JOB SIMULATIONS:
TRYING OUT FOR A FEDERAL JOB

U.S. MERIT SYSTEMS PROTECTION BOARD
1615 M Street, NW
Washington, DC 20419-0001

September 2009

The President
President of the Senate
Speaker of the House of Representatives

Dear Sirs and Madam:

In accordance with the requirements of 5 U.S.C. 1204(a)(3), it is my honor to submit this Merit Systems Protection Board (MSPB) report, *Job Simulations: Trying Out for a Federal Job*. The purpose of this report is to examine the utility of job simulation assessments for use in the Federal hiring process.

We define job simulation as an assessment that presents applicants with realistic, job-related situations and documents their behaviors or responses to help determine their qualifications for the job. Job simulation assessments can be a critical measure in ensuring that agencies are selecting the best candidates. Simulations can do a better job of predicting which applicants will perform well on the job than many other commonly used assessments, and they can provide a greater degree of fairness in the process. However, their potentially high development cost is a key drawback.

Job simulations may not work in every situation. It is important, therefore, for agencies to have a good grasp of the job for which they are hiring, the competencies needed for that job, and knowledge about which assessments would best fit their specific needs. This report identifies a number of factors for agencies to consider when making decisions about their assessment process. Also, it identifies a 5-step assessment strategy we believe should help agencies develop and implement an assessment program that assists them in selecting employees on the basis of relative ability, knowledge, and skills, as set forth in the merit system principles.

I believe that you will find this report useful as you consider issues affecting the Federal Government's ability to assess and select a high-quality workforce now and in the future.

Respectfully,

Neil A. G. McPhie

JOB SIMULATIONS:

TRYING OUT FOR A FEDERAL JOB

U.S. Merit Systems Protection Board

Neil A. G. McPhie, *Chairman*

Mary M. Rose, *Vice Chairman*

Office of Policy and Evaluation

Director
John Crum, Ph.D.

Project Manager
Laura Shugrue

Table of Contents

The U.S. Merit Systems Protection Board (MSPB) has published a number of reports that highlight the importance of using good assessment tools. We have examined several assessment tools to determine their ability to improve the quality of agencies' employee selections, including structured interviews, reference checks, and the probationary period. Job simulations, may also be an effective way for agencies to identify high-quality applicants.

The purpose of this report is to examine the utility of job simulation assessments for use in the Federal hiring process. *A **job simulation** is an assessment that presents applicants with realistic, job-related situations and documents their behaviors or responses to help determine their qualifications for the job.* Job simulations include, but are not limited to, work samples, situational judgment tests, assessment centers, and job tryout procedures.

Findings

Job simulations present a number of important advantages as well as several significant disadvantages.

Advantages

- *Higher Validity.* Job simulation assessments tend to have higher predictive validity than many other types of assessments, meaning they should be better at predicting future job performance.
- *Better Job Fit.* By exposing applicants to the types of events, scenarios, and challenges confronted on the job, job simulations help applicants determine if the job is well suited to their knowledge, skills, abilities, and interests.
- *Positive Applicant Perceptions.* Because job simulations replicate the types of tasks performed in the actual job, studies have found that applicants are more likely to view them as being fair and job-related.
- *Greater Degree of Fairness.* Research generally supports the premise that job simulation assessments have lower rates of adverse impact (i.e., a different rate of employment selection that works to the disadvantage of members of a race, sex, or ethnic group), as well as a lower degree of exposure to discrimination lawsuits based on the selection procedure.

Disadvantages

- *Cost*. Job simulations are sometimes more costly because of the resources necessary to develop and administer the tests. The more advanced simulations require greater expertise to develop than other, less complicated assessments, and this expertise results in higher developmental costs. Plus, these advanced assessments can require more staff and training to administer and assess the results.
- *Limited Scope.* While job simulations can be used to assess multiple competencies, a single simulation exercise will often focus on a limited number of tasks or duties performed on the job. Therefore, job simulations are best used as one assessment in a series of valid selection tools.
- *Not Suited to All Jobs*. Many job simulations require the applicant to already have a certain level of knowledge, skills, or abilities (KSAs) to complete the assessment. Therefore, they may not be appropriate for some entry-level or generalist occupations.

Agency Use of Job Simulations

Job simulations are not being used extensively in the Federal Government, with less than half of the organizations we questioned indicating that they use them. The top barriers cited were related to a lack of resources and knowledge rather than amenability. This finding suggests that agencies might increase the use of job simulations if they had more knowledge or training in developing and administering these assessments and more resources to devote to these steps.

Organizations that use job simulations the most tend to use them for mission-critical occupations or specialty areas (such as leadership). They also tend to be fairly homogenous agencies that do volume hiring into one or two entry-level, mission-critical occupations. Finally, agencies that use them tend to have budgets committed to assessment activities, dedicated assessment staff, and/or leaders who support devoting resources to workforce management issues.

In terms of who develops and administers job simulations, agency selecting officials and subject matter experts seem to take the lead. Unfortunately, evidence suggests that they may not receive the training necessary to take on these responsibilities.

Most of the organizations that used job simulations in the past year believed that the simulation tests increased the quality of the agency's selections. However, few organizations reported actually measuring the impact.

Formulating an Assessment Strategy

There are many factors to consider when choosing the type of assessment an agency should use to evaluate applicant qualifications. The use of job simulations may or may not be appropriate. Therefore, instead of providing specific recommendations about whether to use job simulations, we offer an overall assessment strategy that will help agencies determine which assessments— simulation or non-simulation— would be best for the specific hiring situation.

(1) **Conduct a job analysis.**

The typical job analysis will determine the relative importance of the job's duties and tasks and identify the KSAs or competencies necessary to perform them. In addition, the job analysis will identify other factors that impact the ability to carry out the job's duties, such as tools and technology, work environment and relationships, and training and licensure requirements. With this information, the organization can identify which assessments— simulation or non-simulation— would best measure the identified competencies. We found anecdotal evidence suggesting that some agencies are doing an inadequate job of conducting job analysis for vacancies they are filling, a problem that impacts the succeeding steps in the assessment strategy.

(2) **Identify the assessments.**

When identifying the assessment tools for the specific hiring situation and how they can best be used in the process, organizations should consider such factors as: the availability of resources; the skills being tested; the quality and validity of the assessment tool; the number of assessments to be used; the number of expected applicants; and applicant burden.

(3) **Develop testing procedures.**

After selecting the assessment tools to be used, the organization must choose who will develop the testing procedures. Those chosen should be qualified test developers with experience not only developing tests, but also validating and defending them legally when necessary.

(4) **Select qualified assessors.**

Next, the organization needs to select qualified assessors, considering such factors as the individuals' commitment to the process, time constraints, knowledge of the job being filled, understanding of the assessment, and knowledge of how to effectively score candidates. Diversity in terms of not only race, ethnicity, and gender, but also position

should be considered. Because unqualified assessors will undermine the quality and value of the assessment, only well-trained assessors should be used. If the individuals used to administer the test are different from those who score it, they should also be well-trained.

(5) Measure the success of the assessment procedures.

Finally, evaluating the impact of the assessment procedures on the selection is a critical step to allow the organization to determine which assessments identify the best candidates and, therefore, where assessment funding and resources should be targeted. Further, measuring the costs, benefits, and results of the procedures can help justify resource allocation to agency leadership, a critical step in this time of limited resources.

Introduction

Background

The Federal Government is currently experiencing a surge in interest from potential applicants. As the number of applicants increases, agencies need to be careful to use good assessment tools to distinguish the most qualified applicants. Job simulations may help accomplish this goal.

While many once considered the Government as an employer of last choice, many applicants now view a Federal job as a good employment opportunity. Some view it as a chance to effect change, perhaps because of President Barack Obama's encouragement to young people to get involved in Government at every level and his determination "to make Government cool again."[1] Others view Federal employment favorably because the economy has not been kind to many private and nonprofit sector employees, and they are looking for the type of job security and stability the Government can often provide. Whatever the reason, the Federal Government has seen an increase in applicant interest.

