimal number ranging from 0 to 1.00, where 0 means the scores consist entirely of error. A reliability of 1.00 would mean the scores are free of any random error. In practice, scores always contain some amount of error and their reliabilities are less than 1.00. For most assessment applications, reliabilities above .70 are likely to be regarded as acceptable.

The practical importance of consistency in assessment scores is they are used to make important decisions about people. As an example, assume two agencies use similar versions of a writing skills test to hire entry-level technical writers. Imagine the consequences if the test scores were so inconsistent (unreliable) applicants who applied at both agencies received low scores on one test but much higher scores on the other. The decision to hire an applicant might depend more on the reliability of the assessments than his or her actual writing skills.

Reliability is also important when deciding which assessment to use for a given purpose. The test manual or other documentation supporting the use of an assessment should report details of reliability and how it was computed. The potential user should review the reliability information available for each prospective assessment before deciding which to implement. Reliability is also a key factor in evaluating the validity of an assessment. An assessment that fails to produce consistent scores for the same individuals examined under near-identical conditions cannot be expected to make useful predictions of other measures (e.g., job performance). Reliability is critically important because it places a limit on validity.

Validity. Validity refers to the relationship between performance on an assessment and performance on the job. Validity is the most important issue to consider when deciding whether to use a particular assessment tool because an assessment that does not provide useful information about how an individual will perform on the job is of no value to the organization.

There are different types of validity evidence. Which type is most appropriate will depend on how the assessment method is used in making an employment decision. For example, if a work sample test is designed to mimic the actual tasks performed on the job, then a content validity approach may be needed to establish the content of the test matches in a convincing way the content of the job, as identified by a job analysis. If a personality test is intended to forecast the job success of applicants for a customer service position, then evidence of predictive validity may be needed to show scores on the personality test are related to subsequent performance on the job.

The most commonly used measure of predictive validity is a correlation (or validity) coefficient. Correlation coefficients range in absolute value from 0 to 1.00. A correlation of 1.00 (or -1.00) indicates two measures (e.g., test scores and job performance ratings) are perfectly related. In such a case, you could perfectly predict the actual job performance of each applicant based on a single assessment score. A correlation of 0 indicates two measures are unrelated. In practice, validity coefficients for a single assessment rarely exceed .50. A validity coefficient of .30 or higher is generally considered useful for most circumstances (Biddle, 2005).[3]

When multiple selection tools are used, you can consider the *combined* validity of the tools. To the extent the assessment tools measure different job-related factors (e.g., reasoning ability and honesty) each tool will provide unique information about the applicant's ability to perform the job. Used together, the tools can more accurately predict the applicant's job performance than either tool used alone. The amount of predictive validity one tool adds relative to another is often referred to as the incremental validity of the tool. The incremental validity of an assessment is important to know because even if an assessment has low validity by itself, it has the potential to add significantly to the prediction of job performance when joined with another measure.

Just as assessment tools differ with respect to reliability, they also differ with respect to validity. The following table provides the estimated validities of various assessment methods for predicting job performance (represented by the validity coefficient), as well as the incremental

---

[3] Biddle, D. (2005). *Adverse Impact and Test Validation: A Practitioner's Guide to Valid and Defensible Employment Testing.* Burlington, VT: Gower Publishing.

validity gained from combining each with a test of general cognitive ability. Cognitive ability tests are used as the baseline because they are among the least expensive measures to administer and the most valid for the greatest variety of jobs. The second column is the correlation of the combined tools with job performance, or how well they collectively relate to job performance. The last column shows the percent increase in validity from combining the tool with a measure of general cognitive ability. For example, cognitive ability tests have an estimated validity of .51 and work sample tests have an estimated validity of .54. When combined, the two methods have an estimated validity of .63, an increase of 24% above and beyond what a cognitive ability test used alone could provide.

Table 1: Validity of Various Assessment Tools Alone and in Combination

Assessment method	Validity of method used alone	Incremental (combined) validity	% increase in validity from combining tool with cognitive ability
Tests of general cognitive ability	.51		
Work sample tests	.54	.63	24%
Structured interviews	.51	.63	24%
Job knowledge tests	.48	.58	14%
Accomplishment record*	.45	.58	14%
Integrity/honesty tests	.41	.65	27%
Unstructured interviews	.38	.55	8%
Assessment centers	.37	.53	4%
Biodata measures	.35	.52	2%
Conscientiousness tests	.31	.60	18%
Reference checking	.26	.57	12%
Years of job experience	.18	.54	6%
Training & experience point method	.11	.52	2%
Years of education	.10	.52	2%
Interests	.10	.52	2%

Note: Table adapted from Schmidt & Hunter (1998). Copyright © 1998 by the American Psychological Association. Adapted with permission.[4]
*Referred to as the training & experience behavioral consistency method in Schmidt & Hunter (1998).

Technology. The technology available is another factor in determining the appropriate assessment tool. Agencies that receive a large volume of applicants for position announcements may benefit from using technology to narrow down the applicant pool, such as online screening of resumes or online biographical data (biodata) tests. Technology can also overcome distance challenges and enable agencies to reach and interview a larger population of applicants.

However, because technology removes the human element from the assessment process, it may be perceived as "cold" by applicants, and is probably best used in situations that do not rely

---

[4] 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.

heavily on human intervention, such as collecting applications or conducting applicant screening. Technology should not be used for final selection decisions, as these traditionally require a more individualized and in-depth evaluation of the candidate (Chapman and Webster, 2003).[5]

Legal Context of Assessment. Any assessment procedure used to make an employment decision (e.g., selection, promotion, pay increase) can be open to claims of adverse impact based on subgroup differences. Adverse impact is a legal concept used to determine whether there is a "substantially different" passing rate (or selection rate) between two groups on an assessment procedure (see www.uniformguidelines.com for a more detailed discussion). Groups are typically defined on the basis of race (e.g., Blacks compared to Whites), gender (i.e., males compared to females), or ethnicity (e.g., Hispanics compared to Non-Hispanics). Assessment procedures having an adverse impact on any group must be shown to be job-related (i.e., valid).

What is a "substantially different" passing rate? The Uniform Guidelines provide a variety of statistical approaches for evaluating adverse impact. The most widely used method is referred to as the 80% (or four-fifths) rule-of-thumb. The following is an example where the passing rate for females is 40% and the passing rate for males is 50%. The Uniform Guidelines lay out the following steps for computing adverse impact:

- Divide the group with the lowest rate (females at 40%) by the group with the highest rate (males at 50%)
- In this case, divide 40% by 50% (which equals 80%)
- Note whether the result is 80% or higher

According to the 80% rule, adverse impact is not indicated as long as the ratio is 80% or higher. In this case, the ratio of the two passing rates is 80%, so evidence of adverse impact is not found and the passing rate of females is not considered substantially different from males.

Agencies are encouraged to consider assessment strategies to minimize adverse impact. When adverse impact is discovered, the assessment procedure must be shown to be job-related and valid for its intended purpose.

Face Validity/Applicant Reactions. When applicants participate in an assessment process, they are not the only ones being evaluated; the agency is being evaluated as well. Applicants who complete an assessment process leave with impressions about the face validity and overall fairness of the assessment procedure. Their impressions can also be impacted by whether they believe they had a sufficient opportunity to display their job-related competencies. The quality of the interactions between the applicant and agency representatives can also affect applicant reactions. Agencies using grueling assessment procedures may end up alienating applicants. It is important to recognize applicants use the assessment process as one means to gather information about the agency. Failure to act on this fact can be very costly to agencies, particularly if top candidates are driven to look elsewhere for employment opportunities.

---

[5] Chapman, D. S., & Webster, J. (2003). The use of technologies in the recruiting, screening, and selection processes for job candidates. *International Journal of Selection and Assessment, 11,* 113-120.

Designing a Selection Process

The design of an assessment strategy should begin with a review of the critical competencies identified from the job analysis results. Once you decide *what* to assess, you must then determine *how* to structure the personnel assessment process. In designing a selection process, a number of practical questions must be addressed, such as:

- How much money is available?
- What assessment tool(s) will be selected?
- If using multiple tools, in what order should they be introduced?
- Are trained raters needed, and if so, how many (e.g., for conducting interviews)?
- How many individuals are expected to apply?
- What is the timeframe for filling vacancies?

For example, if your budget is tight, you will need to rule out some of the more expensive methods such as assessment centers or work simulation tests. If you are expecting to receive thousands of applications (based on projections from similar postings), you will need to develop an effective screening mechanism ahead of time. If you need to fill a vacancy and only have a few weeks to do so, then a multi-stage process will probably not be feasible. In working out answers to these questions, it is usually helpful to think in terms of the entire selection process, from beginning to end.

One key consideration is the number of assessment tools to include in the process. Using a variety of assessments tends to improve the validity of the process and will provide information on different aspects of an applicant's likely job performance. Using a single measure will tend to identify applicants who have strengths in a specific area but may overlook applicants who have high potential in other areas. Assessing applicants using multiple methods will reduce errors because people may respond differently to different methods of assessment. For example, some applicants who excel at written tests may be too nervous to do well in interviews, while others who suffer from test anxiety may give impressive interviews. Another advantage of using a variety of assessment methods is a multiple hurdle approach can be taken. The least expensive assessments can be used first to pare down the applicant pool. More labor-intensive and time-consuming procedures can be introduced at a later stage when there are fewer candidates to evaluate.

Considering which assessment methods best measure which competencies at which stage in the process should help you develop a process well suited to your agency's hiring needs.

Ensuring an Effective Assessment Process

Agencies are encouraged to standardize and document the assessment process through the following steps:

- Treat all individuals consistently. This is most easily accomplished by adopting a standardized assessment and decision-making process. "Standardizing" means making a

process uniform to ensure the same information is collected on each individual and is used in a consistent manner in employment decisions.

- Ensure the selection tool is based on an up-to-date job analysis and is supported by strong validity evidence. A validation study can verify applicants who score well on the selection device are more likely to do well on the job and contribute to organizational success. Agencies not familiar with validation research methodology are encouraged to consult a measurement expert.

- To ensure applicants perceive the process as fair, agencies are encouraged to:

    a) Offer applicants a realistic job preview before the assessment process
    b) Discuss with applicants the rationale for using the selection device, as well as what it assesses and why these competencies are important to the job
    c) Provide applicants the opportunity to ask questions about the job and the selection process
    d) Treat individuals with respect, sensitivity, and impartiality during the process
    e) Provide feedback about all hiring decisions in a timely and courteous manner
    f) Elicit feedback from applicants (those selected and those not selected) on the selection process

- Ensure all persons involved in the selection process (e.g., administrators, interviewers, assessors) understand their roles and responsibilities

(Information adapted from Gilliland, S.W., & Cherry, B., 2000).[6]

Sources of Additional Information

For a more in-depth introduction to personnel assessment practices, including measurement techniques and related considerations (e.g., reliability, validity, job analyses, and legal requirements), refer to *Essentials of Personnel Assessment and Selection* by Guion and Highhouse (2006).[7]

For a non-technical summary of the research literature on the value of commonly used assessment methods, see *Selection Methods: A Guide to Implementing Formal Assessments to Build a High Quality Workforce* (Pulakos, 2005).[8]

More information about designing and implementing a selection process can be found in *Competency-based Recruitment and Selection: A Practical Guide* by Wood and Payne (1998).[9]

---

[6] Gilliland, S. W., & Cherry, B. (2000). Managing customers of selection. In J. K. Kehoe (Ed.), *Managing Selection in Changing Organizations* (pp. 158-196). San Francisco: Jossey-Bass.

[7] Guion, R. M., & Highhouse, S. (2006). *Essentials of Personnel Assessment and Selection*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

[8] Pulakos, E. D. (2005). *Selection Methods: A Guide to Implementing Formal Assessments to Build a High Quality Workforce.* Alexandria, VA: SHRM Foundation.

[9] Wood, R., & Payne, T. (1998). *Competency-based Recruitment and Selection: A Practical Guide*. Hoboken, NJ: Wiley.

Section III: Overview of Assessment Methods

This section of the Assessment Decision Guide describes the following assessment methods:

1. Accomplishment Records
2. Assessment Centers
3. Biographical Data (Biodata) Tests
4. Cognitive Ability Tests
5. Emotional Intelligence Tests
6. Integrity/Honesty Tests
7. Job Knowledge Tests
8. Personality Tests
9. Reference Checking
10. Situational Judgment Tests
11. Structured Interviews
12. Training and Experience (T & E) Evaluations
13. Work Samples and Simulations

Each description includes information on the following considerations:

- Validity – The extent to which the assessment method has been shown to accurately measure a job-related competency and/or predict successful performance on the job. For each assessment method, validity is typically discussed in terms of content validity, criterion-related validity, or construct validity.

- Face Validity/Applicant Reactions – Face validity refers to applicants' judgments regarding the degree to which the content of an assessment tool is relevant and representative of the content of the job. Applicant reactions include perceptions of the face validity of the test as well as overall reactions to the assessment process.

- Administration Method – Information regarding how an assessment method can be used. Typically this includes the size of applicant groups that can be assessed at one time, and whether the assessment can be administered using paper-and-pencil or electronic formats.

- Subgroup Differences – The extent to which the assessment method has been shown to result in different pass (or selection) rates, average scores, or prediction errors across groups, typically based on race, ethnicity, or gender.

- Development Costs – The amount and type of resources required to develop an assessment, in terms of time, money, and technical expertise.

- Administration Costs – The amount and type of resources required to administer an assessment, in terms of time, money, staff, equipment, facilities, and information technology support.

- Utility/Return on Investment (ROI) – The extent to which the benefits gained from using the assessment method outweigh the costs of development and administration.

- Common Uses – A description of the occupations and/or situations for which the assessment method is typically used.

The following table provides a summary of how each of the assessment methods discussed in this section rank on five of the assessment considerations covered in this section. Readers are encouraged to consult the specific section for a particular assessment method to fully interpret the rankings presented in this table.

Table 2: Summary of Assessment Methods on Five Considerations

Assessment Method	Assessment Consideration				
	Validity	Face Validity	Subgroup Differences	Development Costs	Administration Costs
Accomplishment Records	High	Moderate	Low	Moderate	Moderate
Assessment Centers	Moderate	High	Low	High	High
Biographical Data (Biodata) Tests	Moderate	Moderate	Moderate	High	Low
Cognitive Ability Tests	High	Moderate	High	Moderate	Low
Emotional Intelligence Tests	Moderate	Moderate	Low	High	Low
Integrity/Honesty Tests	Moderate	Moderate	Low	High	Low
Job Knowledge Tests	High	High	High	High	Low
Personality Tests	Moderate	Moderate	Low	High	Low
Reference Checking	Moderate	Moderate	Low	Moderate	Low
Situational Judgment Tests	Moderate	High	Moderate	High	Low
Structured Interviews	High	High	Low	Moderate	Moderate
Training and Experience (T & E) Evaluations	Low	Moderate	Low	Low	Low
Work Samples and Simulations	High	High	Low	High	High

Most of the assessment methods in this section require considerable test development and measurement expertise to develop in-house. Measurement specialists can assist in selecting or developing valid, fair, and effective assessment tools to meet specific hiring needs. Many vendors offer professionally-developed assessments. Each agency is responsible for ensuring assessments meet all regulatory and legal criteria. Agencies should develop and maintain the necessary documentation to support their selection processes.

Accomplishment Records: The accomplishment record is a systematic procedure used to collect information about applicants' training, education, experience, and past achievements related to critical job competencies. The accomplishment record is based on the [behavioral consistency](#) principle that past behavior is the best predictor of future behavior. Applicants are typically asked to submit information on personal accomplishments to best illustrate their proficiency on critical job competencies (generally between four and eight).

Specifically, applicants are often required to provide written descriptions of what was accomplished, including detailed information about the problem or situation, the specific actions taken, and the results or outcomes achieved by those actions. The name and contact information of an individual who can verify the statements is also usually required. Accomplishments do not need to be limited to those demonstrating previous experience on the specific job in question. Rather, experience gained from other jobs or through community service, school, volunteer work, military service, or even hobbies may also be used to provide examples of accomplishments relevant to the targeted position.

Accomplishment statements are evaluated by a panel of trained raters using competency-based benchmarks created for the targeted occupation. The competency-based benchmarks typically provide specific behavioral examples of what constitutes high, medium, and low levels of proficiency. Scoring is typically based on the degree to which the behaviors and outcomes described within the accomplishments reflect the benchmark levels of proficiency. The length of the rating process, generally between two and six weeks, is determined by the number of applicants and the number of competencies being assessed. Because the accomplishment descriptions are in the form of a written narrative, the method assumes applicants are able to communicate in writing.

Variations of the traditional accomplishment record method involve the collection of alternative types of applicant proficiency or experience information. For example, applicants may be asked to complete a self-report measure by checking off job-related tasks they have performed, rating their degree of proficiency in performing job-related tasks, or rating the extent to which they possess a critical job competency. This approach is also considered a variation on the [training and experience evaluation](#) method, discussed later in this section. Often, accomplishments are later collected to support the self-reported information. In cases where an accomplishment record cannot be implemented, self-report questionnaires are sometimes used as an alternative pre-screen tool. It is important to note the [validity](#) and reliability evidence for some of these self-report measures have not been substantiated by research, and may not be comparable to levels associated with traditional accomplishment records.