For example, by December 2008, 330,000 people had applied for positions within the Obama Administration. That compares to 44,000 applicants in President George W. Bush's 2000-01 transition and 135,000 in President Clinton's 1992-93 transition.[2] At the same time, the Office of Personnel Management (OPM) reported that the number of visitors per month to USAJOBS (www.usajobs.gov)—the Federal Government's official jobs Web site—almost doubled from the previous year, to 13.6 million visits. Many agencies have noted increased interest as well. For example, the Government Accountability Office (GAO) saw a 150-percent increase in applications, including 1,300 applicants for positions related to studying the Federal bailout.[3] Finally, a 2009 survey of university career center officials found that college students are becoming increasingly interested in Federal employment.[4]

---

[1] Barack Obama's comments at the "Service Nation Presidential Candidates Forum," Columbia University, New York, NY, September 11, 2008.

[2] Mimi Hall, "330,000 Applicants Vie for Administration Jobs," *USATODAY.com*, December 12, 2008.

[3] Stephen Losey, "HR Staffs Deluged by Millions More Applicants," *FederalTimes.com*, December 21, 2008.

[4] Steve Vogel, "Survey Finds College Students More Interested in Federal Jobs," *Washington Post*, May 1, 2009.

These facts present a snapshot in time. We do not know how job seekers will rate the Government in the future. A year or two from now, Federal recruiters may find themselves standing alone at job fairs while applicants once again swarm Wall Street or the next cutting edge employer that offers huge salaries, signing bonuses, and prestige. What we do know, however, is that the Government always needs qualified people who have the ability and motivation to carry out its many missions. For this reason, the Government needs to ensure it is hiring the right people, and that is where good assessment practices come into play.

Purpose and Methodology

The merit system principles (MSPs), found in 5 U.S.C. § 2301, constitute the framework for Federal human resources (HR) management. The second MSP states that "selection and advancement [for Federal jobs] should be determined solely on the basis of relative ability, knowledge, and skills."[5] Assessment is the phase in the hiring process when agencies determine the extent to which applicants possess the knowledge, skills, and abilities required to do the job. The purpose of assessment is to identify the best-qualified applicants.

The MSPB is responsible for conducting special studies of Federal HR practices, policies, and procedures to ensure that they adhere to the merit system principles and do not result in prohibited personnel practices. As part of this mission, we have published a number of reports that highlight the importance of using good assessment tools. Specifically, we have looked at several assessment tools to determine whether they improve the quality of agencies' employee selections, including structured interviews, reference checks, and the probationary period.[6]

The purpose of this report is to examine the utility of job simulation assessments for use in the Federal hiring process. For the purpose of this report, ***a job simulation*** *means an assessment that presents applicants with realistic, job-related situations and documents their behaviors or responses to help determine their qualifications for the job.* Job simulations include but are not limited to work samples, situational judgment tests, assessment centers, and job tryout procedures. We focus specifically on how agencies use these assessments for hiring new employees rather than for promoting employees. This report:

---

[5] 5 U.S.C. § 2301(b)(1).

[6] Go to www.mspb.gov and select "MSPB Studies" to read MSPB studies and newsletters. Specifically, see the following reports we have published concerning assessment methods: *The Federal Selection Interview: Unrealized Potential*, February 2003; *Reference Checking in Federal Hiring: Making the Call*, September 2005; and *The Probationary Period: A Critical Assessment Opportunity*, August 2005.

- Examines the advantages and disadvantages of using job simulations;
- Reviews how Federal agencies develop and use job simulations to assess applicant qualifications;
- Determines whether this type of assessment can help agencies improve the quality of their employee selections; and
- Introduces readers to simulation tools that they may not be familiar with and directs them to resources with more information about how to develop and use them.

This study relied primarily on the following sources of information:

- **Literature Review.** We conducted a review of literature related to job simulation assessments, including work samples, situational judgment tests, assessment centers, and job tryout procedures. In addition, we examined professional journals and other articles related to overall trends and innovations in employee assessment.
- **Agency Questionnaires.** We sent questionnaires to 28 Federal agencies asking them about their assessment practices and, in particular, how they use job simulation assessments.[7] We asked agencies to ensure that individuals with knowledge of the agencies' assessment practices complete the questionnaires. We received responses from 20 different Federal agencies, including 12 Cabinet level departments and 8 independent agencies. Because agency assessment programs are often delegated to the sub-component level (e.g., the Veterans Health Administration in the Department of Veterans Affairs or the Internal Revenue Service in the Department of the Treasury), we received multiple responses from some agencies. Therefore, we analyzed a total of 35 questionnaire responses, including responses from the department and/or sub-component level.
- **Interviews with Employers.** To follow up on the results from the questionnaire responses, we interviewed four organizations that had particularly interesting examples of how they have used job simulation assessments. These examples are highlighted in Appendix A.
- **Interviews with OPM.** We conducted interviews with colleagues in two OPM offices to gain more information on research and policy work that OPM has done regarding job simulations, as well as to find out what experiences they have had helping agencies implement these types of assessments.

---

[7] See Appendix B for a copy of the interrogatory and the instructions.

Benefits of Good Assessments

MSPB's research has shown that Federal selecting officials are generally satisfied with the employees they hire.[8] However, we know from our past research that hiring poor performers can have a much greater negative impact on the workforce than their small numbers would suggest.[9] Given this finding, there are a number of benefits that agencies can realize by using better assessment tools.

Studies have shown that making selections based on the applicant's ability to do the work can lead to higher organizational performance and increased financial benefits.[10] Specifically, utility studies indicate that increasing the predictive ability of an assessment will increase the percent of new hires who will perform satisfactorily on the job.[11] For instance, a well known assessment study found that a "superior skilled worker" produces 32 percent more output than an average worker, and a superior manager or professional produces 48 percent more output.[12] Another study in a related line of research suggests that employing a superior, rather than average, performer can result in sizeable economic and productivity gains.[13]

---

[8] For a summary of these findings, see *Testimony of John Crum, Director, Office of Policy and Evaluation, U.S. Merit Systems Protection Board, Before the Senate Subcommittee on Oversight of Government Management, the Federal Workforce, and the District of Columbia*, May 8, 2008, http://hsgac.senate.gov/public/_files/CrumTestimony050808.pdf.

[9] U.S. Merit Systems Protection Board, *Federal Supervisors and Poor Performers*, July 1999.

[10] For instance, see David E. Terpstra and Elizabeth J. Rozell, "The Relationship of Staffing Practices to Organizational Level Measures of Performance," *Personnel Psychology,* Spring 1993, Vol. 46, and Watson Wyatt Worldwide, *The Human Capital Index: Linking Human Capital and Shareholder Value*, 1999.

[11] H.C. Taylor and J.T. Russell, "The Relationship of Validity Coefficients to the Practical Effectiveness of Tests in Selection: Discussion and Tables," *Journal of Applied Psychology*, Vol. 23, 1939, pp. 565-578. The Taylor-Russell tables calculate the increase in performance using the base rate of success, the level of assessment validity, and the selection ratio.

[12] Frank L. Schmidt and John E. Hunter, "The Validity and Utility of Selection Methods in Personnel Psychology: Practical and Theoretical Implications of 85 Years of Research Findings," *Psychological Bulletin*, the American Psychological Association, Inc., Vol. 124, No. 2, September 1998, p. 263.

[13] For instance, see Lyle Spencer's work regarding the economic value of competency-based HR applications. For example, "The Economic Value of Emotional Intelligence Competencies and EIC-Based HR Programs," *The Emotionally Intelligent Workplace: How to Select for, Measure, and Improve Emotional Intelligence in Individuals, Groups and Organizations*, C. Cherniss and D. Goleman, eds, Jossey-Bass/Wiley: San Francisco, CA, 2001.