Another variation of the accomplishment record is a process requiring formal verification of the statements (e.g., via references) made by applicants in their written accomplishments (and self-report information, if applicable). This technique is intended to discourage applicants from inflating or otherwise distorting their submitted accomplishment descriptions.

14

Considerations:

- Validity – If developed properly, the critical dimensions of job performance to which applicants respond will be representative of those required for the job (i.e., they have a high degree of content validity) and scores on the assessment will relate strongly to measures of overall job performance (i.e., they have a high degree of criterion-related validity)

- Face Validity/Applicant Reactions – Reactions from professionals who feel they should be evaluated on their experience is typically favorable; Less favorable reactions may be observed for entry-level applicants having relatively brief employment histories; When applied to entry-level positions, it is important to give credit for accomplishments gained through other than paid employment (e.g., school, volunteer work, community service); Some prospective applicants who dislike writing detailed narratives may be discouraged from applying

- Administration Method – Can be administered individually via paper and pencil or electronically to a large group of applicants at one time

- Subgroup Differences – Generally little or no performance differences are found between men and women or applicants of different races, although the presence of subgroup differences may depend on the specific competencies being assessed

- Development Costs – Accomplishment records can be developed for any occupation within two to four weeks, depending on the number of dimensions measured; Time and cost requirements are associated mainly with benchmark development, scoring procedures, and rater training

- Administration Costs – Highly time consuming for applicants to complete and the scoring may be more time consuming compared to other assessment methods with clear right or wrong answers (e.g., job knowledge tests); The length of the rating process depends on the number of applicants and competencies measured

- Utility/ROI – High return on investment for managerial, professional, or other jobs where applicants may prefer to be evaluated on the basis of their actual work experience rather than an impersonal, standardized test; Investment of time and effort to develop and administer may not be worthwhile in situations where applicant reactions to traditional tests are not a concern

- Common Uses – Commonly used when negative applicant reactions to traditional tests or test "look-alikes" such as biodata are expected; Also commonly used as a screening device prior to an interview

References: (See Section VI for a summary of each article)

Hough, L. M. (1984). Development and evaluation of the "accomplishment record" method of selecting and promoting professionals. *Journal of Applied Psychology, 69*(1), 135-146.

Hough, L. M., Keyes, M. A., & Dunnette, M. D. (1983). An evaluation of three "alternative" selection procedures. *Personnel Psychology, 36*(2), 261-276.

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*(4), 302-318.

Schmidt, F. L., Caplan, J. R., Bemis, S. E., Decuir, R., Dunn, L., & Antone, L. (1979). *The behavioral consistency method of unassembled examining*. Washington, DC: U.S. Office of Personnel Management, Personnel Resources and Development Center.

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*(2), 262-274.

Assessment Centers: The assessment center is not a place as its name seems to suggest, nor is it a single process or method. Rather, an assessment center employs multiple assessment methods and exercises to evaluate a wide range of competencies used to make a variety of employment decisions (e.g., employee selection, career development, promotion). Assessment centers can be used to assess small groups of people at relatively the same time. Many assessment center exercises resemble work sample tests designed to simulate the actual challenges found on the job.

Assessment center exercises can be used to measure many different types of job related competencies, including interpersonal skills, oral and written communication, planning and evaluating, and reasoning and problem solving abilities. A frequently used assessment center exercise is the in-basket test. A typical in-basket test is designed to simulate administrative tasks. During this exercise, applicants are asked to play the role of a person new to the job and are instructed to read and react to a pile of memos, messages, reports, and articles.

Some assessment center exercises can be used to evaluate groups and individual behaviors in group situations. For example, in a leaderless group discussion, a group of applicants is tasked with solving a problem or a series of problems in a limited amount of time. Other assessment center exercises include, but are not limited to, job knowledge tests, personality tests, and structured interviews. Applicant performance is usually observed and evaluated by multiple assessors (i.e., raters). When used for internal promotion purposes, assessment centers are frequently designed to reflect values and practices specific to an organization, but when used to assess external applicants, assessment centers should be designed to focus on the job and level of the job (e.g., manager) rather than practices unique to the organization. While assessment centers can be designed for various types of jobs, they are particularly effective for assessing higher-level managerial and leadership competencies. Assessment centers require extensive experience to develop, considerable logistical planning to set up, and numerous personnel to administer. Highly trained assessors are needed to observe and evaluate applicant performance on the group and individual exercises.

Considerations:

- Validity – Overall, assessment center scores do a good job predicting occupational success (i.e., they have a high degree of criterion-related validity), but the level of predictive validity can vary depending on the purpose of the assessment, the extent of assessor training, and the assessment methods used (See Gaugler, Rosenthal, Thornton & Bentson, 1987); Generally, there is little evidence assessment centers provide useful information about the relative strengths and weaknesses of a given individual. So while assessment centers are highly useful for making selection decisions, they are less useful for providing comprehensive developmental feedback

- Face Validity/Applicant Reactions – Applicants typically react favorably to assessment center exercises and often perceive the process as being very fair (i.e., as having a high degree of face validity); Exercises simulating actual job tasks provide effective realistic job previews

- Administration Method – Used to assess small groups of people at more or less the same time; Can assess individual performance either alone or in a team environment; Enables "hands-on" performance by the applicant and typically in a simulated work setting

- Subgroup Differences – Generally little or no performance differences are found between men and women or applicants of different races, although the presence of gender and/or racial differences may depend on the competencies being assessed

- Development Costs – Often costly to develop, both in terms of time and money; Usually requires frequent updating because the scenarios and problems used in the exercises are often remembered by the applicants long after the administration (raising potential test security issues) and because exercise content may become outdated over time (e.g., memos might be sent via e-mail rather than Fax)

- Administration Costs – Usually expensive to administer; Requires several assessors to observe and rate applicant performance and may require a spacious testing location conducive to rating many applicants at one time; Administration time often depends on number of applicants

- Utility/Return On Investment (ROI) – Productivity gains realized by selecting managers and skilled individuals average well above administrative costs

- Common Uses – Can be used for promotion or selection purposes; Used to measure many types of job related skills, but most widely used to assess candidates for leadership, managerial, customer service, and sales positions; May require a pre-screen to limit the number of applicants scheduled for the labor-intensive assessment center process

References: (See Section VI for a summary of each article)

Arthur, W. Jr., Day, E. A., McNelly, T. L., & Edens, P. S. (2003). A meta-analysis of the criterion-related validity of assessment center dimensions. *Personnel Psychology, 56*, 125-154.

Caldwell, C., Thornton, G. C., & Gruys, M. (2003). Ten classic assessment center errors: Challenges to selection validity. *Public Personnel Management, 32*(1), 73-88.

Gaugler, B. B., Rosenthal, D. B., Thornton, G. C., & Bentson, C. (1987). Meta-analysis of assessment center validity. *Journal of Applied Psychology, 72*(3), 493-511.

Testing and Assessment: An Employer's Guide to Good Practices. (2000). Washington, DC: U.S. Department of Labor, Employment and Training Administration. Note: Article can be accessed at http://www.onetcenter.org/guides.html.

Woehr, D., & Winfred, A. (2003). The construct-related validity of assessment center ratings: A review and meta-analysis of the role of methodological factors. *Journal of Management, 29*(2), 231-258.

Zedeck, S. (1986). A process analysis of the assessment center method. In B. M. Staw & L. L. Cummings (Eds.), *Research in organizational behavior, 8,* 259-296.

The following Society for Industrial and Organizational Psychology (SIOP) website contains information on Assessment Centers:
http://www.siop.org/workplace/employment%20testing/employment_testing_toc.aspx

Biographical Data (Biodata) Tests: Biodata measures are based on the measurement principle of behavioral consistency, that is, past behavior is the best predictor of future behavior. Biodata measures include items about past events and behaviors reflecting personality attributes, attitudes, experiences, interests, skills and abilities validated as predictors of overall performance for a given occupation.

Often, biodata test items are developed through behavioral examples provided by subject matter experts (SMEs). These items specify situations likely to have occurred in a person's life, and ask about the person's typical behavior in the situation. In addition, biodata items reflect external actions that may have involved, or were observable by, others and are objective in the sense there is a factual basis for responding to each item. An item might ask "How many books have you read in the last 6 months?" or "How often have you put aside tasks to complete another, more difficult assignment?" Test takers choose one of several predetermined alternatives to best match their past behavior and experiences.

A response to a single biodata item is of little value. Rather, it is the pattern of responses across several different situations that give biographical data the power to predict future behavior on the job. For this reason, biodata measures often contain between 10 and 30 items and some wide-ranging instruments may contain a hundred or more items. Response options commonly use a 5-point scale (1 = Strongly Disagree to 5 = Strongly Agree). Once a group of biodata items is pre-tested on a sample of applicants, the responses are used to group the items into categories or scales. Biodata items grouped in this way are used to assess how effectively applicants performed in the past in competency areas closely matched to those required by the job.

A more recent development is targeted biodata instruments. In contrast to traditional biodata measures developed to predict overall job performance, targeted biodata measures are developed to predict individual differences in specific job-related behaviors of interest. Similar to the developmental process used for traditional biodata, the content of a targeted biodata measure is often driven by SME-generated behavioral examples relevant to the specific behavior(s) of interest.

An example of a targeted biodata measure is a job compatibility measure (sometimes referred to as a suitability measure) which focuses on the prediction of counterproductive or deviant behaviors. Counterproductive behavior is often defined as on-the-job behavior that is (a) harmful to the mission of the organization, (b) does not stem from a lack of intelligence, and (c) is willful or so seriously careless it takes on the character of being willful. Previous criminal misconduct (e.g., theft), employment misconduct (e.g., sexual harassment, offensiveness to customers, and disclosure of confidential material), fraud, substance abuse, or efforts to overthrow the Government are some major factors that may be relevant to suitability determinations. A job compatibility index is typically used to screen out applicants who are more likely to engage in counterproductive behavior if they are hired. Job compatibility measures are less costly to implement than other procedures typically used to detect counterproductive behaviors (e.g., interviews, polygraphs) and are beneficial for positions requiring employees to interact frequently with others or handle sensitive information or valuable materials.

Considerations:

- Validity – Biodata measures have been shown to be effective predictors of job success (i.e., they have a moderate degree of criterion-related validity) in numerous settings and for a wide range of criterion types (e.g., overall performance, customer service, team work); Biodata measures have also appeared to add additional validity (i.e., incremental validity) to selection systems employing traditional ability measures

- Face Validity/Applicant Reactions – Because some biodata items may not appear to be job related (i.e., low face validity) applicants may react to biodata tests as being unfair and invasive

- Administration Method – Administered individually but can be administered to large numbers of applicants via paper and pencil or electronically at one time

- Subgroup Differences – Typically have less adverse impact on minority groups than do many other types of selection measures; Items should be carefully written to avoid stereotyping and should be based on experiences under a person's control (i.e., what a person did rather than what was done to the person)

- Development Costs – The development of biodata items, scoring strategies, and validation procedures is a difficult and time-consuming task requiring considerable expertise; Large samples of applicants are needed to develop as well as validate the scoring strategy and additional samples may be needed to monitor the validity of the items for future applicants

- Administration Costs – Can be cost effective to administer and generally not time consuming to score if an automated scoring system is implemented

- Utility/ROI – High predictive ability can allow for the identification and selection of top performers; Benefits (e.g., savings in training, high productivity, decreased turnover) can outweigh developmental and administrative costs

- Common Uses – Commonly used in addition to cognitive ability tests to increase validity and lower adverse impact

References: (See Section VI for a summary of each article)

Elkins, T., & Phillips, J. (2000). Job context, selection decision outcome, and the perceived fairness of selection tests: Biodata as an illustrative case. *Journal of Applied Psychology, 85*(3), 479-484.

Hough, L. M., & Oswald, F. L. (2000). Personnel selection: Looking toward the future—Remembering the past. *Annual Review of Psychology, 51*, 631-664.

Mount, M. K., Witt, L. A., & Barrick, M. R. (2000). Incremental validity of empirically keyed biodata scales over GMA and the five factor personality constructs. *Personnel Psychology, 53*(2), 299-323.

Rothstein, H. R., Schmidt, F. L., Erwin, F. W., Owens, W. A., & Sparks, C. P. (1990). Biographical data in employment selection: Can validities be made generalizable? *Journal of Applied Psychology, 75*(2), 175-184.

Schmitt, N., Cortina, J. M., Ingerick, M. J., & Wiechmann, D. (2003). Personnel selection and employee performance. *Handbook of Psychology: Industrial and Organizational Psychology, 12*, 77-105. New York, NY: John Wiley & Sons, Inc.

The following Society for Industrial and Organizational Psychology (SIOP) website contains information on Biographical Data:
http://www.siop.org/workplace/employment%20testing/employment_testing_toc.aspx

Cognitive Ability Tests: Cognitive ability tests assess abilities involved in thinking (e.g., reasoning, perception, memory, verbal and mathematical ability, and problem solving). Such tests pose questions designed to estimate applicants' potential to use mental processes to solve work-related problems or to acquire new job knowledge.

Traditionally, the general trait measured by cognitive ability tests is called "intelligence" or "general mental ability." However, an intelligence test often includes various item types which measure different and more specific mental factors often referred to as "specific mental abilities." Examples of such items include arithmetic computations, verbal analogies, reading comprehension, number series completion, and spatial relations (i.e., visualizing objects in three-dimensional space).

Some cognitive ability tests sum up the correct answers to all of the items to obtain an overall score that represents a measure of general mental ability. If an individual score is computed for each of the specific types of abilities (e.g., numeric, verbal, reasoning), then the resulting scores represent measures of the specific mental abilities.

Traditional cognitive tests are well-standardized, contain items reliably scored, and can be administered to large groups of people at one time. Examples of item formats include multiple choice, sentence completion, short answer, or true-false. Many professionally developed cognitive tests are available commercially and may be considered when there is no significant need to develop a test that refers specifically to the particular job or organization.

Considerations:

- Validity – Tests of general cognitive ability are good predictors of job performance and training success for a wide variety of jobs (i.e., they have a high degree of criterion-related validity); The more complex the job or training demands, the better these tests work; Other predictors may add only small amounts of incremental validity over cognitive tests

- Face Validity/Applicant Reactions – Tests developed to refer explicitly to specific jobs or types of jobs within the hiring organization may be viewed as more highly related to the job (i.e., high face validity) than commercially developed tests

- Administration Method – Can be administered via paper and pencil or electronically

- Subgroup Differences – Cognitive ability tests typically produce racial and ethnic differences larger than other valid predictors of job performance such as biodata, personality tests, and structured interviews; The use of other assessment methods (e.g., interviews, biodata instruments) in combination with cognitive ability tests is recommended to lower any potential adverse impact

- Development Costs – Cost of purchasing a cognitive test is typically less expensive than developing a customized test

- Administration Costs – Generally inexpensive, requires few resources for administration, and does not require skilled administrators

- Utility/ROI – High return on investment if you need applicants who possess particular cognitive abilities or have high potential to acquire job knowledge or benefit from training; Cost effectiveness of developing own test over purchasing a commercial test is lower when face validity is not an issue

- Common Uses – Best used for jobs requiring particular cognitive abilities for effective job performance and for more complex jobs

References: (See Section VI for a summary of each article)

Hunter, J. E. (1986). Cognitive ability, cognitive aptitude, job knowledge, and job performance. *Journal of Vocational Behavior, 29*(3), 340-362.

Murphy, K. R., Cronin, B. E., & Tam, A. P. (2003). Controversy and consensus regarding the use of cognitive ability testing in organizations. *Journal of Applied Psychology, 88*(4), 660-671.

Outtz, J. L. (2002). The role of cognitive ability tests in employment selection. *Human Performance, 15*(1-2), 161-172.

Ree, M. J., Earles, J. A., & Teachout, M. S. (1994). Predicting job performance: Not much more than g. *Journal of Applied Psychology, 79*(4), 518-524.

Schmidt, F. L., & Hunter, J. (2004). General mental ability in the world of work: Occupational attainment and job performance. *Journal of Personality & Social Psychology, 86*(1), 162-173.

NOTE: The following Society for Industrial and Organizational Psychology (SIOP) website contains information on Cognitive Ability Tests:
http://www.siop.org/workplace/employment%20testing/employment_testing_toc.aspx

Emotional Intelligence Tests: Emotional intelligence (EI) is defined as a type of social competence involving the ability to monitor one's own and others' emotions, to discriminate among them, and to use the information to guide one's thinking and actions. EI is a fairly specific ability that connects a person's knowledge processes to his or her emotional processes. As such, EI is different from emotions, emotional styles, emotional traits, and traditional measures of intelligence based on general mental or cognitive ability (i.e., IQ). EI involves a set of skills or abilities that may be categorized into five domains:

- Self-awareness: Observing yourself and recognizing a feeling as it happens.

- Managing emotions: Handling feelings so they are appropriate; realizing what is behind a feeling; finding ways to handle fears and anxieties, anger, and sadness.

- Motivating oneself: Channeling emotions in the service of a goal; emotional self-control; delaying gratification and stifling impulses.

- Empathy: Sensitivity to others' feelings and concerns and taking their perspective; appreciating the differences in how people feel about things.

- Handling relationships: Managing emotions in others; social competence and social skills.

The typical approach to measuring EI ability involves administering a set of questions to applicants and scoring the correctness of those responses based on expert judgment (expert scoring) or consensus among a large number of people (consensus scoring). For example, one EI ability test requires the applicant to view a series of faces and report how much of each of six emotions is present, answer questions about emotional scenarios and responses (e.g., predict how an anxious employee will react to a significantly increased workload), and solve emotional problems (e.g., decide what response is appropriate when a friend calls you upset over losing his or her job).

Some tests of EI use a self-report method. Self-report questionnaires are commonly used to measure personality traits (e.g., extroversion, agreeableness, conscientiousness). Self-report assessments have been around for decades and serve a very useful purpose. As a way to measure EI abilities, they have some drawbacks. Using a self-report approach has been compared to estimating typing skill by asking applicants a series of questions about how quickly and accurately they can type. Does this mean self-report measures of emotional intelligence should not be used? If the objective is to measure a person's self-perceived competence or self-image, then this may be the preferred approach. If the objective is to measure EI as a set of abilities, skills, or emotional competencies, then self-report may not be the best method to use. To the extent employers are concerned with fakability of self-reports, ability models of EI will be more acceptable.

Considerations:

- Validity – Ability-based tests of emotional intelligence have been shown to contribute to the prediction of job performance, particularly when the maintenance of positive interpersonal relations is important to job success

- Face Validity/Applicant Reactions – Test items appearing to measure social skill generally have good face validity (e.g., identifying emotions expressed in a photograph of a person's face); Applicants may have a difficult time determining the best answer on some of the items; Some items may not appear to be directly work-related

- Administration Method – Can be administered via paper and pencil or electronically

- Subgroup Differences – There is some evidence women tend to score better than men on tests of emotional intelligence, which is consistent with other research showing women are more skilled at reading facial expressions of emotions than are men

- Development Costs – Cost of purchasing an emotional intelligence test is typically far less expensive than developing a customized test

- Administration Costs – Generally inexpensive, requires few resources for administration, and does not require skilled administrators

- Utility/ROI – High return on investment if applicants are needed who possess strong interpersonal skills

- Common Uses – Used with occupations requiring high levels of social interaction, cooperation, and teamwork

References: (See Section VI for a summary of each article)

Brackett, M. A., Rivers, S. E., Shiffman, S., Lerner, N., & Salovey, P. (2006). Relating emotional abilities to social functioning: A comparison of self-report and performance measures of emotional intelligence. *Journal of Personality and Social Psychology*, *91*, 780-795.

Frost, D. E. (2004). The psychological assessment of emotional intelligence. In J. C. Thomas & M. Hersen (Eds.), *Comprehensive handbook of psychological assessment, Volume 4: Industrial and organizational psychology* (pp. 203-215). Hoboken, NJ: John Wiley & Sons.

Mayer, J. D., Salovey, P., & Caruso, D. R. (2004). Emotional intelligence: Theory, findings, and implications. *Psychological Inquiry, 15*, 197-215.

Salovey, P., & Mayer, J. D. (1990). Emotional intelligence. *Imagination, Cognition, and Personality*, *9*, 185-211.

Integrity/Honesty Tests: An integrity test is a specific type of personality test designed to assess an applicant's tendency to be honest, trustworthy, and dependable. A lack of integrity is associated with such counterproductive behaviors as theft, violence, sabotage, disciplinary problems, and absenteeism. Integrity tests have been found to measure some of the same factors as standard personality tests, particularly conscientiousness, and perhaps some aspects of emotional stability and agreeableness.

Integrity tests can also be valid measures of overall job performance. This is not surprising because integrity is strongly related to conscientiousness, itself a strong predictor of overall job performance. Like other measures of personality traits, integrity tests can add a significant amount of validity to a selection process when administered in combination with a cognitive ability test. In addition, few, if any, integrity test performance differences are found between men and women or applicants of different races or ethnicities. Integrity tests will not eliminate dishonesty or theft at work, but the research does strongly suggest that individuals who score poorly on these tests tend to be less suitable and less productive employees.

Overt integrity tests (also referred to as clear-purpose tests) are designed to directly measure attitudes relating to dishonest behavior. They are distinguished from personality-based tests in that they make no attempt to disguise the purpose of the assessment. Overt tests often contain questions that ask directly about the applicant's own involvement in illegal behavior or wrongdoing (e.g., theft, illicit drug use). Such transparency can make guessing the correct answer obvious. Applicant faking is always a concern with overt integrity tests. The score results from such tests should be interpreted with caution.

Considerations:

- Validity – Integrity tests have been shown to be valid predictors of overall job performance as well as many counterproductive behaviors such as absenteeism, illicit drug use, and theft; The use of integrity tests in combination with cognitive ability tests can substantially enhance the prediction of overall job performance (i.e., high degree of incremental validity)

- Face Validity/Applicant Reactions – May contain items that do not appear to be job related (i.e., low face validity) or seem to reveal applicants' private thoughts and feelings; Applicants may react to integrity tests as being unnecessarily invasive, but strong negative reactions have been found to be rare; Some item types may be highly transparent making it easy for applicants to fake or distort test scores in their favor

- Administration Method – Can be administered via paper and pencil or electronically

- Subgroup Differences – Generally, few, if any, average score differences are found between men and women or applicants of different races or ethnicities, therefore it is beneficial to use an integrity measure when another measure with greater potential for adverse impact (e.g., a cognitive test) is included in the selection process; Both overt and personality-based integrity test scores seem to be correlated with age indicating younger

27

individuals have the potential to be more counterproductive employees, possibly because of a youthful tendency towards drug experimentation and other social deviance

- Development Costs – The cost of purchasing an integrity test is typically less expensive than developing a customized test

- Administration Costs – Generally inexpensive, requires few resources for administration, and does not require skilled administrators

- Utility/ROI – High return on investment in settings where counterproductive behaviors (e.g., theft of valuable property or sensitive information, absenteeism) are highly disruptive to organizational functioning

- Common Uses – Typically used to measure whether applicants have the potential to be successful in jobs where performance requires a high level of honesty and dependability; Frequently administered to large groups of applicants as a screen-out measure

References: (See Section VI for a summary of each article)

Cullen, M. J., & Sackett, P. R. (2004). Integrity testing in the workplace. In J. C. Thomas & M. Hersen (Eds.), *Comprehensive handbook of psychological assessment, Volume 4: Industrial and organizational psychology* (pp. 149-165). Hoboken, NJ: John Wiley & Sons.

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*, *78,* 679-703.

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*(4), 787-829.

NOTE: The following Society for Industrial and Organizational Psychology (SIOP) website contains information on Integrity Tests:
http://www.siop.org/workplace/employment%20testing/employment_testing_toc.aspx

Job Knowledge Tests: Job knowledge tests, sometimes referred to as achievement or mastery tests, typically consist of questions designed to assess technical or professional expertise in specific knowledge areas. Job knowledge tests evaluate what a person knows at the time of taking the test. Unlike cognitive ability tests, there is no attempt to assess the applicant's learning potential. In other words, a job knowledge test can be used to inform employers what an applicant currently knows, but not whether the individual can be relied on to master new material in a timely manner. Job knowledge tests are not appropriate when applicants will be trained after selection in the critical knowledge areas needed for the job.

Job knowledge tests are used in situations where applicants must already possess a body of learned information prior to being hired. They are particularly useful for jobs requiring specialized or technical knowledge that can only be acquired over an extended period of time. Examples of job knowledge tests include tests of basic accounting principles, computer programming, financial management, and knowledge of contract law. Job knowledge tests are often constructed on the basis of an analysis of the tasks that make up the job. While the most typical format for a job knowledge test is a multiple choice question format, other formats include written essays and fill-in-the-blank questions.

Licensing exams, agency certification, and/or professional certification programs are also job knowledge tests. Licensure and certification are both types of credentialing—the process of granting a designation that indicates competence in a subject or area. Licensure is more restrictive than certification and typically refers to the mandatory Governmental requirement necessary to practice in a particular profession or occupation. A passing score on a job knowledge test is typically a core requirement to obtain a professional license. Licensure implies practice and title protection. This means only individuals who hold a license are permitted to practice and use a particular title. For example, to practice law, a law school graduate must apply for admission into a state bar association that requires passing the bar licensure examination. Certification is usually a voluntary process instituted within a nongovernmental or single Governmental agency in which individuals are recognized for advanced knowledge and skill. As with licensure, certification typically requires a passing score on a job knowledge exam.

Considerations:

- Validity – Knowledge areas tested are very representative of those required to perform the job (i.e., high degree of content validity); Performance on job knowledge tests relates highly to performance on the job (i.e., high degree of criterion-related validity); Can add a substantial amount of incremental validity above and beyond the validity provided by general cognitive ability tests; Customized job knowledge tests have been shown to have slightly higher validity than off-the-shelf tests

- Face Validity/Applicant Reactions – Applicants often perceive job knowledge tests as being very fair (i.e., as having a high degree of face validity) because such tests are typically designed to measure knowledge directly applied to performance of the job

- Administration Method – Can be administered via paper and pencil or electronically

- Subgroup Differences – Tend to produce race and ethnic group differences larger than other valid predictors of job performance (e.g., work sample tests, personality tests)

- Development Costs – Typically expensive and time consuming to develop; Frequent updates to the test content and validation may be needed to keep up with changes in the job; Cost of purchasing an off-the-shelf job knowledge test is typically less expensive than developing a customized test

- Administration Costs – Generally inexpensive and requires few resources for administration

- Utility/ROI – High return on investment if you need applicants who possess technical expertise in specific job knowledge areas; Utility is lower when the job knowledge test contributes little to the prediction of job performance above and beyond inexpensive and readily available cognitive ability tests

- Common Uses – Best used for jobs requiring specific job knowledge on the first day of the job (i.e., where the knowledge is needed upon entry to the position)

References: (See Section VI for a summary of each article)

Dubois, D., Shalin, V. L., Levi, K. R., & Borman, W. C. (1993). Job knowledge test design: A cognitively-oriented approach. *U.S. Office of Naval Research Report*, *Institute Report 241*, i-47.

Dye, D. A., Reck, M., & McDaniel, M. A. (1993). The validity of job knowledge measures. *International Journal of Selection and Assessment, 1,* 153-157.

Ree, M. J., Carretta, T. R., & Teachout, M. S. (1995). Role of ability and prior job knowledge in complex training performance. *Journal of Applied Psychology, 80*(6), 721-730.

Roth, P. L., Huffcutt, A. I., & Bobko, P. (2003). Ethnic group differences in measures of job performance: A new meta-analysis. *Journal of Applied Psychology, 88*(4), 694-706.

Sapitula, L., & Shartzer, M. C. (2001). Predicting the job performance of maintenance workers using a job knowledge test and a mechanical aptitude test. *Applied H.R.M. Research, 6*(1-2), 71-74.

NOTE: The following Society for Industrial and Organizational Psychology (SIOP) website contains information on Job Knowledge Tests:
http://www.siop.org/workplace/employment%20testing/employment_testing_toc.aspx

Personality Tests: Personality tests are designed to systematically elicit information about a person's motivations, preferences, interests, emotional make-up, and style of interacting with people and situations. Personality measures can be in the form of interviews, in-basket exercises, observer ratings, or self-report inventories (i.e., questionnaires).

Personality self-report inventories typically ask applicants to rate their level of agreement with a series of statements designed to measure their standing on relatively stable personality traits. This information is used to generate a profile used to predict job performance or satisfaction with certain aspects of the work.

Personality is described using a combination of traits or dimensions. Therefore, it is ill-advised to use a measure that taps only one specific dimension (e.g., conscientiousness). Rather, job performance outcomes are usually best predicted by a combination of personality scales. For example, people high in integrity may follow the rules and be easy to supervise but they may not be good at providing customer service because they are not outgoing, patient, and friendly. The personality traits most frequently assessed in work situations include: (1) Extroversion, (2) Emotional Stability, (3) Agreeableness, (4) Conscientiousness, and (5) Openness to Experience. These five personality traits are often referred to collectively as the Big Five or the Five-Factor Model. While these are the most commonly measured traits, the specific factors most predictive of job performance will depend on the job in question. When selecting or developing a personality scale, it is useful to begin with inventories that tap the Big Five, but the results from a validity study may indicate some of these traits are more relevant than others in predicting job performance.

It is important to recognize some personality tests are designed to diagnose psychiatric conditions (e.g., paranoia, schizophrenia, compulsive disorders) rather than work-related personality traits. The Americans with Disabilities Act considers any test designed to reveal such psychiatric disorders as a "medical examination." Examples of such medical tests include the Minnesota Multiphasic Personality Inventory (MMPI) and the Millon Clinical Multi-Axial Inventory (MCMI). Generally speaking, personality tests used to make employment decisions should be specifically designed for use with normal adult populations. Under the Americans with Disabilities Act, personality tests meeting the definition of a medical examination may only be administered after an offer of employment has been made.

Considerations:

- Validity – Personality tests have been shown to be valid predictors of job performance (i.e., they have an acceptable level of criterion-related validity) in numerous settings and for a wide range of criterion types (e.g., overall performance, customer service, team work), but tend to be less valid than other types of predictors such as cognitive ability tests, assessment centers, and work samples and simulations

- Face Validity/Applicant Reactions – May contain items that do not appear to be job related (i.e., low face validity) or seem to reveal applicants' private thoughts and feelings; Applicants may react to personality tests as being unnecessarily invasive; Items may also

be highly transparent, making it easy for applicants to fake or distort test scores in their favor

- Administration Method – Can be administered via paper and pencil or electronically

- Subgroup Differences – Generally, few, if any, average score differences are found between men and women or applicants of different races or ethnicities, therefore it is beneficial to use a personality measure when another measure with greater potential for adverse impact (e.g., cognitive ability test) is included in the selection process

- Development Costs – Cost of purchasing a personality test is typically less expensive than developing a customized test

- Administration Costs – Generally inexpensive, requires few resources for administration, and does not require skilled administrators

- Utility/ROI – High return on investment if you need applicants who possess strong interpersonal skills or other job-related specific personality traits

- Common Uses – Typically used to measure whether applicants have the potential to be successful in jobs where performance requires a great deal of interpersonal interaction or work in team settings; Less useful for highly scripted jobs where personality has little room to take effect; Frequently administered to large groups of applicants as a screen

References: (See Section VI for a summary of each article)

Barrick, M. R., & Mount, M. K. (1991). The Big Five personality dimensions and job performance: A meta-analysis. *Personnel Psychology, 44*, 1-26.

Hogan, R., Hogan, J., & Roberts, B. W. (1996). Personality measurement and employment decisions: Questions and answers. *American Psychologist, 51*, 469-477.

Hough, L. M., Eaton, N. K., Dunnette, M. D., Kamp, J. D., & McCloy, R. A. (1990). Criterion-related validities of personality constructs and the effect of response distortion on those validities. *Journal of Applied Psychology, 75*, 581-595.

Hough, L. M., & Oswald, F. L. (2000). Personnel selection: Looking toward the future— Remembering the past. *Annual Review of Psychology, 51*, 631-664.

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.

NOTE: The following Society for Industrial and Organizational Psychology (SIOP) website contains information on Personality Tests:
http://www.siop.org/workplace/employment%20testing/employment_testing_toc.aspx

Reference Checking: Reference checking is an objective evaluation of an applicant's past job performance based on information collected from key individuals (e.g., supervisors, peers, subordinates) who have known and worked with the applicant. Reference checking is primarily used to:

- Verify the accuracy of information given by job applicants through other selection processes (e.g., résumés, occupational questionnaires, interviews)

- Predict the success of job applicants by comparing their experience to the competencies required by the job

- Uncover background information on applicants that may not have been identified by other selection procedures

Job applicants may attempt to enhance their chances of obtaining a job offer by distorting their training and work history information. While résumés summarize what applicants claim to have accomplished, reference checking is meant to assess how well those claims are backed up by others. Verifying critical employment information can significantly cut down on selection errors. Information provided by former peers, direct reports, and supervisors can also be used to forecast how applicants will perform in the job being filled. Reference data used in this way is based on the behavioral consistency principle that past performance is a good predictor of future performance.

As a practical matter, reference checking is usually conducted near the end of the selection process after the field of applicants has been narrowed to only a few competitors. Most reference checks are conducted by phone. Compared to written requests, phone interviews allow the checker to collect reference data immediately and to probe for more detailed information when clarification is needed. Phone interviews also require less time and effort on the part of the contact person and allow for more candid responses about applicants.

Reference checking has been shown to be a useful predictor of job performance (as measured by supervisory ratings), training success, promotion potential, and employee turnover. As with employment interviews, adding structure to the reference checking process can greatly enhance its validity and usefulness as an employee selection procedure. Strategies for structuring reference checking include basing questions on a job analysis, asking applicants the same set of questions, and providing interviewers with standardized data collection and rating procedures.

Conducting reference checks can reduce the risk of lawsuits for negligent hiring—the failure to exercise reasonable care when selecting new employees. Providing accurate information when called as a reference for a former employee is equally important, but many employers refuse to give negative information about former employees, fearing a lawsuit for defamation. This is generally not deemed a serious problem for Federal reference providers and reference checkers because of legal protections provided under the Federal Tort Claims Act.