Research also indicates that selecting the wrong applicants can increase costs and decrease productivity. Some estimates show that the cost of hiring a person who does not have the right KSAs is up to three times the employee's salary in terms of wasted hiring time, training costs, salary, benefits, and severance pay.[14] Poor selections can also result in lost productivity because of the time it takes to correct performance and conduct problems. For instance, a 2004 study found that managers in the United States spend about 13 percent of their time correcting employees' mistakes.[15] Other research has found that poor selection can lead to increased turnover and absenteeism.[16]

To reduce the costs associated with bad selections and improve the organization's ability to carry out its mission, it is important to employ good assessment strategies that help identify the best candidates for the job.

What Makes a Good Assessment?

Having reviewed the professional literature, there are several factors that appear to serve as the foundation of good employee assessment:

The assessment should be reliable. Reliability means that the score an applicant receives on the assessment is consistent. For instance, if the applicant were to take the test more than once, with no performance intervention, the score should be similar each time. In short, reliability ensures that random sources of error are minimized during the assessment so that there is consistency and repeatability.

The assessment should be valid. Validity is critical to determining which assessments will produce the greatest results. Validity is the relationship between performance on an assessment and performance on the job. Specifically, validity demonstrates if the assessment measures a job-related characteristic and how well it measures that characteristic. There are three key methods used to determine the validity of an assessment:

- *Content-related validation* is evidence that the questions or tasks included in an assessment actually represent the knowledge, skills, or abilities necessary to perform the job. Content validity is generally based on the results of a job analysis and subject matter expert judgment. Content validity is the easiest to achieve and is, therefore, a popular method used in Federal assessment practices.

---

[14] Corporate Leadership Council, Literature Review, "Employee Selection Tests," Catalog No. 070-198-213, Washington, DC, March 1998, p. 2.

[15] The Future Foundation and SHL, *Getting the Edge in the New People Economy*, 2004, p. 31.

[16] The Partnership for Public Service, *Asking the Wrong Questions: A Look at How the Federal Government Assesses and Selects Its Workforce*, Washington, DC, October 2004, p. 3.

- *Construct-related validation* requires a demonstration that the assessment being used measures the construct (i.e., the psychological or physical characteristics that distinguish between people and how they will perform on the job, such as intelligence or sociability) it was intended to measure.

- *Criterion-related validation* is the proven statistical relationship between performance on the assessment and performance on the job. It requires a validation study in which job incumbents are tested and the tests are compared to performance evaluation results (concurrent validity) or job applicants are tested and tests are later compared to performance results on the job (predictive validity). Because of its ability to best predict an applicant's future performance on the job, criterion-related validity should be what organizations strive to achieve. As previously indicated, studies have shown that an increase in predictive validity will increase the percentage of new hires who will perform satisfactorily on the job.[17] However, a validation study also requires a certain level of expertise to achieve, making criterion-related validation more resource intensive.

   The most commonly used measure of predictive validity is the validity coefficient. This coefficient ranges in absolute value from 0 to 1.00. The closer the score is to 1, the stronger the relationship between the assessment tool results and job performance. A validity coefficient of .30 or higher is generally considered desirable for purposes of employee assessment.

The assessment should be used for the appropriate target population. Different assessments are appropriate for different situations. For instance, a job knowledge test that is designed to test an applicant's knowledge of a specific field should not be used for an entry-level position that does not require prior job knowledge or experience. Another example is the multiple hurdle approach that MSPB has recommended in earlier reports. As part of this process, agencies use fast, accessible, and low-cost instruments in the early part of the assessment process, thereby saving assessments that have a higher cost to administer for later in the process when more precision is needed to identify the best candidate.[18]

Applicants should have positive perceptions about the assessment process. An applicant's perception that an assessment is fair and job-related promotes cooperation and acceptance of the assessment process.[19] Specifically, research shows that applicant perceptions of the selection process are important to their overall perception of the

---

[17] Taylor and Russell, "The Relationship of Validity Coefficients to the Practical Effectiveness of Tests in Selection: Discussion and Tables," pp. 565-578.

[18] U.S. Merit Systems Protection Board, *Reforming Federal Hiring: Beyond Faster and Cheaper*, July 2006, pp. 27-28

[19] U.S. Office of Personnel Management, *Assessment Decision Guide*, http://apps.opm.gov/ADT.

organization and whether they want to work for it. Results from a meta-analysis of 86 independent studies demonstrate that applicants with favorable perceptions of the selection process were more likely to view the organization favorably. Therefore, they were more likely to accept job offers and to recommend the employer to others.[20] We know from previous MSPB research that word-of-mouth is a key way applicants find out about Federal employment opportunities, so these recommendations are important.

The assessment should be fair and unbiased. It is important to use assessment tools that are as unbiased and fair as possible and that do not result in adverse impact. Adverse impact means that the assessment creates a substantially different rate of selection for individuals of a particular ethnicity, race, or gender. To help hiring organizations comply with this principle and make proper use of tests and other selection procedures, the Equal Employment Opportunity Commission, the Department of Labor, the Department of Justice, and the former Civil Service Commission developed the *Uniform Guidelines on Employee Selection Procedures*.[21] It is important to ensure that assessment practices conform to these guidelines, not only to help improve applicant perceptions of the assessment process, but also to help save the organization from unnecessary litigation over selection practices.

In the next section, we discuss the different kinds of assessment methods that fall under our definition of "job simulation."

---

[20] John P. Hausknecht, David V. Day, and Scott C. Thomas, "Applicant Reactions to Selection Procedures," *Personnel Psychology*, Autumn 2004, Vol. 57, No. 3, p. 639.

[21] http://www.uniformguidelines.com/uniformguidelines.html.

Definition

There is an American television series called *House* about a brilliant, but somewhat unbalanced, doctor who works with a team of diagnosticians to diagnose some of the most rare and difficult medical cases. At one point in the series, Dr. House's team disbanded, and the hospital administrator forced him to hire two new fellows. She provided Dr. House with 40 resumes of qualified physicians from which to choose. Dr. House hired all 40. He then spent several episodes observing and evaluating the new doctors' performance, firing those who did not perform to his standards.

This type of flexibility in the hiring process would be ideal. After all, it seems that hiring all qualified applicants and evaluating their actual performance on the job over time would be the best way to determine who is the best qualified. However, Dr. House is a fictional character in a fictional hospital that does not have the same kind of resource limitations real organizations have. As mentioned previously, hiring—and then firing—the wrong person for the job is costly. Furthermore, applicants generally would not tolerate such a life-disrupting hiring process.

There are other options, though. The next best thing to hiring all qualified applicants and evaluating their on-the-job performance may be to evaluate applicants' performance on exercises that replicate as closely as possible different responsibilities of the job. In job simulations, the employer presents applicants with realistic, job-related situations and documents their behaviors to help determine their qualifications for the job.

Job simulations can be high fidelity or low fidelity in their administration. High fidelity simulations use real materials, equipment, and/or scenarios to represent the duties of the job. For example, to assess applicants for a customer service position, an employer may sit applicants at an actual customer service work station with a telephone, computer, manual of standard operating procedures, and other work materials and have them role-play one or more scenarios. A trained assessor may call them with a customer service complaint and rate applicants on how they respond to the situation.

A low fidelity simulation, on the other hand, uses exercises that rely more on verbal or written instructions and do not require the applicant to actually perform or act out the situation. A low fidelity assessment for the customer service job may have the employer

present a verbal or written description of a scenario and ask applicants to choose from a set of responses which they feel is most appropriate, rather than ask them to role-play.