Considerations:

- Validity – Reference checks are useful for predicting applicant job performance, better than years of education or job experience, but not as effective as cognitive ability tests; Reference checks can add incremental validity when used with other selection procedures, such as cognitive ability and self-report measures of personality; Adding structure (as is done with employment interviews) can enhance their effectiveness

- Face Validity/Applicant Reactions – Some applicants may view reference checks as invasive

- Administration Method – Reference checks are typically collected by phone using a structured interview format; Written requests for work histories typically result in low response rates and less useful information

- Subgroup Differences – Generally little or no score differences are found between men and women or applicants of different races; Employers should be especially careful to avoid asking questions not directly related to the job

- Development Costs – Costs are generally low and depend on the complexity of the job, the number of questions needed, competencies measured, and development and administration of checker/interviewer training

- Administration Costs – Generally inexpensive, structured telephone reference checks take about 20 minutes to conduct per contact, a minimum of three contacts is recommended

- Utility/ROI – Used properly, reference checks can reduce selection errors and enhance the quality of new hires at a minimal cost to the agency

- Common Uses – Best used in the final stages of a multiple-hurdle selection process when deciding among a handful of finalists

References: (See Section VI for a summary of each article)

Aamodt, M. G. (2006). Validity of recommendations and references. *Assessment Council News*, February, 4-6.

Taylor, P. J., Pajo, K., Cheung, G. W., & Stringfield, P. (2004). Dimensionality and validity of a structured telephone reference check procedure. *Personnel Psychology, 57*, 745-772.

U.S. Merit Systems Protection Board. (2005). Reference checking in federal hiring: Making the call. Washington, DC: Author. Note: Report available at: http://www.mspb.gov/netsearch/viewdocs.aspx?docnumber=224106&version=224325&application=ACROBAT

Situational Judgment Tests: Situational judgment tests (SJTs) present applicants with a description of a work problem or critical situation related to the job they are applying for and ask them to identify how they would handle it. Because applicants are not placed in a simulated work setting and are not asked to perform the task or behavior (as would be the case in an assessment center or a work sample), SJTs are classified as low-fidelity simulations.

SJTs measure effectiveness in social functioning dimensions such as conflict management, interpersonal skills, problem solving, negotiation skills, facilitating teamwork, and cultural awareness. SJTs are particularly effective measures of managerial and leadership competencies.

SJTs can be developed to present scenarios and collect responses using a variety of formats. One alternative is to present a situation and then ask respondents to answer several questions about the situation. More often, SJTs present a new situation for each question. To respond to this type of SJT item, applicants may be asked: a) what they would do in the particular situation, b) what they would be most and least likely to do in the situation, c) what response is the best response among several options, d) what response is the best and second-best among several options, or e) what would most likely occur next in a certain situation or as a result of a certain decision.

SJTs can be presented in either a linear or interactive format. With a linear format, all respondents are presented with the same questions and in the same order. With an interactive (usually computer administered) format, SJTs can be structured according to a branching process in which the scenarios and response options presented later in the test depend on how applicants responded to questions presented earlier in the test. SJT questions and alternatives are typically based on critical incidents generated by subject matter (i.e., job) experts. Scores are based on subject matter experts' judgments of the best and worst alternatives.

Considerations:

- Validity – The tasks and activities described in the SJT scenarios are very representative of the tasks and activities found on the job (i.e., they have a high degree of content validity) and performance on the tests moderately relates to performance on the job (i.e., they have a moderately high degree of criterion-related validity)

- Face Validity/Applicant Reactions – Applicants often perceive SJTs as being very fair (i.e., the tests have a high degree of face validity)

- Administration Method – Possible to administer in paper and pencil, computer-based, or video-based format

- Subgroup Differences – Subgroup differences are typically moderate; Racial differences in test scores may be smaller than those typically observed for cognitive ability tests

- Development Costs – Generally, developmental costs are less than high-fidelity alternatives (e.g., assessment centers) and depend on costs related to use of subject matter experts

- Administration Costs – Administration costs are typically low when delivered via paper and pencil, but may be more costly via computer or video; No special administrator expertise is needed

- Utility/ROI – High return on investment if you need applicants who possess a high level of social and interpersonal skills upon entry into the job; If the skills measured by the tests can be learned on the job or are not highly critical, then the return on investment will be significantly lower

- Common Uses – SJTs can be developed for a variety of jobs, but are typically used for managerial positions or other jobs requiring effective interpersonal interactions

References: (See Section VI for a summary of each article)

Hanson, M. A., Horgen, K. E., & Borman W. C. (1998, April). Situational judgment tests (SJT) as measures of knowledge/expertise. Paper presented as the 13th Annual Conference of the Society for Industrial and Organizational Psychology, Dallas, TX.

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

McDaniel, M. A., Whetzel, D. L., & Nguyen, N. T. (2006). *Situational judgment tests for personnel selection.* Alexandria, VA: IPMA Assessment Council.

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

Motowidlo, S. J., & Tippins, N. (1993). Further studies of the low-fidelity simulation in the form of a situational inventory. *Journal of Occupational and Organizational Psychology, 66,* 337-344.

Weekley, J. A., & Jones, C. (1999). Further studies of situational tests. *Personnel Psychology*, *52*(3), 679-700.

Structured Interviews: The employment interview is one of the most widely used methods of assessing job applicants. Due to its popularity, a great deal of research on improving the reliability and validity of the interview has been conducted. This body of research has demonstrated that structured interviews, which employ rules for eliciting, observing, and evaluating responses, increase interviewers' agreement on their overall evaluations by limiting the amount of discretion an interviewer is allowed.

The level of structure in an interview can vary according to the constraints placed on the questions asked and evaluation criteria. Interviews with a low degree of structure place no constraints on the questions asked and allow for global evaluation of applicant responses. Interviews with a very high level of structure involve asking all applicants the same exact set of pre-defined lead and probe (i.e., follow-up questions) and are scored according to benchmarks of proficiency. Interviews with higher degrees of structure show higher levels of validity, rater reliability, rater agreement, and less adverse impact.

Interviews also vary according to the specific competencies being measured. Employment interviews can focus on past, present, or future behavior, beliefs, opinions, or attitudes of the applicant. Information may also include behavior observed during the interview itself (e.g., oral communication), work experience, training, education, and career aspirations. Research shows interview questions based on specific job competencies identified through job analysis as being critical to job success demonstrate high levels of validity, rater reliability, and rater agreement.

The most common methods for developing specific, job-related questions are based on either the situational or behavioral description format. Situational interview questions ask applicants to describe what they would do or how they would behave in a situation similar to those encountered on the job. An example of a situational question is, "You have been assigned to work on a project with some of your coworkers. While on the job, you notice several of them goofing off. You know you are falling behind schedule to complete the work by the deadline. What would you do?" This format relies on applicants' ability to project what they might do in a future situation. Behavioral description interview questions ask applicants to describe a past behavior demonstrated in a situation relevant to the competency of interest. An example of this type is, "Describe a situation where you analyzed and interpreted information." This type of interview is based on the behavioral consistency principle that past behavior is the best predictor of future behavior.

Both methods rely on the development of job-related critical incidents (e.g., examples of notably good or notably bad job performance). Both types of interview formats have proven to be effective. Behavioral description interviews have shown higher levels of validity where the nature of the work is highly complex (e.g., professional and managerial level jobs). Structured interviews are usually scored by a panel in which each member rates applicant responses individually and then participates in a group discussion to resolve significant scoring discrepancies. Faking occurs less frequently during structured interviews than paper-and-pencil inventories measuring the same competencies.

Considerations:

- Validity – Situations presented in structured interview questions are highly representative of the situation encountered on the job (i.e., a high degree of content validity), performance on structured interviews relates highly to performance on the job (i.e., a high degree of criterion-related validity), and show moderate relationships with measures of cognitive ability and personality (i.e., construct validity); Can add validity beyond other selection measures (i.e., a high degree of incremental validity), such as cognitive ability and personality variables

- Face Validity/Applicant Reactions – Interviews typically result in more favorable applicant reactions compared to other popular selection measures, however, interviewers and applicants tend to favor less structured formats

- Administration Method – Individual administration only; Can be conducted face-to-face or over the telephone or via video conference

- Subgroup Differences – Generally little or no performance differences are found between men/women or applicants of different races; Some differences in interviewer ratings have been found for different races (dependent on race of candidate and interviewer(s), job complexity, and types of questions asked)

- Development Costs – Costs are generally low and depend on the complexity of the job, the number of questions used, dimensions measured, and development and administration of interviewer/rater training

- Administration Costs – Typically not costly to administer but may depend on costs related to train interviewers, rater time required, and number of applicants to assess

- Utility/ROI – High return on investment if you need applicants who possess specific, critical competencies upon entry into the job; If the competencies measured by the interview can be learned on the job or are not highly critical, then the return on investment will be significantly lower

- Common Uses – Used for recruitment, selection and promotion purposes; Frequently used late in the assessment process as a final screen or in situations where the applicant pool is moderate or small in size

References: (See Section VI for a summary of each article)

Campion, M. A., Palmer, D. K., & Campion, J. E. (1997). A review of structure in the selection interview. *Personnel Psychology, 50*(3), 655-702.

Conway, J. M., Jako, R. A., & Goodman, D. F. (1995). A meta-analysis of interrater and internal consistency reliability of selection interviews. *Journal of Applied Psychology, 80*(5), 565-579.

Huffcutt, A I., & Arthur, W. (1994). Hunter and Hunter (1984) revisited: Interview validity for entry-level jobs. *Journal of Applied Psychology*, *79*(2), 184-190.

Huffcutt, A. I., & Roth, P. L. (1998). Racial group differences in employment interview evaluations. *Journal of Applied Psychology, 83*(2), 179-189.

Huffcutt, A. I., Weekley, J. A., Wiesner, W. H., DeGroot, T. G., & Jones, C. (2001). Comparison of situational and behavior description interview questions for higher-level positions. *Personnel Psychology*, *54*(3), 619-644.

McFarland, L. A., Ryan, A. M., Sacco, J. M., Kriska, S. D. (2004). Examination of structured interview ratings across time: The effects of applicant race, rater race, and panel composition. *Journal of Management, 30*(4), 435-452.

Taylor, P., & Small, B. (2002). Asking applicants what they would do versus what they did do: A meta-analytic comparison of situational and past behavior employment interview questions. *Journal of Occupational & Organizational Psychology, 75*(3), 277-294.

The following Society for Industrial and Organizational Psychology (SIOP) website contains information on Interviews:
http://www.siop.org/workplace/employment%20testing/employment_testing_toc.aspx

Training and Experience (T & E) Evaluations: A traditional T & E evaluation, sometimes called a crediting plan or rating schedule, is a systematic method used to assess previous experience, education, and training information provided by job applicants. These assessment factors are based on critical job requirements and competencies identified through a job analysis.

Rating factors generally include the amount and quality of the applicant's previous job-related experience, as well as any other information deemed important to performing the duties of the position. Typically, information on the assessment factors is reported by applicants as a supplement to the application blank. This information is evaluated against education and experience benchmarks to generate scores for selection purposes. Benchmarks are often developed by Human Resource Specialists familiar with the occupations covered with the T & E evaluation.

T & E evaluations are relatively easy to develop and may apply to multiple occupations sharing the same requirements and competencies. For the most part, these assessments are used for entry level positions. Most often, T & E evaluations are used as a screen early in the selection process to identify applicants who meet the minimum proficiency levels on the rating factors. While most rating factors are usually broad, more specific factors tailored to a particular occupation or organization can be developed.

A variation of the traditional rating schedule based on training and experience rating factors is a task-based rating method. The task-based method is used to assess applicants' training and experience in relation to descriptions of tasks performed on the job to be filled. Specifically, the task-based rating schedule is developed from a list of tasks performed by incumbents in the target job. Applicants read each task statement and indicate whether they have ever performed such activities. Some versions ask applicants to also indicate the level of proficiency at which the task was performed. Generally, the more tasks performed, the higher an applicant's score will be.

As with most self-report instruments, applicant inflation or distortion can threaten the validity of a T & E evaluation. Two approaches can be taken to combat the problem of rating inflation: (1) creating applicant expectations that responses will be verified, and (2) carrying out verification procedures, making adjustments to scores based on the findings.

Other self-report measures that collect additional types of training and experience information are available as alternatives to the traditional T & E evaluation. An example of such an alternative is the competency-based self-report method. This method functions much like a traditional rating schedule in terms of ease of administration and scoring. However, in addition to rating the extent to which a critical job competency is demonstrated, accomplishments, (e.g., written statements of personal accomplishments that best illustrate an applicant's proficiency on critical job dimensions) are collected to support the self-reported information. This is very similar to the accomplishment records method discussed earlier in this section. Another option with the competency-based self-report method is the inclusion of a process requiring formal verification (e.g., via reference checking) of the information provided by the applicants in their written self-ratings and/or accomplishments. This verification information is often used to limit,

as much as possible, the rating inflation typically observed with applicant self-reports of accomplishments.

Considerations:

- Validity – The content of the training and experience items on a traditional rating schedule and the task items on a task-based rating schedule are often highly representative of actual job performance (i.e., they show a high degree of content validity); Generally, performance on rating schedules does not relate well to performance on the job (i.e., they show a low degree of criterion-related validity), with length and recency of education, academic achievement, and extracurricular activities demonstrating the weakest relation to job performance

- Face Validity/Applicant Reactions – Reactions from professionals who feel they should be evaluated on their experience is typically favorable; Less favorable reactions may be seen if used for younger, less experienced applicants with few previous related experiences to describe

- Administration Method – Can be administered via paper-and-pencil or electronically

- Subgroup Differences – Generally little or no performance differences are found between men and women or applicants of different racial or ethnic backgrounds

- Development Costs – Takes less time to develop than other measures of training and experience (e.g., the accomplishment record)

- Administration Costs – Takes a very short time to administer and for applicants to complete; Administration time is shorter than other measures of training and experience (e.g., accomplishment record)

- Utility/ROI – Return on investment for training and experience measures can be moderate to high if the same rating schedule instrument can be used to assess for various positions

- Common Uses – Commonly used as a screening device prior to another selection tool (e.g., structured interview) for both entry level positions across various professional occupations (e.g., trainee positions) and jobs requiring prior preparation.

References: (See Section VI for a summary of each article)

Lyons, T. J. (1989). *Validity of Education and Experience Measured in Traditional Rating Schedule Procedures: A Review of the Literature*. Office of Personnel Research and Development, U.S. Office of Personnel Management, Washington, DC, OPRD-89-02.

Lyons, T. J. (1988). *Validity Research on Rating Schedule Methods: Status Report.* Office of Personnel Research and Development, U.S. Office of Personnel Management,

Washington, DC, OED-88-17.

McCauley, D. E. (1987). *Task-Based Rating Schedules: A Review.* Office of Examination Development, U.S. Office of Personnel Management, Washington, DC, OED 87-15.

McDaniel, M. A., Schmidt, F. L., & Hunter, J. E. (1988). A meta-analysis of the validity of methods for rating training and experience in personnel selection. *Personnel Psychology, 41*, 283-309.

Schwartz, D. J. (1977). A job sampling approach to merit system examining. *Personnel Psychology, 30*(2), 175-185.

Sproule, C. F. (1990). *Personnel Assessment Monographs: Recent Innovations in Public Sector Assessment* (Vol 2, No. 2)*.* International Personnel Management Association Assessment Council (IPMAAC).

Work Samples and Simulations: Work sample tests require applicants to perform tasks or work activities that mirror the tasks employees perform on the job. For instance, applicants for an Administrative Assistant position may be asked to transcribe an internal memo using a word processor or to accurately file a stack of paperwork. Because work samples require applicants to perform tasks identical or highly similar to tasks from the job, great care is taken in trying to mimic the work environment to the greatest extent possible. For example, applicants to the Administrative Assistant position may perform tasks at a workstation highly similar, if not identical, to that found on the job. As with job knowledge tests, work sample tests should only be used in situations where applicants are expected to possess the required competencies upon entry into the position. If training on how to perform the work activities will be provided after selection, the work sample method may not be appropriate.

In addition to work samples, which attempt to re-create specific work scenarios, performance tests can also be designed to mirror very broad aspects of the job that may draw on fundamental competencies needed to perform a wide range of job tasks. For example, the Administrative Assistant position mentioned above may require individuals to routinely find specific materials to answer various questions posed by upset or hostile customers. Rather than re-create a large number of scenarios to cover a wide array of situations, employers may design a single exercise to measure the general competencies in question (e.g., an interactive role-play between the applicant and a well-trained actor that measures applicant's problem solving, communication, and interpersonal skills). Applicant scores on work sample tests are generated by trained assessors who observe the applicant's behavior and/or by measuring task outcomes (e.g., the degree of interpersonal skills demonstrated or the number of errors made in transcribing an internal memo).

Considerations:

- Validity – Tasks applicants are asked to perform are very representative of the tasks performed on the job (i.e., they have a high degree of content validity) and performance on the tests relates highly to performance on the job (i.e., a high degree of criterion-related validity)

- Face Validity/Applicant Reactions – Applicants often perceive work samples as being very fair (i.e., a high degree of face validity)

- Administration Method – Often individual administration only (i.e., may not be suitable for group administrations); "Hands-on" performance by the applicant in a simulated work environment

- Subgroup Differences – Generally little or no performance differences are found between men and women or applicants of different races, although the presence of gender and/or racial differences depends on the competencies being assessed

- Development Costs – May be costly to develop, both in terms of time and money; May require periodic updating (e.g., if the task was using a typewriter to draft a document and

the organization becomes fully automated such that documents are now drafted using word processors)

- Administration Costs – May be time consuming and expensive to administer; Requires individuals to observe, and sometimes rate, applicant performance

- Utility/ROI – High return on investment if you need applicants who possess specific, critical competencies upon entry into the job; If the competencies measured by the tests can be learned on the job or are not highly critical then the return on investment will be significantly lower

- Common Uses – Best used for positions for which the measured competencies are highly critical for successful performance on the job, there is a limited number of applicants to test, and only a small number of prospective applicants are expected to have the needed competencies

References: (See Section VI for a summary of each article)

Campion, J. E. (1972). Work sampling for personnel selection. *Journal of Applied Psychology, 56*(1), 40-44.

Gilliland, S. W. (1995). Fairness from the applicants' perspective: Reactions to employee selection procedures. *International Journal of Selection and Assessment, 3*(1), 11-19.

Lance, C. E., Johnson, C. D., Douthitt, S. S., Bennett, W., & Harville, D. L. (2000). Good news: Work sample administrators' global performance judgments are (about) as valid as we've suspected. *Human Performance, 13*(3), 253-277.

Robertson, I. T. & Kandola, R. S. (1982). Work sample tests: Validity, adverse impact and applicant reaction. *Journal of Occupational Psychology, 55*(3), 171-183.

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*(2), 262-274.

The following Society for Industrial and Organizational Psychology (SIOP) website contains information on Work Samples and Simulations:
http://www.siop.org/workplace/employment%20testing/employment_testing_toc.aspx

Section IV: Other Assessment Methods

The following section presents some assessment methods outside the competency-based assessment methods included in Section III of the guide. Each method may play an important part in the recruitment and selection process, and summary information is provided for each.

Background Evaluation/Investigation: Background evaluations, sometimes referred to as background investigations, seek information about an applicant's employment, criminal, and personal history in an effort to investigate behavioral reliability, integrity, and personal adjustment. Background evaluations are conducted to determine whether there are any historical facts that would interfere with an applicant's ability to perform the job, including violations of statutes, regulations, or laws. It is important to note background evaluations are a different process than competency-based assessments and are typically handled apart from the traditional assessments (e.g., cognitive ability tests, accomplishment record). Depending on the extensiveness of the background evaluation, you may be required to gain the applicant's permission.

Background evaluation data are primarily used when screening personnel for positions of trust in which integrity and positive psychological adjustment is particularly desirable. Such occupations include law enforcement, private security industry, and positions requiring government-issued security clearances. The appointment of any civilian employee to a position in the Federal Government is subject to a background investigation.

Examples of factors investigated with a background evaluation are an applicant's employment history, past illegal drug use, and previous criminal records. In addition to collecting background information directly from an applicant, background information is sometimes collected from other sources who know the applicant such as former employers and co-workers, friends, and neighbors.

References: (See Section VI for a summary of each article)

Hilliard, P. A. (2001). Comparison of the predictive validity of a written test, an integrity test, a conscientiousness questionnaire, a structured behavioral interview and a personality inventory in the assessment of job applicants' background investigations, and subsequent task or contextual performance. *Dissertation Abstracts International: Section B: The Sciences & Engineering*, *62*(6-B), 2981.

McDaniel, M. A. (1989). Biographical constructs for predicting employee suitability. *Journal of Applied Psychology, 74*(6), 964-970.

McFadden, K. L. (1997). Policy improvements for prevention of alcohol misuse by airline pilots. *Human Factors, 39*(1), 1-8.

General Questions and Answers about OPM Background Investigations (2002). Retrieved February 5, 2007, from http://www.opm.gov/extra/investigate/QABackground.asp.

Job-Fit Measures: Job-fit measures (sometimes referred to as organization fit, person-organization fit, person-environment fit, or "fit check" tools), compare applicant personality, interest, value, or organizational culture preference information to the characteristics of the job or organization. The concept behind job-fit instruments is individuals are attracted to, and seek employment with, organizations which exhibit characteristics similar to their own.

The most common organizational characteristic used in job-fit measures is the organizational culture (e.g., innovative, detail oriented, team oriented). Although job-fit can be measured with interviews or other instruments, job-fit instruments are typically administered to applicants in the form of self-report questionnaires or surveys. Technological advancements of the Internet have made it easier to administer job-fit measures on-line, or as a possible feature to an agency or company's website. An example item from a job-fit measure is: "I prefer a work environment which doesn't demand constant adaptation" (1 = strongly disagree, 5 = strongly agree).

Based on their responses to the job-fit items, applicants are often offered tailored feedback regarding their likely fit with the job or organization. Moreover, those who perceive or receive feedback which indicates they are not a good fit with the job or organization are more likely to voluntarily withdraw from the application process. For this reason, job-fit measures that give applicants the opportunity to self-select out are typically administered before all traditional assessments (e.g., cognitive ability tests, accomplishment record).

Job-fit measures can also be used as a screen-out tool (such as traditional assessments); however, the research (e.g., validity, methodology, utility) regarding the use of job-fit measures in this regard is still in its infancy.

References: (See Section VI for a summary of each article)

Arthur, W., Jr., Bell, S. T., Villado, A. J., & Doverspike, D. (2006). The use of person-organization fit in employment decision making: An assessment of its criterion-related validity. *Journal of Applied Psychology, 91,* 786-801.

Cable, D. M., & Judge, T. A. (1997). Interviewers' perceptions of person-organization fit and organizational selection decisions. *Journal of Applied Psychology, 82,* 546-561.

Dineen, B. R., Ash, S. R., & Raymond, N. A. (2002). A web of applicant attraction: Person-organization fit in the context of web-based recruitment. *Journal of Applied Psychology*, *87*(4), 723-734.

Judge, T. A., & Cable, D. M. (1997). Applicant personality, organizational culture, and organizational attraction. *Personnel Psychology, 50,* 359-394.

Kristof, A. L. (1996). Person-organization fit: An integrative review of its conceptualizations, measurement, and implications. *Personnel Psychology, 49,* 1-49.

Martinez, M. N. (2000). Get job seekers to come to you. *HR Magazine, 45*, 45-52.

Physical Ability Tests: Physical ability tests typically ask individuals to perform job-related tasks requiring manual labor or physical skill. These tasks measure physical abilities such as strength, muscular flexibility, and stamina. Examples of physical ability tests include:

- Muscular Tension Tests - Tasks requiring pushing, pulling, lifting
- Muscular Power Tests - Tasks requiring the individual to overcome some initial resistance (e.g., loosening a nut on a bolt)
- Muscular Endurance Tests - Tasks involving repetitions of tool use (e.g., removing objects from belts)
- Cardiovascular Endurance Tests - Tasks assessing aerobic capacity (e.g., climbing stairs)
- Flexibility Tests - Tasks where bending, twisting, stretching or reaching of a body segment occurs (e.g., installing lighting fixtures)
- Balance Tests - Tasks in which stability of body position is difficult to maintain (e.g., standing on rungs of a ladder)

While some physical ability tests may require electronically monitored machines, equipment needs can often be kept simple. For example, stamina can be measured with a treadmill and an electrocardiograph, or with a simple set of steps. However, a possible drawback of using simpler methods is less precise measurement.

Many factors must be taken into consideration when using physical ability tests. First, employment selection based on physical abilities can be litigious. Legal challenges have arisen over the years because physical ability tests, especially those involving strength and endurance, tend to screen out a disproportionate number of women and some ethnic minorities. Therefore, it is crucial to have validity evidence justifying the job-relatedness of physical ability measures. Second, physical ability tests involving the monitoring of heart rate, blood pressure, or other physiological factors are considered medical exams under the Americans with Disabilities Act. Administering medical exams to job applicants prior to making a job offer is expressly prohibited. Finally, there is the concern of candidates injuring themselves while performing a physical ability test (e.g., a test involving heavy lifting may result in a back injury or aggravate an existing medical condition).

References: (See Section VI for a summary of each article)

Arvey, R. D., Maxwell, S. E., & Salas, E. (1992). Development of physical ability tests for police officers: A construct validation approach. *Journal of Applied Psychology, 77,* 996-1009.

Arvey, R. D., Nutting, S. M., & Landon, T. E. (1992). Validation strategies for physical ability testing in police and fire settings. *Public Personnel Management, 21,* 301-312.

Campbell, W. J., & Fox, H. R. (2002). Testing individuals with disabilities in the employment context: An overview of issues and practices. In R. B. Ekstrom & D. K. Smith (Eds.) *Assessing Individuals with Disabilities in Educational, Employment, and Counseling Settings* (1st ed, p. 198). Washington, DC: American Psychological Association.

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

Hogan, J. (1991). The structure of physical performance in occupational tasks. *Journal of Applied Psychology, 76,* 495-507.

Realistic Job Previews: A Realistic Job Preview (RJP) is a recruiting tool used to communicate both the good and bad aspects of a job. Essentially, it is used to provide a prospective employee a realistic view of what the job entails. This measure, much like the job-fit measure, is to provide candidates a richer description of the agency and the job (e.g., work environment, duties, expectations) to help them decide if they are a good match. While the RJP can be useful for reducing turnover, it should be used as a candidate self-evaluation tool rather than a traditional selection device (e.g., cognitive ability test, accomplishment record).

In creating a RJP, there are many factors to consider, including:

- How the RJP will be created (e.g., structured observations, meetings with current employees)
- How the RJP will be distributed (e.g., written material, video, interview)
- How to present both positive and negative aspects of the job (e.g., always follow a negative item with a positive item)
- When to introduce the RJP (i.e., early or late in the recruiting process)

References: (See Section VI for a summary of each article)

McEvoy, G. M., & Cascio, W. F. (1985). Strategies for reducing employee turnover: A meta-analysis. *Journal of Applied Psychology, 70*(2), 342-353.

Pitt, L. F., & Ramaseshan, B. (1995). Realistic job information and salesforce turnover: An investigative study. *Journal of Managerial Psychology, 10*(5), 29-36.

Saks, A. M, Wiesner, W. H., & Summers, R. (1996). Effects of job previews and compensation policy on applicant attraction and job choice. *Journal of Vocational Behavior*, *49,* 68-85.

Wanous, J. P. (1989). Installing a realistic job preview: Ten tough choices. *Personnel Psychology, 42*(1), 117-133.

Section V: Glossary

Adverse impact	A substantially different rate of selection in hiring which works to the disadvantage of members of any race, sex, or ethnic group. *Back*
Behavioral consistency method	Based on the principle that past behavior is the best predictor of future behavior. In practice, the method involves describing previous accomplishments gained through work, training, or other experience (e.g., school, community service, hobbies) and matching those accomplishments to the competencies required by the job. *Back*
Concurrent validity	In a *concurrent* study, job incumbents (i.e., current employees) are tested and their job performance is evaluated at the same time. The relation between current performance on the assessment and on the job can then be examined. Evidence of concurrent validity is often substituted for predictive validity. Whether this is appropriate will depend on the type of measure and how similar the incumbent sample is to the applicant population. *Back*
Construct validity	A construct refers to the underlying trait (e.g., intelligence, sociability) assumed to be measured by an assessment. Construct validation involves collecting evidence to determine whether the assessment does indeed measure the trait it was intended to measure. *Back*
Content validity	Evidence (based on job analysis and expert judgment) the choice of items or tasks included in the assessment logically match or represent those tasks or competencies required by the job. *Back*
Criterion-related validity	The degree to which performance on an assessment procedure predicts (or is statistically related to) an important criterion such as job performance, training success, or productivity. There are two major types of criterion-related validity, concurrent and predictive. *Back*

Face validity	An applicant's perception of how valid a measure is based on simple visual inspection. Though face validity alone cannot be used to support the use of an assessment, it is important because it promotes cooperation and acceptance of the assessment process on the part of applicants. *Back*
Incremental validity	The extent to which a new assessment adds to the prediction of job success above and beyond the forecasting powers of an existing assessment. *Back*
Job analysis	A systematic examination of the tasks performed in a job and the competencies required to perform them. *Back*
Predictive validity	In a *predictive* study, job applicants are tested and their performance evaluated at a later time, usually after being on the job for 6 months or more. The relation between performance on the assessment and on the job can then be examined. *Back*
Validity	The extent to which assessment scores are related to current or future job performance (or some other work-related outcome such as training success, productivity, absenteeism, turnover). For types of validity evidence, see content validity, construct validity, and criterion-related validity. *Back*

Section VI: Annotated References

[Accomplishment Records](#) References:

Hough, L. M. (1984). Development and evaluation of the "accomplishment record" method of selecting and promoting professionals. *Journal of Applied Psychology, 69*(1), 135-146.
Developed the accomplishment record method, which applies the principle of behavioral consistency to selection and promotion and results in self-reported descriptions of accomplishments in highly relevant, behavioral job dimensions. These protocols were reliably rated (.82) using specially prepared rating scales and guidelines. Scores on the accomplishment record inventory were unrelated to traditional psychological measures, such as aptitude tests, grades, and honors, but were correlated with job performance. Correlations with self-perceptions of success, hard work, and self-assurance, and with length of time spent practicing a profession are also provided.

Hough, L. M., Keyes, M. A., & Dunnette, M. D. (1983). An evaluation of three "alternative" selection procedures. *Personnel Psychology, 36*(2), 261-276.
A content-oriented strategy was used to develop three alternative selection inventories designed to reflect the content domain of positions held by attorneys employed with a large Federal agency. These inventories and three traditional inventories were completed by 329 attorneys of the agency as part of a concurrent validation study. Criterion-related validities of an accomplishment record inventory with a background inventory and an interest and opinion inventory were both statistically and practically significant. The special features and advantages of the accomplishment record inventory as an alternative selection procedure are discussed.

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*(4), 302-318.
The authors describe the nature of the issues faced by practitioners when trying to optimize both the performance and ethnic diversity of chosen individuals. A review of research on different strategies to address these issues (e.g., adverse impact) is provided. The authors recommend using selection materials assessing the full range of relevant attributes using a format minimizing 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. The authors conclude that regardless of strategy adopted, it is difficult to maximize both the performance and ethnic diversity of selected individuals.

Schmidt, F. L., Caplan, J. R., Bemis, S. E., Decuir, R., Dunn, L., & Antone, L. (1979). *The behavioral consistency method of unassembled examining.* **Washington, DC: U.S. Office of Personnel Management, Personnel Resources and Development Center.**
This report describes the history, rationale, and development of the behavioral consistency procedure, a methodology very similar to the accomplishment record in format and is based on an applicant's past achievements rather than on credentials. It also describes the results of a study conducted to compare the behavioral consistency method to two traditional rating tools. The results indicated the behavioral consistency method (1) shows greater interrater reliability

than the other two methods, (2) is feasible in terms of cost and time requirements, and (3) measures factors different than those measured by the traditional rating procedures.

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*(2), 262-274.
This article summarizes the practical and theoretical implications of 85 years of research in personnel selection. On the basis of meta-analytic findings, this article presents the validity of 19 selection procedures for predicting job performance and training performance and the validity of paired combinations of cognitive ability tests with the 18 other selection procedures. The practical utility of the implications of this article's summary findings are substantial. In addition, the implications of these research findings for the development of theories of job performance are discussed.

[Assessment Center](#) **References:**

Arthur, W. Jr., Day, E. A., McNelly, T. L., & Edens, P. S. (2003). A meta-analysis of the criterion-related validity of assessment center dimensions. *Personnel Psychology, 56*, 125-154.
Used meta-analytic procedures to investigate the criterion-related validity of assessment center dimension ratings. By focusing on dimension-level information, the authors were able to assess the extent to which specific constructs account for the criterion-related validity of assessment centers. From a total of 34 articles reporting dimension-level validities, the authors collapsed 168 assessment center dimension labels into an overriding set of 6 dimensions: (1) consideration/awareness of others, (2) communication, (3) drive, (4) influencing others, (5) organizing and planning, and (6) problem solving. Results show a range of estimated criterion-related validities from .25 to .39.

Caldwell, C., Thornton, G. C., & Gruys, M. (2003). Ten classic assessment center errors: Challenges to selection validity. *Public Personnel Management, 32*(1), 73-88.
This paper summarizes ten classic errors associated with selection and promotion related assessment center administration. Critical errors covered are: (1) poor planning, (2) inadequate job analysis, (3) weakly defined dimensions, (4) poor exercises, (5) no pre-test evaluations, (6) unqualified assessors, (7) inadequate assessor training, (8) inadequate candidate preparation, (9) sloppy behavior documentation and scoring, and (10) misuse of results. Reducing and/or eliminating the errors in this list will allow more efficient and effective assessment center administration.

Gaugler, B. B., Rosenthal, D. B., Thornton, G. C., & Bentson, C. (1987). Meta-analysis of assessment center validity. *Journal of Applied Psychology, 72*(3), 493-511.
A meta-analysis of 50 assessment center studies revealed higher validities were found in studies in which potential ratings were the criterion, and lower validities were found in promotion studies. Validities were higher when the percentage of female assessees was high, when several evaluation devices were used, when assessors were psychologists rather than managers, when peer evaluation was used, and when the study was methodologically sound. Age of assessees, whether feedback was given, days of assessor training, days of observation, percentages of

minority assessees, and criterion contamination did not moderate assessment center validities. The findings suggest assessment centers show both validity generalization and situational specificity.