Overview of Advantages and Disadvantages

HR consultants report that the use of job simulations in the United States has risen steadily this decade, particularly in manufacturing, health care, and call center occupations.[22] A recent assessment trends report found that 83 percent of studied companies prefer using realistic assessments in their hiring process, and simulated assessments were the top "plan-to-use" assessments identified by U.S. companies (31 percent).[23] Job simulations can include, but are not limited to, work samples, situational judgment tests, assessment centers, and job tryout procedures.[24] We will provide more detailed descriptions of each of these assessments shortly, but first we will summarize some of the general advantages and disadvantages of job simulations as discussed by the professional literature.

Advantages

Higher Validity. Job simulation assessments, like those mentioned above, tend to have higher predictive validity than many other typical assessments, such as unstructured interviews, reference checks, work experience, and education level. This means that they should be better at predicting future job performance.

Better Job Fit. Job simulations tend to provide applicants with a realistic job preview, a result that can increase employee retention, productivity, and job satisfaction.[25] By exposing applicants to the types of events, scenarios, and challenges to be confronted on the job, they will be better able to determine if the job is well suited to their knowledge, skills, abilities, and interests. In hiring for its client services department, one company moved from typing tests that measured only speed and accuracy to simulations that measured those characteristics plus the quality of client interaction. The company found that its 90-day success rate went from in the 70s up to 90 percent.[26]

---

[22] Eilene Zimmerman, "Use of Job Simulations Rising Steadily," *Workforce Management,* http://www.workforce.com/section/06/feature/24/18/59/index.html, October 10, 2005.

[23] Sarah S. Fallaw and Andrew L. Solomonson of Previsor Talent Measurement, *2009 Global Assessment Trends Report*, p. 16.

[24] Some researchers and agency representatives considered situational interviews to be job simulations as well. Because MSPB has published a report on structured, situational interviews, we did not include them as part of this study. For more information on this type of assessment, see *The Federal Selection Interview: Unrealized Potential*, February 2003.

[25] For more information on realistic job previews, see U.S. Merit Systems Protection Board, "Realistic Job Preview," *Issues of Merit*, September 2008.

[26] Zimmerman, "Use of Job Simulations Rising Steadily."

Positive Applicant Perceptions. Because job simulations replicate the types of tasks performed in the actual job, applicants are more likely to view them as being fair and job-related. As previously noted, positive perceptions of the assessment can improve the result of the hiring process. In addition, studies have shown that applicants who view the assessment as being relevant to the job may work harder to perform better on the assessment.

Greater Degree of Fairness. Research generally supports the premise that job simulation assessments have lower rates of adverse impact, as well as a lower degree of exposure to discrimination lawsuits, based on the selection procedure. This not only saves the employer resources having to defend assessment tools in court, but it also reinforces applicants' positive perceptions of the process.

Overcome Faults of Other Assessments. Job simulations can help overcome the faults of non-simulated assessments, such as an applicant's distortion of his or her own abilities or "faking" competence. Also, job simulations rely on actual behavior rather than the verbal descriptions offered during an interview, which makes them more accurate measures of actual ability.[27]

Disadvantages

Cost. Job simulations can cost more. The more advanced simulations require added expertise to develop than other, less complicated assessments, and this expertise would result in higher developmental costs. Plus, they can require more staff and training to administer and assess the results. The use of job-related material or equipment may also add to the cost.

Limited Scope. While job simulations can be used to assess multiple competencies, a single simulation exercise will often, out of design necessity, focus on a limited number of tasks or duties performed on the job. If a work sample is used for a clerical position to measure an applicant's ability to type a memo, the assessment could miss other key aspects of the job, such as customer service skills. Therefore, job simulations are best used as one assessment in a multiple hurdle process, where a number of assessments are used in succession to measure different dimensions of the applicant's ability.

Not Suited to All Jobs. Many job simulations require the applicant to already have a certain level of knowledge, skills, or abilities to complete the assessment. Therefore, they may not be appropriate for some entry-level or generalist occupations that do not require pre-existing knowledge.

---

[27] Robert D. Gatewood and Hubert S. Feild, *Human Resource Selection, 5th edition*, Harcourt College Publishers, 2001, pp. 633-634.

Limitations of the Research. Some of the research reviewed for this study pointed to the fact that there were still many areas that need further research in terms of the utility of different job simulation methods. For instance, while some of the most referenced studies on work samples cite high levels of validity and low levels of adverse impact, some recent studies have questioned these research findings. We hope that bringing more attention to job simulation assessments will encourage the pursuit of additional research in this area.

Types of Job Simulations

Below, we provide basic information on typical job simulations that organizations may want to add to their assessment tool kit. Appendix A of this report contains a more in-depth look at each assessment. The discussions in this report are not meant to be exhaustive. Many books and research studies have looked at different aspects of work samples, situational judgment tests, assessment centers, and job tryouts. We are providing context and generalized research findings so that readers can become familiar with the advantages and disadvantages of these assessments and make educated decisions on what is appropriate to pursue for their organization.

WORK SAMPLE TEST

Definition	Examples	Advantages	Disadvantages
Evaluates applicants' job-related skills by having them perform actual activities or tasks that are physically or mentally similar to the duties they would perform on the job.	A work sample assessment for a welder may include an onsite welding exercise. Most applicants will already have the basic knowledge to perform the task. Therefore, test developers just need to ensure the equipment is available and develop measurements to evaluate the products produced. A work sample for a customer service job might involve role-play between the applicant and an "angry customer" (the assessor). The applicant may sit at a desk with a telephone, computer, and access to necessary records and standard operating procedures to be able to respond to the scenario. The caller would call with a complaint and evaluate the applicant's responses against pre-determined benchmarks.	They are generally viewed as having a high validity rating (.54), meaning that they are good predictors of future job performance. They are generally viewed as having low adverse impact and stand up well in selection discrimination litigation. Applicants are less able to "fake" proficiency. Applicants view them as fair because they can see the relationship to the job. Work sample tests provide applicants with a job preview to better inform their decision on whether they are a good fit for the job.	They generally do not measure aptitude or future potential. Their scope is limited to only the competencies needed for the specific activity carried out during the test. They are not very useful for tasks that take a long time to complete.

SITUATIONAL JUDGMENT TEST (SJT)

Definition	Examples	Advantages	Disadvantages
Presents applicants with a description of a work-related scenario and asks them to exercise their judgment by choosing, constructing, or evaluating alternative courses of action given the situation.	An applicant may be provided a paper and pencil test that contains several detailed scenarios. After reading through the scenarios, the applicant is asked to respond to the scenario. The following is a brief example of a possible SJT question: You have noticed that there is a lack of cooperation and trust among the employees in the organization you manage. This counter-productive behavior is adversely affecting both the quality and the quantity of the work produced by the employees. You would most likely… a. Inform the employees that failure to cooperate and work harmoniously will result in disciplinary action. b. Bring in someone to do some team building with the employees. c. Meet with the employees to discuss your observations and identify the cause of the interpersonal problems. d. Change some of the work assignments in order to attain a more cohesive and cooperative work group.*	They have moderately high levels of predictive validity (about .34). SJTs are easily administered and scored, especially when using video-based or digital technology to administer and record answers. Applicants can see the link between SJT scenarios and the work. As a result, they take the test seriously and try to do well. Also, the test provides a good preview of what the job will be like. SJTs are generally not susceptible to "practice effects" (i.e., improved performance on the assessment if the applicants complete it more than once). It is harder for dishonest applicants to remember and disclose to other applicants the longer SJT questions. SJTs have lower adverse impact than many written tests (e.g., cognitive ability tests), particularly when video or digital technology is used to administer the test.	SJTs help measure applicants' responses to ambiguous conditions. Good SJT scenarios contain a rich set of details, only some of which help determine the correct answer. If the test provides insufficient detail, the question does not fully test an applicant's ability to discern the relevant information and respond accordingly. If correct answers are too obvious, the SJT can become a test of what applicants know they "should" do on the job rather than what they would actually do. For some scenarios, four or five answer options might not cover the full range of actions, not capturing what the applicant might really do— right or wrong— on the job.

*Example taken from Dennis A. Joiner, "Assessment Center Trends," Presentation at IPMA-HR Assessment Council Conference on Personnel Assessment, June 23, 2004.

ASSESSMENT CENTER (AC)

Definition	Examples	Advantages	Disadvantages
"A process employing multiple techniques and multiple assessors to produce judgments regarding the extent to which a participant displays selected competencies."*	Assessment centers always use more than one exercise to measure the different applicant dimensions under review. Some of the more common exercises used in ACs include in-basket exercises (in which an applicant manages a set of tasks provided in a simulated "in-basket"), leaderless group discussions, structured interviews, and oral presentations.	They have moderately high validity ratings, meaning that they have been found to be good predictors of job performance, especially in terms of leadership abilities. They have low adverse impact and have actually been mandated by some courts to overcome selection bias in other instruments.** Applicants view them as fair because they can see the relationship between the exercises and the job. Assessment centers provide applicants with a job preview to better inform their decision on whether they are a good fit for the job.	The key disadvantage to assessment centers is that they are resource intensive. They take time and expertise to develop and organize. They require multiple, trained raters. They require space, equipment, and materials to administer. All of these resources amount to a fairly significant cost.

*International Task Force on Assessment Center Guidelines, *Guidelines and Ethical Considerations for Assessment Center Operations*, http://www.assessmentcenters.org/pdf/00guidelines.pdf, May 4, 2000.

**William C. Byham, "Section 2: Validity and Fairness," *What is an Assessment Center? The Assessment Center Method, Applications, and Technologies*, http://www.assessmentcenters.org/articles/whatisassess2.asp.

JOB TRYOUTS

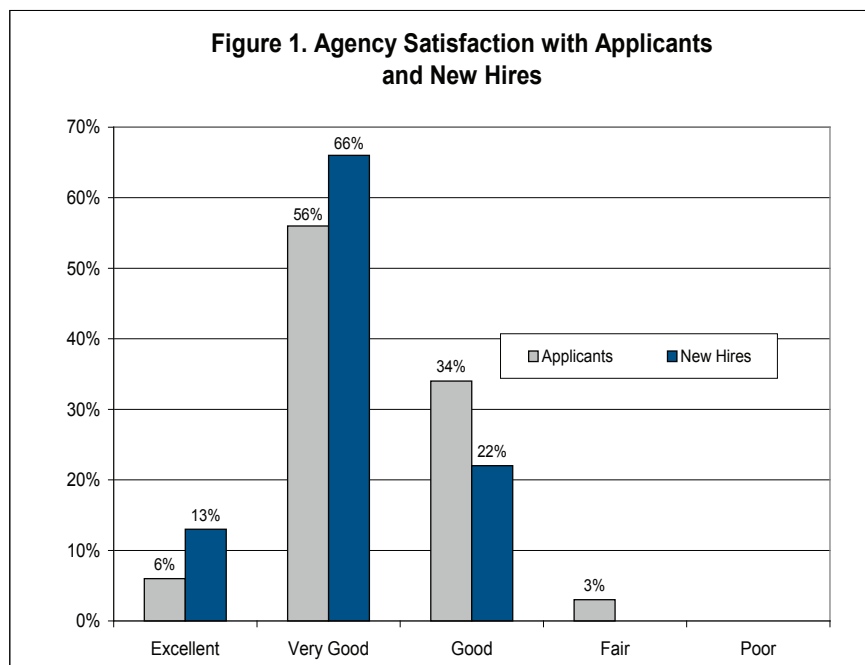| Definition	Examples	Advantages	Disadvantages
Applicants are hired with minimal screening and are given an evaluation period that determines if they meet the established levels of satisfactory job performance.	In the Federal Government, new employees are generally subject to a probationary or trial period. The purpose of the probationary or trial period is to provide the Government the opportunity to evaluate the employee on the job to determine if the appointment should become final at the end of the period. Until the probationary period has been completed, a probationer is still just an applicant for an appointment. If used correctly, the probationary or trial period could be considered a job tryout procedure.	Job tryouts have been found to have rather high levels of predictive validity (.44).\n\nThey are generally considered fair and objective when carried out appropriately.\n\nThey have high levels of face validity because the tryout period is on-the-job work.	There is a high cost to terminating lightly screened, low performers.\n\nThe key to ensuring an effective, predictive job tryout is following through on terminating employees who do not meet the standards for satisfactory performance. Research shows this often does not happen.*

*MSPB, *The Probationary Period: A Critical Assessment Opportunity*, p.7.

Job Simulations: How Federal Agencies Use Them

Federal agencies have a number of tools available to them to assist in evaluating candidate qualifications. Some are better than others in predicting applicants' future performance on the job. To use these tools effectively, agencies need to have adequate resources, expertise, and an understanding of employee assessment principles and practices. To understand how agencies are using job simulation assessments, we queried them about their typical assessment practices.

Nature of Assessment Programs

Figure 1. Agency Satisfaction with Applicants and New Hires

	Excellent	Very Good	Good	Fair	Poor
Applicants	6%	56%	34%	3%	
New Hires	13%	66%	22%		

Satisfaction with Applicants and New Hires. In the questionnaire we sent to Federal agencies, we asked about the quality of the agency's applicants and new hires. As seen in Figure 1, respondents were generally positive about the quality of the applicants for their organizations' vacancies, with 62 percent rating them as very good to excellent. Similarly, respondents were positive about the quality of the new employees selected by their organization. Over three-fourths rated them as very good to excellent. The numbers for new employees are in line with past research the MSPB has done with

supervisors having reported similarly high satisfaction with new hires.[28] While the levels of satisfaction depicted in Figure 1 are impressive, there is still room for improvement. For instance, it would be in the Government's interest if more agencies rated applicants and new hires in the "excellent" category. Furthermore, one-third of agencies rated applicants only as "good," and almost a quarter rated new employees as only "good."[29] In this regard, MSPB research has found that Federal agencies often use selection methods that are not very capable of distinguishing between the good candidates and the superstars.[30] It is possible that if agencies used better assessment methods, selecting officials would be even more impressed with applicants and new hires.

Roles and Responsibilities. Ideally, HR staffs and selecting officials work closely together to identify, develop, and administer the tools used to assess applicant qualifications because HR staffs are often the assessment and selection experts in the agency and selecting officials are the subject matter experts (SMEs). We asked questionnaire respondents a number of questions related to the roles and responsibilities of HR staffs and selecting officials in the assessment process.

First, we asked what level of knowledge HR staffs and selecting officials have about assessment tools. Keeping in mind that the questionnaires were generally completed by HR staff members, the results indicate that HR staffs are more knowledgeable about assessment tools than selecting officials. Eighty-four percent said HR was knowledgeable or very knowledgeable while 44 percent rated selecting officials the same.

Second, we asked who is typically responsible for deciding what assessment tools will be used. This is an area in which we would like, and expect, to see shared responsibility among the different groups involved in the selection process, and our expectation was borne out by the data. Few agencies relied on only one group to carry out this responsibility. Generally, HR had the highest level of responsibility for deciding what assessment tools would be used to measure applicant qualifications (79 percent of respondents said HR is typically responsible for deciding what assessment tools will be used), but there was some sharing of this role with selecting officials (68 percent) and subject matter experts and/or line staff (52 percent).[31] A few respondents noted that they have Industrial/Organizational (I/O) psychologists or other testing staff who participate (18 percent). Finally, almost one-quarter (24 percent) reported that OPM typically assists in this role.

---

[28] For instance, see U.S. Merit Systems Protection Board, *The Probationary Period: A Critical Assessment Opportunity*, 2005; *Building a High-Quality Workforce: The Federal Career Intern Program*, 2005; *Accomplishing Our Mission: Results of the Merit Principles Survey 2005*, 2007; and *In Search of Highly Skilled Professionals: A Study on the Hiring of Upper-Level Employees from Outside the Federal Government*, 2007.