Testing and Assessment: An Employer's Guide to Good Practices. (2000). Washington, DC: U.S. Department of Labor, Employment and Training Administration. Note: Article can be accessed at http://www.onetcenter.org/guides.html.
This guide is written as a source of guidance to managers and Human Resource (HR) professionals. It provides basic yet essential concepts on employment testing and assessment to help managers and HR professionals evaluate and select the most appropriate assessment tool for a specific situation, administer and score assessment tools, interpret assessment results, and understand professional and legal standards that must be followed when conducting personnel assessments. An overview of the development and administration of assessment centers is provided.

Woehr, D., & Winfred, A. (2003). The construct-related validity of assessment center ratings: A review and meta-analysis of the role of methodological factors. *Journal of Management, 29*(2), 231-258.
The present study provides a systematic review of the assessment center literature with respect to specific design and methodological characteristics that potentially moderate the validity of assessment center ratings. In addition, the results of a meta-analysis of the relationship between these characteristics and construct-related validity outcomes are presented. These results provide little if any support for the view assessment center ratings lack construct-related validity, while at the same time demonstrating criterion-related validity. The implications of these findings for assessment center construct-related validity are discussed.

Zedeck, S. (1986). A process analysis of the assessment center method. In B. M. Staw & L. L. Cummings (Eds.), *Research in organizational behavior, 8,* 259-296.
Discusses the dynamics operating in observing the behavior of managerial candidates in simulated exercises and in processing information for the evaluation of candidates. These dynamics are viewed from 3 perspectives: (1) information processing, (2) categorization and social cognition, and (3) group dynamics. Concepts such as categories and management behavior schema are used to explain how assessors recall information and make predictions and judgments.

Biographical Data (Biodata) Tests References:

Elkins, T., & Phillips, J. (2000). Job context, selection decision outcome, and the perceived fairness of selection tests: Biodata as an illustrative case. *Journal of Applied Psychology, 85*(3), 479-484.
The present study aimed to verify and extend Gilliland's (1993) proposed model of perceived selection fairness by investigating the role of job context in the formation of fairness perceptions of biodata. A sample of 255 students completed an operational biodata instrument, believing it would be used to hire persons for either international, local, or unspecified entry-level managerial positions. Participants were then presented with outcome information (selected or rejected for further consideration). Consistent support was found for the research hypotheses

derived from the Gilliland model. Participants' perceptions of the fairness and job relatedness of biodata were affected by the selection context and decision outcome.

Hough, L. M., & Oswald, F. L. (2000). Personnel selection: Looking toward the future— Remembering the past. *Annual Review of Psychology, 51*, 631-664.
Reviews personnel selection research from 1995-1999. Areas covered are job analysis; performance criteria; cognitive ability and personality predictors; interview, assessment center, and biodata assessment methods; measurement issues; meta-analysis and validity generalization; evaluation of selection systems in terms of differential prediction, adverse impact, utility, and applicant reactions; emerging topics on team selection and cross-cultural issues; and finally professional, legal, and ethical standards. Three major themes are revealed: (1) better taxonomies produce better selection decisions; (2) the nature and analyses of work behavior are changing, influencing personnel selection practices; (3) the field of personality research is healthy, as new measurement methods, personality constructs, and compound constructs of well-known traits are being researched and applied to personnel selection.

Mount, M. K., Witt, L. A., & Barrick, M. R. (2000). Incremental validity of empirically keyed biodata scales over GMA and the five factor personality constructs. *Personnel Psychology, 53*(2), 299-323.
Sought to determine whether empirically keyed, cross-validated biodata scales account for incremental variance over that accounted for by the five factor model (FFM) of personality and cognitive ability predictors. A concurrent validation study was employed using 376 employees in a clerical job (222 in the developmental sample and 146 in the cross-validation sample). Building on the F. A. Mael and A. C. Hirsch (1993) and M. A. McManus and M. L. Kelly (1999) studies, the authors examined the joint use of cognitive ability, biodata, and personality as predictors of four different criteria: quantity and quality of work, problem solving, interpersonal facilitation, and retention probability. Results for the cross-validation sample provided support for the hypothesis that biodata predictors accounted for substantial incremental variance beyond that accounted for by the FFM predictors and cognitive ability for three of the four criteria. Support was also found for the hypothesized zero-order correlations between cognitive ability, FFM, and biodata predictors and the four criteria.

Rothstein, H. R., Schmidt, F. L., Erwin, F. W., Owens, W. A., & Sparks, C. P. (1990). Biographical data in employment selection: Can validities be made generalizable? *Journal of Applied Psychology, 75*(2), 175-184.
The hypothesis was examined that organizational specificity of biodata validity results from the methods typically used to select and key items. In this study, items were initially screened for job relevance, keying was based on large samples from multiple organizations, and items were retained only if they showed validity across organizations. Cross-validation was performed on approximately 11,000 first-line supervisors in 79 organizations. The resulting validities were meta-analyzed across organizations, age levels, sex, and levels of education, supervisory experience, and company tenure. In all cases, validities were generalizable. Validities were also stable across time and did not appear to stem from measurement of knowledge, skills, or abilities acquired through job experience. Finally, these results provide additional evidence against the hypothesis of situational specificity of validities, the first large-sample evidence in a noncognitive domain.

Schmitt, N., Cortina, J. M., Ingerick, M. J., & Wiechmann, D. (2003). Personnel selection and employee performance. *Handbook of Psychology: Industrial and Organizational Psychology, 12*, 77-105. New York, NY: John Wiley & Sons, Inc.
The authors of this chapter suggest personnel selection research has clearly expanded from its early interest in documenting predictor-criterion relationships. They discuss progress made in considering a broader range of predictors, testing more sophisticated performance models, and in paying more attention to the social significance of personnel selection and the reactions of examinees. Biodata measures are discussed as part of this new trend in personnel selection and employee performance.

[Cognitive Ability Tests](#) References:

Hunter, J. E. (1986). Cognitive ability, cognitive aptitude, job knowledge, and job performance. *Journal of Vocational Behavior, 29*(3), 340-362.
A research review indicates general cognitive ability (GCA) predicts supervisor ratings and training success as well as objective, rigorously content-valid work sample performance. Analyses carried out in several previous studies by the present author showed much of this predictive power stemmed from the fact GCA predicted job knowledge and job knowledge predicted job performance. However, GCA predicted performance to a greater extent, verifying job analyses showing most major cognitive skills are used in everyday work. Evidence showing GCA and not specific cognitive aptitudes predict performance is discussed. Findings support classic learning theory over behaviorist theories of learning and performance.

Murphy, K. R., Cronin, B. E., & Tam, A. P. (2003). Controversy and consensus regarding the use of cognitive ability testing in organizations. *Journal of Applied Psychology, 88*(4), 660-671.
Over 700 members of the Society for Industrial and Organizational Psychology indicated agreement or disagreement with 49 propositions regarding cognitive ability tests in organizations. There was consensus that cognitive ability tests are valid and fair, they provide good but incomplete measures, different abilities are necessary for different jobs, and diversity is valuable. Items dealing with the unique status of cognitive ability were most likely to generate polarized opinions. The data represented two factors: (1) societal concerns over the consequences of ability testing and (2) emphasis on the unique status of cognitive ability.

Outtz, J. L. (2002). The role of cognitive ability tests in employment selection. *Human Performance, 15*(1-2), 161-172.
Cognitive ability tests correlate with measures of job performance across many jobs. However, cognitive ability tests produce racial differences three to five times larger than other predictors-- such as biodata, personality inventories, and the structured interview--that are valid predictors of job performance. Given (a) cognitive ability tests can be combined with other predictors such that adverse impact is reduced while overall validity is increased, and (b) alternative predictors with less adverse impact can produce validity coefficients comparable to those obtained with cognitive ability tests alone, sole reliance on cognitive ability tests when alternatives are available is unwarranted.

Ree, M. J., Earles, J. A., & Teachout, M. S. (1994). Predicting job performance: Not much more than g. *Journal of Applied Psychology, 79*(4), 518-524.

The roles of general cognitive ability and specific abilities or knowledge were investigated as predictors of work sample job performance criteria in seven jobs for U.S. Air Force enlistees. The interaction of general cognitive ability and specific experience were defined by scores on the first and subsequent principal components of the enlistment selection and classification test (the Armed Services Vocational Aptitude Battery). Analyses revealed cognitive ability was the best predictor of all criteria and specific abilities or knowledge added a statistically significant but practically small amount to predictive efficiency. These results are consistent with those of previous studies, most notably Army Project A. The study also extends the findings to other jobs and uses traditionally more acceptable estimates of cognitive ability.

Schmidt, F. L., & Hunter, J. (2004). General mental ability in the world of work: Occupational attainment and job performance. *Journal of Personality & Social Psychology, 86*(1), 162-173.

The psychological construct of general mental ability (GMA), otherwise known as cognitive ability, introduced by C. Spearman (1904) nearly 100 years ago, has enjoyed a resurgence of interest and attention in recent decades. This article presents research evidence GMA predicts both occupational level attained and performance within one's chosen occupation and does so better than any other ability, trait, or disposition, and better than job experience. The sizes of these relationships with GMA are also larger than most found in psychological research. Evidence is presented that weighted combinations of specific aptitudes tailored to individual jobs do not predict job performance better than GMA alone, disconfirming specific aptitude theory. A theory of job performance is described explaining the central role of GMA in the world of work.

[Emotional Intelligence Tests](#) **References:**

Brackett, M. A., Rivers, S. E., Shiffman, S., Lerner, N., & Salovey, P. (2006). Relating emotional abilities to social functioning: A comparison of self-report and performance measures of emotional intelligence. *Journal of Personality and Social Psychology*, *91*, 780-795.

Two distinct approaches to measuring emotional intelligence (EI) have emerged: ability-based (performance) tests and self-report questionnaires. The purpose of this study was to explore the relationship between these two different measures of EI and whether one approach is better than the other at predicting an important outcome measure (namely, social competence). The findings indicated ability-based measures and self-rated EI were not strongly related, suggesting a self-assessment of one's own EI may not constitute an accurate measure of EI. These findings were compared to other research showing individuals are also bad judges of their own mental abilities. In a follow-up study, the researchers examined the extent to which self-rated and ability-based measures of EI were able to predict a key outcome variable, social competence. The results showed women score higher than men on ability-based measures of EI and self-rated measures came out about the same for the two groups. The self-report measures of EI did not predict social competence for either men or women. The ability-based measure of EI predicted social competence scores for men but not women. Possible explanations for the gender differences were considered.

Frost, D. E. (2004). The psychological assessment of emotional intelligence. In J. C. Thomas & M. Hersen (Eds.), *Comprehensive handbook of psychological assessment, Volume 4: Industrial and organizational psychology* **(pp. 203-215). Hoboken, NJ: John Wiley & Sons.**
The author first describes the growing public and scientific interest in the concept of emotional intelligence (EI) and the historical and social context behind EI's recent emergence. Debates concerning the proper definition of EI are considered and whether the concept is little more than a renaming of personality variables as aspects of traditional intelligence (i.e., IQ). The author concludes most researchers and practitioners see the psychological construct of EI as being made up of distinct mental abilities or capacities. The rest of the chapter describes attempts to measure EI and its many workplace applications (e.g., selection, placement, and training). The author also considers what types of occupations would be most suitable for EI assessment.

Mayer, J. D., Salovey, P., & Caruso, D. R. (2004). Emotional intelligence: Theory, findings, and implications. *Psychological Inquiry, 15***, 197-215.**
This article provides a detailed and comprehensive review of findings regarding emotional intelligence (EI) measured as an ability. The first major conclusion is EI is distinct from other commonly measured variables related to personality and intelligence. Evidence for this is based on studies finding very low correlations between EI and other types of intelligence. Also, EI has rather low correlations with measures of social and emotional traits. A growing body of evidence indicates EI, measured as an ability, predicts a variety of important outcomes. For example, studies have shown people higher in EI are more likely to avoid drug problems, interpersonal arguments, and violence. Such individuals are also more satisfied with their social networks and appear to receive more social support. These predictive relationships are at levels typically found for similarly complex concepts (e.g., personality variables). The authors argue that because high EI involves the more successful resolution of personal conflict and lower levels of aggression, it is a highly desirable, and often personally valuable, attribute to possess.

Salovey, P., & Mayer, J. D. (1990). Emotional intelligence. *Imagination, Cognition, and Personality***,** *9***, 185-211.**
Introduces the original framework for the emotional intelligence concept, a set of skills hypothesized to contribute to the accurate appraisal and expression of emotion, the effective regulation of emotion, and the use of feelings to motivate, plan, and achieve. Adaptive versus maladaptive qualities of emotion are discussed. The literature on intelligence, and especially social intelligence, is reviewed to examine the place of emotion in traditional intelligence conceptions. The article also provides a framework for integrating the diverse research on emotion-related skills.

Integrity/Honesty Tests References:

Cullen, M. J., & Sackett, P. R. (2004). Integrity testing in the workplace. In J. C. Thomas & M. Hersen (Eds.), *Comprehensive handbook of psychological assessment, Volume 4: Industrial and organizational psychology* **(pp. 149-165). Hoboken, NJ: John Wiley & Sons.**
This chapter provides a comprehensive summary of integrity testing research and practice. Topic areas include a review of validity evidence for personality-oriented and overt integrity tests, the relation between integrity tests and other assessments (e.g., cognitive ability tests,

standard personality factors), the effect of faking on integrity test scores, applicant privacy right issues, and subgroup score differences by race and gender. The chapter's central conclusions are integrity tests: (1) show consistent evidence of validity for predicting counterproductive behaviors, as well as overall job performance, (2) do not show subgroup differences by gender, race, or ethnicity, (3) have some overlap with standard personality factors (i.e., the Big Five) but also appear to measure other unique factors, and (4) are subject to faking by applicants but response distortion (i.e., providing socially desirable answers) does not seem to undermine the usefulness of the tests.

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*, *78*, 679-703.
The authors conducted a comprehensive meta-analysis that quantitatively summarized the results of multiple validity studies. Generally, the researchers found integrity tests were useful predictors of overall job performance and counterproductive behaviors on the job, such as theft, disciplinary problems, and absenteeism. Both personality-based and overt (i.e., clear purpose) integrity tests proved to be valid predictors of job performance. Results from predictive validity studies conducted on applicants using external performance measures (i.e., excluding self-reports of counterproductive behavior) indicate integrity tests predict an overall measure of counterproductivity (a combination of violence on the job, absenteeism, tardiness, and other disruptive behaviors) better than they predict employee theft alone.

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*(4), 787-829.
This article provides a fairly extensive summary of issues related to the use of integrity tests for personnel selection. The authors conclude personality-based and overt (clear purpose) integrity tests are valid predictors of counterproductive behaviors (admitted theft, dismissals based on theft, various illegal activities, tardiness, absenteeism, and on-the-job violence). Both types of tests also predict measures of overall job proficiency for a wide variety of jobs. Limiting the predictive analysis to only applicants (as opposed to current employees) and actual detection of theft lowers the validity substantially. Integrity test scores are related to some of the Big Five measures of personality, particularly conscientiousness, agreeableness, and emotional stability. Integrity tests do not demonstrate adverse impact against women or minorities and are not generally perceived as negative by applicants. Integrity tests are relatively unrelated to cognitive ability (i.e., intelligence), implying they can be used in conjunction with cognitive ability measures to enhance selection quality while also reducing the adverse impact associated with cognitive ability tests.

Job Knowledge Tests References:

Dubois, D., Shalin, V. L., Levi, K. R., & Borman, W. C. (1993). Job knowledge test design: A cognitively-oriented approach. *U.S. Office of Naval Research Report*, *Institute Report 241*, i-47.
This study applied cognitive methods to the measurement of performance using tests of job knowledge. The research goal was to improve the usefulness of job knowledge tests as a proxy

for hands-on performance. The land navigation skills of 358 Marines were tested with a written job knowledge test consisting of multiple-choice questions, hands-on proficiency tests, and a work-sample performance test. Results indicate cognitively-oriented job knowledge tests show improved correspondence with hands-on measures of performance, compared with existing content-oriented test development procedures.

Dye, D. A., Reck, M., & McDaniel, M. A. (1993). The validity of job knowledge measures. *International Journal of Selection and Assessment, 1,* **153-157.**
The results of this study demonstrated the validity of job knowledge tests for many jobs. Job knowledge was defined as "the cumulation of facts, principles, concepts, and other pieces of information considered important in the performance of one's job" (p. 153). In their meta-analysis of 502 validity coefficients based on 363,528 individuals, they found high levels of validity for predicting training and job performance.

Ree, M. J., Carretta, T. R., & Teachout, M. S. (1995). Role of ability and prior job knowledge in complex training performance. *Journal of Applied Psychology, 80*(6), 721-730.
A causal model of the role of general cognitive ability and prior job knowledge in subsequent job knowledge acquisition and work sample performance during training was developed. Participants were 3,428 U.S. Air Force officers in pilot training. The measures of ability and prior job knowledge came from the Air Force Officer Qualifying Test. The measures of job knowledge acquired during training were derived from classroom grades. Work sample measures came from check flight ratings. The model showed ability directly influenced the acquisition of job knowledge. General cognitive ability influenced work samples through job knowledge. Prior job knowledge had almost no influence on subsequent job knowledge but directly influenced the early work sample. Early training job knowledge influenced subsequent job knowledge and work sample performance. Finally, early work sample performance strongly influenced subsequent work sample performance.