[29] When we refer to "agencies" that responded to the interrogatory, this includes the total 35 responses received from the Department level and major sub-component level organizations.
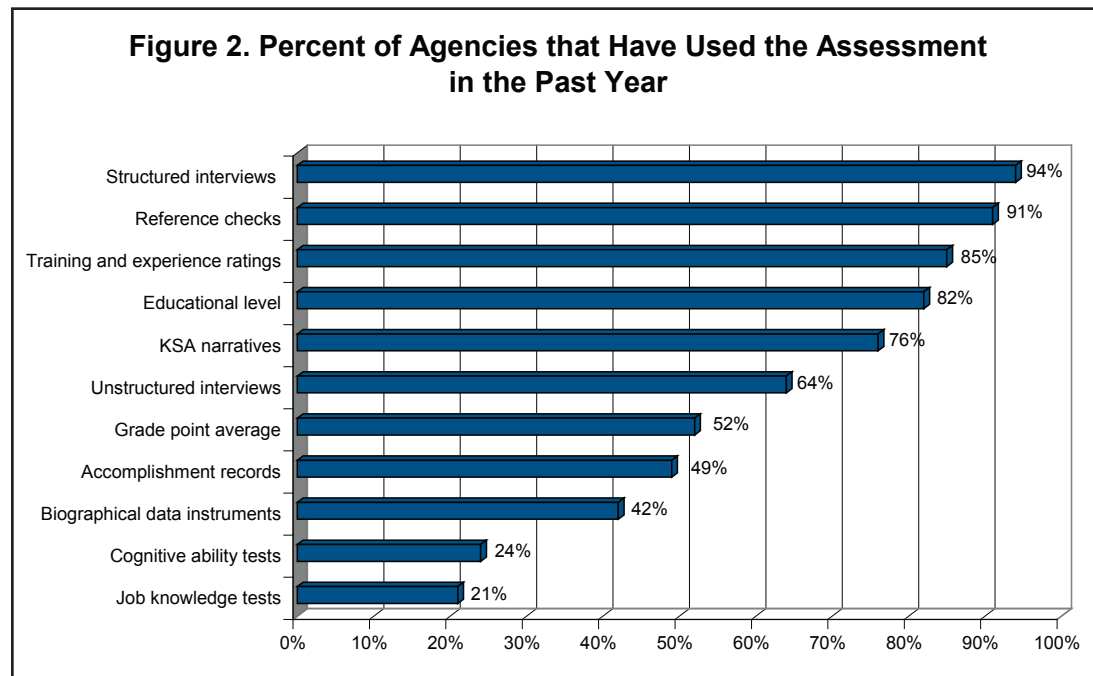
[30] MSPB, *Reforming Federal Hiring: Beyond Faster and Cheaper*.

[31] Because respondents were instructed to select "all that apply," percentages do not total 100.

Selecting officials and SMEs should not be expected to be experts in employee assessment. Therefore, when agency representatives noted that these groups have a role in deciding what assessments to use, we asked how much help they receive from the HR office in making these decisions. Many respondents reported that the HR office consults regularly with selecting officials about their options (48 percent) or provides guidance when contacted by selecting officials who have questions (48 percent).

Types of Assessments Used. We then asked agencies what types of non-simulation assessments they have used in the past year. As shown in Figure 2, the top instrument respondents reported using was the structured interview. This is a promising trend. Structured interviews have received considerable attention in recent years, with both the MSPB and OPM encouraging agencies to use them in place of unstructured interviews because research has found them to be better predictors of future performance.[32]

Reference checks were the second most used assessment. As the MSPB reported in a 2005 study, *Reference Checking in Federal Hiring: Making the Call*, reference checks can be a valuable tool, particularly in helping employers validate information provided by applicants. Given that the next most prevalent assessments are those that ask applicants to self-report their achievements (i.e., through itemization of education and experience; KSA narratives; and unstructured interviews) validation through reference checks becomes that much more important. Job simulations can provide an additional layer of validation.

Figure 2. Percent of Agencies that Have Used the Assessment in the Past Year

Assessment	Percent
Structured interviews	94%
Reference checks	91%
Training and experience ratings	85%
Educational level	82%
KSA narratives	76%
Unstructured interviews	64%
Grade point average	52%
Accomplishment records	49%
Biographical data instruments	42%
Cognitive ability tests	24%
Job knowledge tests	21%

Note: See Appendix B, Question 6 for assessment definitions.

---

[32] Structured interviews employ objective, pre-determined rules for eliciting, observing, and evaluating interview responses. Unstructured interviews include questions that may be unplanned or vary across interviews, and the results are analyzed and applied subjectively.

Use of Job Simulation Assessments

In looking at the assessments being used most by those agencies responding to our questionnaire, it becomes apparent that the assessments rely heavily on self-reported information, such as interviews, training and experience ratings, and KSA narratives. Reference checks can help validate some of that information, but they have their limitations, given that prior employers may be careful in what they divulge for fear of litigation.

As previously discussed, job simulations put applicants in realistic, job-related situations and document their response or behavior to determine their qualifications for the job. Because of this advantage, we explore the extent to which agencies are currently using job simulations, the barriers they face in using them, and the implementation procedures for those who do use them.

Extent to which Job Simulations Are Used. Fewer than half of the organizations responding to our questionnaire (16 respondents) used a job simulation in the past year. Based on prior research and anecdotal information, we were not surprised by this somewhat low number. But we need to be cautious and try not to read too much into the responses from only 16 organizations. Their observations can provide us with some valuable context in considering how effective the use of simulations can be, but their experiences are only a glimpse and should not be generalized to all organizations that use this type of assessment.

Of those that did use job simulations, the work sample was the most popular, as shown in Table 1. After work samples, the next most popular job simulation was the assessment center, which is more complex to design and more resource intensive to administer.

No organizations reported using the job tryout.[33] This finding was not surprising given that job tryouts often rely on minimal applicant screening and assume that the organization can easily separate employees once they are on-board. A job tryout simulation may be appropriate for a limited category of Federal jobs, but it would require the use of tax payer resources to hire applicants after limited screening, put them on the payroll, train them, and separate them if they fail. Therefore, agencies should be cautious in using this type of assessment and only when there is a strong business case to do so.

---

[33] New Federal employees are generally subject to a probationary or trial period during which agencies evaluate the employee to determine if the appointment should become final at the end of the period. If used correctly, the probationary or trial period could be considered a job tryout procedure. Because MSPB's 2005 probationary period report looked at this assessment in-depth, we specifically asked respondents not to include their use of the probationary period as a job tryout.

Table 1. Types of Job Simulations Agencies Used in Past Year	
Simulation	**# of Agencies**
Work Sample	9
Assessment Center	7
Situational Judgment Test	5
Job Tryouts	0

Most organizations that used job simulations used them for less than 10 percent of their vacancies in the past year, as shown in Table 2. These agencies generally have a variety of mission-critical occupations (only some of which are amenable to job simulation assessments) or are using simulations for specialty areas (such as leadership positions). The agencies that use them for more than 50 percent of their vacancies are fairly homogenous agencies that do volume hiring into one or two entry-level, mission-critical occupations, such as law enforcement or foreign service officers.

Table 2. Percent of Vacancies for which Agencies Used Job Simulations in Past Year	
Percent of Vacancies	**Number of Agencies**
Less than 10	10
11-25	2
26-50	1
51-75	1
More than 75	2

Deciding to Use Job Simulations. We asked agencies how they determine whether to use simulations, and they cited a number of factors that influence this decision. First, agencies noted that the results of the job analysis helped determine whether the position would be amenable to a simulation. Once they identified the key tasks of the job and the competencies and skills needed to carry out those tasks, it became easier to see what types of assessments would best measure those competencies.

Budget was another driving force behind the decision. Because simulations generally require more resources to develop and/or administer, agencies have to ensure they have the budget necessary to support the assessment.

Another factor appears to be the relationship between selecting officials and HR staffs. When a consultative relationship exists, there is more communication and discussion of what would be the best assessments to use for the position/occupation. In addition, such a relationship seems to foster more openness to using non-traditional types of assessments.

Jobs and Skills for which Simulations Are Used. We asked agencies to identify the occupations for which they used job simulations and found that agencies used them for a wide variety of occupations. Some agencies used them for support positions such as clerical, human resources, or information technology positions. Other agencies used them for mission-critical occupations, such as those in health care or law enforcement. Some reported using them for all positions above a certain grade level, and others use them only for leadership positions.

The types of competencies for which agencies used simulations were a bit easier to categorize, but still were varied across organizations. Agencies used them to measure competencies such as technical skills, customer service skills, communication, problem solving, decision-making, leadership, interpersonal skills, achieving results, and flexibility.

How Job Simulations Are Used. Over half of the respondents who reported that their organizations had used simulations in the past year indicated that they used them as one hurdle in a multiple hurdle process, as opposed to a stand-alone assessment. A multiple hurdle approach uses a set of valid assessment procedures successively to manage the candidate pool and narrow the field of qualified candidates.[34] Research has found that using good assessment procedures in succession can improve the ability of the assessments to predict job performance, adding to the quality and cost benefit of the process.

The Federal assessment process can be broken into two distinct steps. First, applicants are assessed to ensure that they meet the minimum requirements to perform satisfactorily on the job—what the Government refers to as minimum qualifications. Second, for those who meet minimum qualifications, their relative qualifications are assessed to make distinctions among the qualified applicants. Different assessment tools can be used to carry out these steps. All of our respondents reported using simulations for assessing relative qualifications, not for making minimum qualification determinations. Using simulations in this way is understandable. Because they are generally more resource intensive to administer, job simulations can cost more per applicant to administer. Therefore, using them to distinguish among the smaller pool of the most qualified applicants is a more judicious use of resources.

Implementing Job Simulations. There are some key steps to implementing job simulation assessments: (1) the test developer is selected; if he or she lacks expertise, arrangements should be made for an expert to assist; (2) the test is developed; (3) the test administrator is selected and trained; (4) the test is administered; and (5) the test results are assessed. We asked agencies how they carry out these steps. We found that selecting officials have a lot of responsibility, but not necessarily a lot of assistance.

---

[34] For more information on the multiple hurdle approach, see U.S. Merit Systems Protection Board, "Fewer Hoops, Higher Hurdles," *Issues of Merit*, September 2002; "How Many is Multiple," *Issues of Merit*, July 2003; and *Identifying Talent through Technology*, 2004, pp. 75-77.

Almost half of the organizations that use job simulations indicated that the selecting official is typically responsible for developing the test. In these cases, the selecting official received no assistance from the HR staff or I/O psychologists, though some did receive assistance from a contractor or subject matter expert. This occurred despite the fact that few of the organizations that used simulations indicated that selecting officials are knowledgeable or very knowledgeable about the assessment tools available for measuring applicant qualifications.

It is of concern that assessments are being developed and administered by individuals who may not have the expertise necessary to develop, administer, assess, and defend the tests. However, we must keep in mind that this was reported by only a few organizations and cannot be generalized to all Federal agencies and therefore do not provide enough evidence of a systemic problem. Nonetheless, we note that if selecting officials have significant responsibility for developing assessments, it is especially important for them to have assistance from assessment experts. Without the proper expertise, simulations could lead to erroneous or less than optimal selections and/or adverse impact.

In terms of administering simulations, almost half of the organizations that used job simulations reported that the selecting official was typically responsible for administering the test. SMEs were involved in administering the simulations in about a third of the organizations.

The fact that selecting officials and SMEs have a lot of responsibility in administering assessments is not in and of itself a problem, as long as those participants receive the proper training to carry out their responsibilities. Simulation assessments can require significant training to ensure that administrators understand their role in the assessment process. However, as Table 3 demonstrates, some organizations may not be providing sufficient training in the different phases of the assessment process. For instance, the two organizations that reported that the selecting officials administer the simulations with no assistance also reported that administrators receive no training.

Table 3. Training Provided for Administering Simulations	
Number of Days	**Number of Agencies**
5 or more days	2
2-4 days	2
1 day or less	3
No training	2
Don't know	6

Results of Job Simulations. Most of the organizations that used job simulations in the past year believed that the simulation tests increased the quality of the agency's selections. Unfortunately, few organizations reported actually measuring the impact. The organizations that did measure the effectiveness of these assessments generally

used surveys of HR and selecting officials, measured the rate of successful probationary employees, and/or conducted validation studies.

Barriers to Using Job Simulations. To help gain a better understanding of why few agencies are using job simulations, we asked all of our respondents what obstacles they faced in using job simulations or what factors kept them from using these assessments. Not surprising, the top barriers cited were related to a lack of resources and knowledge rather than amenability, as shown in Table 4. More agencies might be inclined, therefore, to use job simulations if they had more knowledge or training in developing and administering these assessments and more resources to devote to these steps.

TABLE 4. BARRIERS TO USING JOB SIMULATIONS	
Barrier	**Percent of All Respondents**
The agency does not have the time required to develop assessments.	55
Selecting officials do not have enough knowledge about these types of assessments to determine if they would be beneficial.	55
The HR staff does not have enough knowledge about these types of assessments to determine if they would be beneficial.	52
The agency does not have the resources to develop and/or administer them.	48
The agency does not have the expertise to develop and/or administer them.	45
The agency does not have the time required to train those who administer the assessments.	41
Other, such as lack of upper level support; fear of litigation for improperly administered or invalid assessments; potential increase in time to fill the vacancy; and lack of resources for nationwide onsite assessments one agency felt could discourage diversity by biasing the process toward local applicants.	35
Agency uses other assessments they believe better differentiate among candidates.	24
The agency's jobs do not lend themselves well to this type of assessment.	14

FORMULATING AN ASSESSMENT STRATEGY

As explained earlier, the MSPB is responsible for conducting special studies of Federal HR practices, policies, and procedures to ensure that they adhere to the merit system principles and do not result in prohibited personnel practices. Given that mission, we generally write reports that contain definitive recommendations about what agencies should or should not do in carrying out their HR responsibilities. For instance, we have recommended that agencies use a structured format to conduct job interviews and reference checks because those are strategies from which all organizations using those assessments can benefit. We have also recommended that agencies do a better job of using the probationary period to terminate employees who cannot perform at the expected level.

This report is a little different. We are not advocating that all agencies use job simulation assessments, nor are we providing specific recommendations about what assessment methods to use. Instead, we recommend an assessment strategy that agencies should consider adopting. The steps in the strategy will assist agencies in making informed decisions about the type of assessments that will help ensure that they are hiring on the basis of relative ability, knowledge, and skills, in adherence to the merit system principles.

There are many factors to consider when choosing which type of assessment to use to evaluate applicant qualifications, and job simulations may or may not be appropriate. The strategy presented below will help the agency determine what assessments are appropriate to the given situation—whether it is a job simulation or a non-simulation assessment. The strategy comes from a review of professional literature on assessment and selection, lessons learned from the agencies that responded to MSPB's questionnaire, agency interviews, and interviews with OPM officials who have experience working with agencies to improve their assessments.

Some of the steps in this strategy will require agencies to devote more time and resources to the hiring process than they currently do. With the current emphasis on improving the timeliness of the Federal hiring process, these steps may appear counterintuitive. However, they can help improve the quality of agency selections and, therefore, the ability of agencies to effectively achieve their missions. In addition, much of the necessary work can be done before a position becomes vacant, thereby not impacting the timeliness of individual HR actions.