Roth, P. L., Huffcutt, A. I., & Bobko, P. (2003). Ethnic group differences in measures of job performance: A new meta-analysis. *Journal of Applied Psychology, 88*(4), 694-706.
The authors conducted a meta-analysis of ethnic group differences in job performance. Analyses of Black-White differences within categories of job performance were conducted and subgroup differences within objective and subjective measurements were compared. Contrary to one perspective sometimes adopted in the field, objective measures are associated with very similar, if not somewhat larger, standardized ethnic group differences than subjective measures across a variety of indicators. This trend was consistent across quality, quantity, and absenteeism measures. Further, work samples and job knowledge tests are associated with larger ethnic group differences than performance ratings or measures of absenteeism. Analysis of Hispanic-White standardized differences shows they are generally lower than Black-White differences in several categories.

Sapitula, L., & Shartzer, M. C. (2001). Predicting the job performance of maintenance workers using a job knowledge test and a mechanical aptitude test. *Applied H.R.M. Research, 6*(1-2), 71-74.
This study examined the predictive validity of the Job Knowledge Written Test (JKWT) and the Wiesen Test of Mechanical Aptitude (WTMA, J. P. Wiesen, 1997), and the effects of race,

gender, and age on scores. A total of 782 applicants completed the JKWT and the WTMA, and 102 maintenance workers were administered the JKWT, the WTMA, and a job performance appraisal. Results show no significant relationship between job performance ratings and either the JKWT or WTMA. Male applicants scored higher than did female applicants and White applicants scored higher than did minority applicants.

[Personality Tests](#) References:

Barrick, M. R., & Mount, M. K. (1991). The Big Five personality dimensions and job performance: A meta-analysis. *Personnel Psychology, 44*, 1-26.
Investigated the relation of the "Big Five" personality dimensions to three job performance criteria (job proficiency, training proficiency, and personnel data) for five occupational groups (professionals, police, managers, sales, and skilled/semi-skilled). A review of 117 studies yielded 162 samples totaling 23,994 subjects. Conscientiousness showed consistent relations with all job performance criteria for all occupational groups. Extraversion was a valid predictor for two occupations involving social interaction (managers and sales). Also, openness to experience and extraversion were valid predictors of the training proficiency criterion across occupations. Overall, results illustrate the benefits of using the five-factor model of personality to accumulate empirical findings. Study results have implications for research and practice in personnel psychology.

Hogan, R., Hogan, J., & Roberts, B. W. (1996). Personality measurement and employment decisions: Questions and answers. *American Psychologist, 51*, 469-477.
Summarizes information needed to answer the most frequent questions about the use of personality measures in applied contexts. Conclusions are (1) well-constructed measures of normal personality are valid predictors of performance in virtually all occupations, (2) they do not result in adverse impact for job applicants from minority groups, and (3) using well-developed personality measures for pre-employment screening is a way to promote social justice and increase organizational productivity.

Hough, L. M., Eaton, N. K., Dunnette, M. D., Kamp, J. D., & McCloy, R. A. (1990). Criterion-related validities of personality constructs and the effect of response distortion on those validities. *Journal of Applied Psychology, 75*, 581-595.
An inventory of six personality constructs and four response validity scales measuring accuracy of self-description were administered in three contexts: a concurrent criterion-related validity study, a faking experiment, and an applicant setting. Results showed (a) validities were in the .20s against targeted criterion constructs, (b) respondents successfully distorted their self-descriptions when instructed to do so, (c) response validity scales were responsive to different types of distortion, (d) applicants responses did not reflect evidence of distortion, and (e) validities remained stable regardless of possible distortion.

Hough, L. M., & Oswald, F. L. (2000). Personnel selection: Looking toward the future—Remembering the past. *Annual Review of Psychology, 51*, 631-664.
Reviews personnel selection research from 1995-1999. Areas covered are job analysis; performance criteria; cognitive ability and personality predictors; interview, assessment center, and biodata assessment methods; measurement issues; meta-analysis and validity generalization;

evaluation of selection systems in terms of differential prediction, adverse impact, utility, and applicant reactions; emerging topics on team selection and cross-cultural issues; and finally professional, legal, and ethical standards. Three major themes are revealed: (1) better taxonomies produce better selection decisions; (2) the nature and analyses of work behavior are changing, influencing personnel selection practices; (3) the field of personality research is healthy, as new measurement methods, personality constructs, and compound constructs of well-known traits are being researched and applied to personnel selection.

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.
Based on 97 independent samples, a meta-analysis was used to (a) assess overall validity of personality measures as predictors of job performance, (b) investigate moderating effects of several study characteristics on personality scale validity, and (c) appraise predictability of job performance as a function of eight categories of personality content. Results indicated studies using confirmatory research strategies produced corrected mean personality scale validity more than twice as high as studies adopting exploratory strategies. An even higher mean validity was obtained based on studies using job analysis explicitly in selection of personality measures.

Reference Checking References:

Aamodt, M. G. (2006). Validity of recommendations and references. *Assessment Council News*, February, 4-6.
Reference data are subject to inflation and low reliability and generally reach only moderate levels of predictive validity. Even so, organizations are encouraged to check the references of their applicants because of widespread resume fraud and potential liability in the form of negligent hiring.

Taylor, P. J., Pajo, K., Cheung, G. W., & Stringfield, P. (2004). Dimensionality and validity of a structured telephone reference check procedure. *Personnel Psychology, 57*, 745-772.
Reports that reference checking, when properly structured, can prevent defamation litigation and add significant value to the selection process. Specifically tests the hypothesis that utilizing a structured, competency-based approach to reference checking can increase the predictive validity of ratings in much the same way as structuring the employment interview process. A structured job analysis was used to identify the core job-related competencies deemed essential to effective performance in a family of customer-contact jobs within a 10,000-employee service organization. These competencies (Commitment, Teamwork, and Customer Service) were incorporated into a structured reference form and contacts were asked to rate applicants on a number of behavioral indicators within each competency. A structured telephone interview with contacts was then used to obtain evidence of actual occurrences to support the ratings. Results indicated using a structured telephone reference check increased the employer's ability to predict future job performance. Results also indicated a shorter contact-applicant relationship does not undermine predictions of future job performance.

U.S. Merit Systems Protection Board. (2005). Reference checking in federal hiring: Making the call. Washington, DC: Author. Note: Report available at:
http://www.mspb.gov/netsearch/viewdocs.aspx?docnumber=224106&version=224325&application=ACROBAT
Hiring officials should check references. The quality of reference checking can be improved by insisting job applicants provide at least three references who have observed their performance on the job. Supervisors should discuss the performance of their current and former employees with prospective employers. Some former supervisors will only provide basic facts about work histories (e.g., employment dates and positions held) because they are concerned with protecting the privacy of former employees. Their concern is understandable but need not interfere with reference checking. So long as reference checking discussions focus on job-related issues such as performance, reference giving is appropriate and legally defensible. Former supervisors who support reference checking inquiries can reward good employees for their past contributions and avoid "passing on" a problem employee to another agency. Agency human resources personnel can work to remove barriers to effective reference checking. For example, applicants should be required to complete Declaration of Federal Employment (OF-306) forms early in the application process. This form explicitly grants permission to check references. And this sets applicants' expectations appropriately—their performance in previous employment will be investigated.

Situational Judgment Tests References:

Hanson, M. A., Horgen, K. E., & Borman W. C. (1998, April). Situational judgment tests (SJT) as measures of knowledge/expertise. Paper presented as the 13th Annual Conference of the Society for Industrial and Organizational Psychology, Dallas, TX.
This paper discusses the situational judgment test (SJT) methodology and reasons for its popularity. This paper also investigates the nature of the construct(s) measured by these tests, why they are valid, when they are valid, and why they are sometimes not valid. The authors propose the SJT methodology is best suited for measuring knowledge or expertise, and discusses available construct validity evidence consistent with this perspective. This perspective generates several testable hypotheses, and additional research is proposed. Finally, the implications of this perspective for the development of valid and useful SJTs are discussed.

McDaniel, M. A., Morgeson, F. P, Finnegan, E. B, Campion, M. A., & Braverman, E. P. (2001). Use of situational judgment tests to predict job performance: A clarification of the literature. *Journal of Applied Psychology, 86,* **730-740.**
This article reviews the history of situational judgment tests (SJT) and presents the results of a meta-analysis on criterion-related and construct validity. SJTs showed useful levels of validity across all jobs and situations studied. The authors also found a relatively strong relationship between SJTs and cognitive ability and the relationship depended on how the test had been developed. On the basis of the literature review and meta-analytic findings, implications for the continued use of SJTs are discussed, particularly in terms of recent investigations into tacit knowledge.

McDaniel, M. A., Whetzel, D. L., & Nguyen, N. T. (2006). *Situational judgment tests for personnel selection.* **Alexandria, VA: IPMA Assessment Council.**

Employers should take into account several factors before choosing to develop their own in-house situational judgment tests (SJTs). For example, SJT developers must make a number of decisions about the content of items, response options, response instructions, and answer key. This monograph also describes the major steps in building a situational judgment test such as conducting a critical incident workshop, creating item stems from critical incidents, generating item responses, developing item response instructions, and choosing among several scoring key methods.

Motowidlo, S. J., Dunnette, M. D., & Carter, G. W. (1990). An alternative selection procedure: The low-fidelity simulation. *Journal of Applied Psychology, 75,* 640-647.
A low-fidelity simulation was developed for selecting entry-level managers in the telecommunications industry. The simulation presents applicants with descriptions of work situations and five alternative responses for each situation. Applicants select one response they would most likely make and one they would least likely make in each situation. Results indicated simulation scores correlated from .28 to .37 with supervisory ratings of performance. These results show samples of even hypothetical work behavior can predict performance.

Motowidlo, S. J., & Tippins, N. (1993). Further studies of the low-fidelity simulation in the form of a situational inventory. *Journal of Occupational and Organizational Psychology, 66,* 337-344.
Authors examined two studies that extend the results of S. J. Motowidlo et al (1990) by providing further evidence about relations between situational inventory scores, job performance, and demographic factors. Combined results from both studies yield an overall validity estimate of .20, with small differences between race and sex subgroups, and confirm the potential usefulness of the low-fidelity simulation in the form of a situational inventory for employee selection.

Weekley, J. A., & Jones, C. (1999). Further studies of situational tests. *Personnel Psychology*, *52*(3), 679-700.
Results are reported for two different situational judgment tests (SJTs). Across the two studies, situational test scores were significantly related to cognitive ability and experience. In one study, there was a slight tendency for experience and cognitive ability to interact in the prediction of situational judgment, such that cognitive ability became less predictive as experience increased. Situational judgment fully mediated the effects of cognitive ability in one study, but not in the other. SJT race effect sizes were consistent with past research and were smaller than those typically observed for cognitive ability tests. The evidence indicates situational judgment measures mediate a variety of job relevant skills.

Structured Interviews References:

Campion, M. A., Palmer, D. K., & Campion, J. E. (1997). A review of structure in the selection interview. *Personnel Psychology, 50*(3), 655-702.
Reviews the research literature and describes and evaluates the many ways selection interviews can be structured. Fifteen components of structure are identified which may enhance either the content of or the evaluation process in the interview. Each component is critiqued in terms of its impact on numerous forms of reliability, validity, and user reactions. Finally, recommendations

for research and practice are presented. The authors conclude interviews can be easily enhanced by using some of the many possible components of structure, and the improvement of this popular selection procedure should be a high priority for future research and practice.

Conway, J. M., Jako, R. A., & Goodman, D. F. (1995). A meta-analysis of interrater and internal consistency reliability of selection interviews. *Journal of Applied Psychology, 80*(5), **565-579.**
A meta-analysis of 111 inter-rater reliability coefficients and 49 coefficient alphas from selection interviews was conducted. Moderators of inter-rater reliability included study design, interviewer training, and three dimensions of interview structure (standardization of questions, of response evaluation, and of combining multiple ratings) such that standardizing questions increased reliability of ratings more for individual vs. panel interviews, and multiple ratings were useful when combined mechanically (there was no evidence of usefulness when combined subjectively), and standardization of questions and number of ratings made resulted in greater levels of validity. Upper limits of validity were estimated to be .67 for highly structured interviews and .34 for unstructured interviews.

Huffcutt, A I., & Arthur, W. (1994). Hunter and Hunter (1984) revisited: Interview validity for entry-level jobs. *Journal of Applied Psychology*, *79*(2), **184-190.**
By adopting the theory of planned behavior, this study tried to predict human resources managers' intentions toward unstructured and structured interview techniques. Managers evaluated case descriptions of both techniques and were interviewed about their own practices. The data revealed stronger intentions toward unstructured interviewing than toward structured interviewing, which was consistent with their own practices in selecting staff, which appeared to be rather unstructured. Ajzen's (1991) theory appeared to be a useful framework for predicting managers' intentions. In particular, attitudes and subjective norms were predictive of intentions to engage in either method. Only intentions toward the unstructured case were related to managers' actual behavior.

Huffcutt, A. I., & Roth, P. L. (1998). Racial group differences in employment interview evaluations. *Journal of Applied Psychology, 83*(2), **179-189.**
The purpose of this meta-analysis was to research the various factors that can play a role in racial group differences resulting from an interview, such as the level of structure in the interview, job complexity, etc. Results suggest, in general, employment interviews do not affect minorities as much as other assessments (i.e., mental ability tests). Moreover, structured interviews tend to limit or decrease the influence of bias and stereotypes in ratings. High job complexity resulted in mean negative effect sizes for Black and Hispanic applicants, meaning they received higher overall ratings than White applicants. Behavior description interviews averaged smaller group differences than situational interviews, and group differences tended to be larger when there was a larger percentage of a minority (i.e., Black or Hispanic) in the applicant pool.

Huffcutt, A. I., Weekley, J. A., Wiesner, W. H., DeGroot, T. G., & Jones, C. (2001). Comparison of situational and behavior description interview questions for higher-level positions. *Personnel Psychology*, *54*(3), **619-644.**
This paper discusses two structured interview studies involving higher-level positions (military officer and a district manager) and had matching situational interviews and behavior description

interviews (BDI) questions written to assess the same job characteristics. Results confirmed results of previous studies finding situational interviews are less effective for higher-level positions than BDIs. Moreover, results indicated very little correspondence between situational and behavior description questions written to assess the same job characteristic, and a link between BDI ratings and the personality trait Extroversion. Possible reasons for the lower situational interview effectiveness are discussed.

McFarland, L. A., Ryan, A. M., Sacco, J. M., Kriska, S. D. (2004). Examination of structured interview ratings across time: The effects of applicant race, rater race, and panel composition. *Journal of Management, 30*(4), 435-452.
This study looked at the effect of race on interview ratings for structured panel interviews (candidates were interviewed and rated by three raters of varying races). Results indicated panel composition produced the largest effect. Specifically, predominately White panels provided significantly more favorable ratings (of all candidates, regardless of race) than panels which consisting of predominately Black raters. Panel composition also affected ratings, such that Black raters provided higher ratings to Black candidates only when the panel was predominately Black. However, the authors caution these effects were rather small; therefore, the results should be cautiously interpreted.

Taylor, P., & Small, B. (2002). Asking applicants what they would do versus what they did do: A meta-analytic comparison of situational and past behavior employment interview questions. *Journal of Occupational & Organizational Psychology, 75*(3), 277-294.
Criterion-related validities and inter-rater reliabilities for structured employment interview studies using situational interview (SI) questions were compared with those from studies using behavioral description interview (BDI) questions. Validities and reliabilities were further analyzed in terms of whether descriptively-anchored rating scales were used to judge interviewees' answers, and validities for each question type were also assessed across three levels of job complexity. While both question formats yielded high validity estimates, studies using BDI questions, when used with descriptively anchored answer rating scales, yielded a substantially higher mean validity estimate than studies using the SI question format with descriptively-anchored answer rating scales (.63 vs .47). Question type (SI vs. BDI) was found to moderate interview validity. Inter-rater reliabilities were similar for both SI and BDI questions, provided descriptively-anchored rating scales were used, although they were slightly lower for BDI question studies lacking such rating scales.

[Training and Experience (T & E) Evaluations](#) References:

Lyons, T. J. (1989). *Validity of Education and Experience Measured in Traditional Rating Schedule Procedures: A Review of the Literature*. **Office of Personnel Research and Development, U.S. Office of Personnel Management, Washington, DC, OPRD-89-02.**
This paper reviews research on the validity of specific education and experience measures common to traditional rating schedule procedures used by the Federal Government. The validity of each measure is discussed and recommendations for rating schedule use are offered.

Lyons, T. J. (1988). *Validity Research on Rating Schedule Methods: Status Report.* **Office of Personnel Research and Development, U.S. Office of Personnel Management, Washington, DC, OED-88-17.**
This report summarizes the findings from a series of studies conducted on rating schedule validity. The first objective was to investigate the criterion-related validity of rating schedules used in the Federal Government and the second was to study the validity of three rating schedule methodologies. Results indicated little evidence of validity for a rating schedule method based on training and experience at either entry-level or full performance level jobs. Findings supported the validity of a Knowledge, Skills, and Abilities (KSA)-based rating schedule method for full performance level jobs, but not for entry level jobs. Except for one entry level study, results indicated the most promising validity coefficients (in the mid to upper .20's) for rating procedures employing behavioral consistency measures for both entry and full performance level jobs.

McCauley, D. E. (1987). *Task-Based Rating Schedules: A Review.* **Office of Examination Development, U.S. Office of Personnel Management, Washington, DC, OED 87-15.**
This paper reviews the evidence for the validity and practicality of the task-based rating schedule (TBRS), a self-report instrument used to assess applicants' training and experience in relation to job required tasks. A review of the background of the TBRS and the assumptions on which it is based are discussed. In addition, a discussion of meta-analytic results on the predictive validity of the TBRS is provided.

McDaniel, M. A., Schmidt, F. L., & Hunter, J. E. (1988). A meta-analysis of the validity of methods for rating training and experience in personnel selection. *Personnel Psychology, 41*, **283-309.**
This paper discusses a meta-analysis of validity evidence of the methods (point, task, behavioral consistency, grouping, and job element) used to evaluate training and experience (T&E) ratings in personnel selection. Results indicate validity varied with the type of T&E evaluation procedure used. The job element and behavioral consistency methods each demonstrated useful levels of validity. Both the point and task methods yielded low mean validities with larger variability. Partial support was found for both the point and task methods being affected by a job experience moderator. Moderator analyses suggested the point method was most valid when the applicant pool had low mean levels of job experience and was least valid with an experienced applicant pool.

Schwartz, D. J. (1977). A job sampling approach to merit system examining. *Personnel Psychology, 30*(2), **175-185.**
A method for collecting content validity evidence for a merit examining process or rating schedule without violating the principles of content validity is presented. This technique, called the job sampling approach, is a task-based, structured system of eliciting the information necessary to construct the rating schedule from sources most able to provide that information, and for using the information to construct the rating schedule and linking it to job performance. The steps include definition of the performance domain of the job in terms of process statements, identification of the selection and measurement objectives of the organization, development of the measurement domain in relation to the performance domain and to the selection and

measurement objectives, and demonstration that a close match between the performance domain and the measurement domain was in fact achieved.

Sproule, C. F. (1990). *Personnel Assessment Monographs: Recent Innovations in Public Sector Assessment* **(Vol 2, No. 2)***.* **International Personnel Management Association Assessment Council (IPMAAC).**
This report reviews selected assessment methods and procedures (e.g., training and experience measures) used frequently in the public sector during the 1980s. Many of these, including the rating schedule, are still used today. Each section on assessments contains a variety of examples describing public sector methods, as well as a summery of related research findings. Other sections include discussions on selected Federal assessment innovations, application of technology to assessment, use of test scores, legal provisions related to assessment, and employment testing of persons with disabilities.

[Work Samples and Simulations](#) References:

Campion, J. E. (1972). Work sampling for personnel selection. *Journal of Applied Psychology, 56***(1), 40-44.**
The author developed a work sample checklist from supervisors' ratings of important tasks and behaviors expected of maintenance mechanics. The test was given to 34 employees by an outside consulting firm, and checklist evaluations were compared with (a) supervisors' evaluation of each employee on three factors; and (b) employee's responses on the test of mechanical comprehension, the Wonderlic Personnel Test, and the short employment tests. Results indicated performance on the work sample was significantly related to supervisory evaluations of job success, but none of the validity coefficients for the paper-and-pencil tests were significant.

Gilliland, S. W. (1995). Fairness from the applicants' perspective: Reactions to employee selection procedures. *International Journal of Selection and Assessment, 3***(1), 11-19.**
Applicants' reactions to selection procedures were examined in terms of the satisfaction and/or violation of ten procedural justice rules. Critical incidents (n = 237) of fair and unfair treatment during selection were collected from 31 individuals who had recently experienced job search and hiring processes. Incidents were categorized into ten procedural justice rules and the distribution of these incidents was examined for different hiring outcomes and different selection procedures. Dominant procedural concerns reflected selection procedure job relatedness and interpersonal treatment applicants received. Accepted applicants were primarily concerned about consistency of treatment, while rejected applicants were more concerned with timely feedback and blatant bias. Ease of faking was the primary procedural concern of applicants taking honesty and personality tests, while job relatedness was the primary concern with ability and work sample tests. Research issues were discussed and a number of practical suggestions were offered in terms of minimizing applicants' negative reactions to the selection process.

Lance, C. E., Johnson, C. D., Douthitt, S. S., Bennett, W., & Harville, D. L. (2000). Good news: Work sample administrators' global performance judgments are (about) as valid as we've suspected. *Human Performance, 13*(3), 253-277.
Data obtained on over 1,500 first-term U.S. Air Force enlisted personnel indicated work sample administrators' global ratings of work sample performance substantially reflect actual ratee behavior in the work sample, and not potentially biasing factors (e.g., race, gender, amount of recent experience), supporting the "folk wisdom" these global performance judgments are, in fact, valid and unbiased measures of performance.

Robertson, I. T. & Kandola, R. S. (1982). Work sample tests: Validity, adverse impact and applicant reaction. *Journal of Occupational Psychology, 55*(3), 171-183.
Work sample tests are assigned to one of four categories: (1) psychomotor; (2) individual, situational decision making; (3) job-related information; and (4) group discussion/decision making. Validity data drawn from over 60 studies are presented and show psychomotor work sample tests and group discussions predict job performance relatively well when compared with more conventional forms of psychological testing, such as intelligence or personality tests. Data showing other validity relationships are presented, and the importance of point-to-point correspondence between predictors and criteria is discussed. Research on the adverse impact of work sample tests and applicant reaction to such tests is reviewed and suggests the tests may help to reduce adverse impact and produce positive reactions from candidates.

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*(2), 262-274.
This article summarizes the practical and theoretical implications of 85 years of research in personnel selection. On the basis of meta-analytic findings, this article presents the validity of 19 selection procedures for predicting job performance and training performance and the validity of paired combinations of general mental ability (GMA) with the 18 other selection procedures. Overall, the three combinations with the highest multivariate validity and utility for job performance were 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 they can be used for both entry level selection and selection of experienced employees. The practical utility implications of these summary findings are substantial. The implications of these research findings for the development of theories of job performance are discussed.

[Background Evaluation/Investigation](#) References:

Hilliard, P. A. (2001). Comparison of the predictive validity of a written test, an integrity test, a conscientiousness questionnaire, a structured behavioral interview and a personality inventory in the assessment of job applicants' background investigations, and subsequent task or contextual performance. *Dissertation Abstracts International: Section B: The Sciences & Engineering*, *62*(6-B), 2981.
This study was designed to compare the validity of several personnel selection instruments in the prediction of the results of applicants' background investigations, and incumbents' subsequent contextual and task work performance. The selection instruments used were a written test, an

integrity test, a conscientiousness questionnaire, a structured behavioral interview, and a personality inventory. Out of a total of 168 applicants who were interviewed, 23 were subsequently hired, and job performance evaluations were collected for 18 employees. Although there were some statistically significant findings (e.g., interviews and conscientiousness predicted background investigation results and task performance), the primary hypotheses were not supported. An additional findings was many potential applicants screened themselves out of the process early in the proceedings once they realized an extensive background investigation would be conducted.

McDaniel, M. A. (1989). Biographical constructs for predicting employee suitability. *Journal of Applied Psychology, 74*(6), 964-970.
The use of background investigation data in personnel screening is reviewed. Background investigations are routinely conducted on persons seeking positions of trust in law enforcement, the nuclear power industry, and military and civilian occupations requiring government-issued security clearances. The application of background investigation information in personnel screening differs in many important ways from biodata applications developed by psychologists. In this article, these differences are reviewed, and the construct and criterion-related validity of a survey-based measure are examined. This measure taps content areas typically explored in background investigations. Seven background factors were identified. The background scales showed acceptable reliability, informative interscale relationships, and useful levels of criterion-related validity.

McFadden, K. L. (1997). Policy improvements for prevention of alcohol misuse by airline pilots. *Human Factors, 39*(1), 1-8.
Analyzes two strategies for reducing pilot-error aviation accidents: conducting background checks on pilots for driving-while-intoxicated (DWI) convictions, and random preflight alcohol testing of airline pilots. The results of this study are based on analysis of data obtained from the Federal Aviation Administration on the flying performance of 70,164 airline pilots. DWI convictions were found to be associated with a significantly greater risk of a pilot-error accident. In contrast, no evidence was found to validate the assumption a random alcohol testing program could have prevented accidents. The results provide support for improving the existing DWI background check program and for reducing the sampling rate of random alcohol testing for airline pilots. This twofold strategy could result in greater improvements in aviation safety and reduced overall costs.

Job-Fit Measures References:

Arthur, W., Jr., Bell, S. T., Villado, A. J., & Doverspike, D. (2006). The use of person-organization fit in employment decision making: An assessment of its criterion-related validity. *Journal of Applied Psychology, 91,* 786-801.
This article cautions against the use of P-O measures as selection tools. Using meta-analytic analyses, the criterion-related validity of P-O fit as a predictor of job performance was only .15 with a small effect size. Another consideration is if using P-O as a criterion-related selection tool, it needs to abide by the same professional standards and practices (e.g., psychometric and legal) as other selection tests as outlined by the Civil Rights Act (CRA; 1964, 1991) and Equal Employment Opportunity Commission (EEOC; 1978). At this time, the authors suggest it is best

to limit P-O measures for after the candidate has been hired, such as for placement, and not selection.

Cable, D. M., & Judge, T. A. (1997). Interviewers' perceptions of person-organization fit and organizational selection decisions. *Journal of Applied Psychology, 82,* **546-561.**
A model of person-organization fit and organizational hiring decisions is developed and tested, using data from 38 interviewers making hiring decisions about 93 applicants. Results suggest interviewers can assess applicant-organization values congruence with significant levels of accuracy and interviewers compare their perceptions of applicants' values with their organizations' values to assess person-organization fit. Results also suggested interviewers' subjective person-organization fit assessments have large effects on their hiring recommendations relative to competing applicant characteristics, and interviewers' hiring recommendations directly affect organizations' hiring decisions (e.g., job offers).

Dineen, B. R., Ash, S. R., & Raymond, N. A. (2002). A web of applicant attraction: Person-organization fit in the context of web-based recruitment. *Journal of Applied Psychology,* *87*(4), **723-734.**
Applicant attraction was examined in the context of Web-based recruitment. Specifically, the provision of feedback to individuals regarding their potential Person-Organization (P-O) fit with an organization related to attraction was studied. Objective and subjective P-O fit, agreement with fit feedback, and self esteem also were examined in relation to attraction. Results of an experiment that manipulated fit feedback level after a self-assessment provided by a fictitious company Web site found both feedback level and objective P-O fit were related to attraction. In addition, attraction was related to the interaction of objective fit, feedback, and agreement and objective fit, feedback, and self esteem. Implications and future Web-based recruitment research directions are discussed.

Judge, T. A., & Cable, D. M. (1997). Applicant personality, organizational culture, and organizational attraction. *Personnel Psychology, 50,* **359-394.**
Examined the dispositional basis of job seekers' organizational culture preferences and how these preferences interact with recruiting organizations' cultures in their relation to organization attraction. Data were collected from 182 business, engineering, and industrial relations students who were seeking positions at the time of the study. Results obtained from multiple sources suggested the Big Five personality traits generally were related to hypothesized dimensions of culture preferences. Results also suggested both objective person-organization fit and subjective fit (applicant's direct perception of fit) were related to organization attraction.

Kristof, A. L. (1996). Person-organization fit: An integrative review of its conceptualizations, measurement, and implications. *Personnel Psychology, 49,* **1-49.**
Presents a comprehensive definition and conceptual model of person-organization fit that incorporates supplementary as well as complementary perspectives on fit. To increase the precision of the construct's definition, it is also distinguished from other forms of environmental compatibility, such as person-group and person-vocation fit. Measurement as it relates to supplementary and complementary fit is discussed and recommendations are offered regarding the necessity of its use. A distinction is made between the direct measurement of perceived fit

and the indirect measurement of actual person-organization fit, and the debate regarding differences scores is reviewed.

Martinez, M. N. (2000). Get job seekers to come to you. *HR Magazine, 45*, **45-52.**
This article discusses ways organizations can use their website to attract top talent. Interactive self-assessments of job fit are promoted as one effective practice. The author states applicants are desperate for tools that help them determine their "fit" in an organization and the key is to give them results.

[Physical Ability Tests](#) References:

Arvey, R. D., Maxwell, S. E., & Salas, E. (1992). Development of physical ability tests for police officers: A construct validation approach. *Journal of Applied Psychology, 77,* **996-1009.**
A construct validation approach was followed and results indicated eight physical ability test events were significantly related to two important constructs underlying the job performance of police officers: strength and endurance. In addition, the data were examined for potential gender differences and bias. Considerable differences were shown between men and women on both test and performance variables.

Arvey, R. D., Nutting, S. M., & Landon, T. E. (1992). Validation strategies for physical ability testing in police and fire settings. *Public Personnel Management, 21,* **301-312.**
Discusses two issues within the context of selection and staffing for police and fire fighting positions: (1) the increasing litigation challenging the validity of physical ability tests in screening applicants for these positions, and (2) the lack of published literature concerning the use and validation of physical ability tests in these settings. The authors discuss issues associated with court challenges of traditional content validity procedures and suggest there may be an over-reliance on this particular validation strategy. They suggest construct validation procedures may be an alternative procedure to provide evidence concerning the validity of physical ability tests within these contexts. Construct validation is described and illustrated via hypothetical examples.

Campbell, W. J., & Fox, H. R. (2002). Testing individuals with disabilities in the employment context: An overview of issues and practices. In R. B. Ekstrom & D. K. Smith (Eds.) *Assessing Individuals with Disabilities in Educational, Employment, and Counseling Settings* **(1st ed, p. 198). Washington, DC: American Psychological Association.**
This chapter discusses the possible issues that can occur when testing individuals with disabilities. It noted while the Americans with Disabilities Act (ADA) permits the use of physical ability tests, the tests cannot include (or involve) physiological or biological measures. They also warned of the possibility of applicant's hurting themselves while performing a physical ability test. Depending on the extensiveness of the physical ability test, employers may request applicants to provide a certificate from their physicians indicating their ability to safely perform a physical ability test.

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

Central issue of this paper is improvement in personnel selection systems for physically demanding jobs is needed due to equal employment opportunity (EEO) considerations, concern for worker physical well-being, and the lack of alternative procedures. After addressing the special EEO sensitivies of physical abilities selection, the literature is reviewed 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). The structure of physical performance in occupational tasks. *Journal of Applied Psychology, 76,* 495-507.

Two lines of research concerning the dimensionality of physical performance in occupational tasks are described. In the first, the physical requirements of tasks are analyzed as reflected in job analyses. In the second, the structure of physical abilities tests used to predict performance in physically demanding jobs is evaluated. Results of the data analysis suggest the structure of physical abilities has three major components: strength, endurance, and movement quality. This structure appears to be independent of job type or level of incumbents' performance.

[Realistic Job Previews](#) References:

McEvoy, G. M., & Cascio, W. F. (1985). Strategies for reducing employee turnover: A meta-analysis. *Journal of Applied Psychology, 70*(2), 342-353.

In looking at the relative effectiveness of the Realistic Job Preview (RFP) as a turnover-reduction strategy, this meta-analysis found RJPs are about half as effective as job enrichment strategies in reducing turnover. However, it should be noted the job enrichment measures included in this meta-analysis were used with existing workers, whereas the RJPs were used during the application process. Given the low correlation between RJPs and reduction in turnover, the authors suggest managers use other turnover reduction strategies prior to hiring.

Pitt, L. F., & Ramaseshan, B. (1995). Realistic job information and salesforce turnover: An investigative study. *Journal of Managerial Psychology, 10*(5), 29-36.

Salespeople were asked to provide ratings on the volume, personal relevance, depth, and accuracy of information they received when they applied for their current jobs. They were also asked about their intention to quit their current job. The results indicated those applicants (employees) who felt they received more accurate and realistic job information were significantly less likely to indicate (or consider) quitting their current jobs.

Saks, A. M, Wiesner, W. H., & Summers, R. (1996). Effects of job previews and compensation policy on applicant attraction and job choice. *Journal of Vocational Behavior, 49,* 68-85.

Students were asked to rate the attractiveness of jobs after reviewing a Realistic Job Preview (RJP) and a Traditional Job Preview (TJP). Jobs with RJPs were rated as more attractive only when those positions had a higher compensation than the TJP position. Results of this study indicate many other factors (with compensation being studied here) may contribute to an applicant's decision to accept (or reject) a job when coupled with an RJP.

Wanous, J. P. (1989). Installing a realistic job preview: Ten tough choices. *Personnel Psychology, 42*(1), 117-133.
The author reviews the ten different choices one must make when constructing a Realistic Job Preview (RJP). Examples of the choices include whether the content should be descriptive or judgmental, contain high or medium negativity, or whether the results should be shared or kept proprietary. For each choice, the author includes the pros and cons of each. The author concludes the article with personal recommendations for each of the ten choices.