(1) Conduct a Job Analysis

OPM defines job analysis as a "systematic procedure for gathering, documenting, and analyzing information about the content, context, and requirements of the job."[35] The Code of Federal Regulations requires that "[e]ach employment practice of the Federal Government generally, and of individual agencies, shall be based on a job analysis..."[36] The Uniform Employee Selection Guidelines call job analysis essential to the professional practice of employee selection.[37]

The typical job analysis will determine the relative importance of the job's duties and tasks and identify the KSAs necessary to perform them. In addition, the job analysis will identify other factors that impact the ability to carry out the job's duties, such as tools and technology, work environment and relationships, and training and licensure requirements.[38]

Ultimately, the job analysis demonstrates the relationship between the duties and tasks performed on the job, and the competencies and KSAs required to perform the tasks. This information can then be used to inform a number of workforce management decisions, including the development of the position description, compensation, recruitment strategies, onboarding strategies, training and development needs, and deciding on the assessment methods to use. For the purpose of this report, the job analysis should be used to identify job content, the competencies required, and the assessment methods to use, as well as establish minimum standards for assessor training.

A job analysis is carried out by a "job analyst." This person must know the job well and/or have extensive training in carrying out the analysis. If the person has only one of these two attributes, a team approach may be used. In fact, the most effective job analyses are often performed through cooperative ventures in which a subject matter expert provides the job knowledge and the HR staff or I/O psychologist translates that information into usable formats and applications.

Many methods can be used to conduct a job analysis. These include direct observation of employees, incumbent and/or supervisor interviews, structured questionnaires, expert panels, task inventories, checklists, and work logs. The critical incident technique lends

---

[35] Office of Personnel Management, *Delegated Examining Unit Operations Handbook*, 2007.

[36] 5 C.F.R § 300.103

[37] *Uniform Guidelines on Employee Selection Procedures*, http://www.uniformguidelines.com/uniformguidelines.html#88.

[38] OPM's *Delegated Examining Operations Handbook 2007* provides detailed guidance and tools on how to conduct an appropriate job analysis. The handbook can be found at http://www.opm.gov/deu/Handbook_2007/DEO_Handbook.pdf.

itself to developing scenarios for many simulation exercises.[39] A typical method of job analysis would be to collect information from a small group of job incumbents through interviews or focus groups. The resulting duties, tasks, and competencies are then distributed to a larger group of incumbents through a questionnaire. Those incumbents rate the importance and frequency of each duty and task and the importance and need of the related KSAs. The job analyst would then use the information to prepare a job description.

While the method may differ, the job analysis should ultimately collect information regarding:[40]

- The basic *duties and tasks* of the position, including frequency, importance, duration, effort, skill, complexity, equipment, and standards.
- The *competencies and KSAs* required to perform the job. Job analyses typically state the minimal requirements for the job, though some experts recommend identifying the requirements necessary to perform that job at a higher level and focusing on outcomes rather than duties and tasks. In particular, the critical incident technique is a process that identifies behavior or events that make a significant contribution to an activity, such as the KSAs that distinguish satisfactory performance from unsatisfactory performance.
- The *environment* in which the duties are performed, particularly any physical requirements that may pose obstacles to accomplishing the tasks, such as any unpleasant or risky conditions.
- The *tools and equipment* necessary to carry out the duties and a description of how they are used and what skill level they require.
- The *relationships* required to carry out the job, such as supervision that is given or received or relationships with internal or external stakeholders.

The results of the job analysis will identify the critical tasks of the job and the skills necessary to carry out those tasks. Once these have been established, the organization can use that information to identify which assessments would best measure the identified KSAs, including job simulations. For instance, if a job requires already established knowledge of particular tools or equipment, a work sample could help identify those candidates who already possess that knowledge. In another example, the job analysis may identify the critical incidents incumbents may find themselves in that require a high level of judgment. Those situations could be used as the basis of a situational judgment test.

---

[39] For more information on this technique, see John C. Flanagan, "The Critical Incident Technique," *Psychological Bulletin*, Vol. 51, No. 4, July 1954.

[40] HR-guide.com, "Job Analysis: Overview," *Job Analysis Internet Guide*, 1999, http://www.hr-guide.com/data/G000.htm.

The results of the questionnaires and agency interviews we conducted demonstrate that the agencies that effectively use job simulations base them on up-to-date job analyses. Some of them update the analyses on a scheduled basis, some as necessary, but all indicated that the job analysis is an important step in identifying how to best use the simulations.

Anecdotal observations from some of the agencies we interviewed, including OPM, indicated that the lack of an up-to-date job analysis is a key barrier to agencies' ability to effectively establish a new assessment process, such as simulation assessments. A job analysis can take a significant amount of resources to complete; however, it provides the roadmap for the assessment process. Those involved in assessment and selection need to understand the job as it currently stands before they can determine how to fill it.

By using an old job analysis, an organization may miss some key aspects of the job that it needs to consider when hiring a new person. Technology is redefining the work we do and how we perform it. Manual work is declining while knowledge-based work is increasing. These trends could easily impact the competencies needed to effectively carry out the duties of a position. Therefore, using an out-dated job analysis could result in hiring the wrong person with the wrong skills. Ensuring that job analyses are current can reduce the resources it will take later to try to train that person for the job or to have to re-recruit for the position.

(2) Identify the Assessments

Following the job analysis, the agency will know what competencies and KSAs should be tested. The next step is to identify which assessments to use. As discussed earlier, one of the key barriers to using job simulations was lack of knowledge about them. Therefore, this is where organizations need to do their homework and educate the HR staff, selecting officials, and other agency leaders about assessments, including job simulations. The number of assessment tools available to organizations is extensive, and they all have their advantages and disadvantages. They differ in a number of areas, such as: [41]

- Purpose— job selection, promotion, development;
- What they are designed to measure— competencies, work styles, values;
- What they are designed to predict— job performance, managerial potential;
- Format— paper and pencil, video or online, role-play;
- Assessment characteristics— level of standardization, objectivity, and quantifiability; and
- Type of administration— individual versus group.

---

[41] U.S. Department of Labor, Employment and Training Administration, *Testing and Assessment: An Employer's Guide to Good Practices*, http://www.onetcenter.org/dl_files/empTestAsse.pdf.

Helping Choose an Assessment

There are a number of resources available to agencies that discuss the different types of assessments and what to consider when making this decision. Of particular note, OPM has developed the Assessment Decision Guide and Assessment Decision Tool, both of which are available on OPM's Web site at http://apps.opm.gov/ADT/. The Decision Guide summarizes a number of popular assessment methods and discusses the different dimensions of each assessment, including the validity, reliability, resources needed, and time to develop. The Decision Tool is an interactive system that allows the user to input information regarding the hiring situation and receive customized information on which assessment tools may meet the specific hiring need. The user can input details about the hiring situation, including the competencies to be assessed, or ask for information about a particular assessment tool. Both cover simulation assessments as well as non-simulation assessments.

When researching the assessment tools available, there are a number of factors to consider that will help identify which tools are best for the specific hiring situation, such as:

Available resources. As noted earlier, one of the key barriers to using job simulations was the lack of resources and expertise available to develop and administer these types of assessments. The resources needed include the time, money, staff, and expertise available to develop and administer the assessments, including the training and time requirements for each person charged with assessing the applicants. An organization that has an ample budget, dedicated assessment staff, and supportive leadership will generally have the resources necessary to carry out some of the better assessments. Many organizations lack these advantages and may need to find ways to work with limited resources. For instance, if an agency does not have the expertise to develop certain assessments, it may be able to find the money to hire a contractor or to partner with an agency with similar hiring needs to develop and administer the assessment.

Job simulation assessments require different levels of resources. A simple work sample could be fairly easy to develop and administer. For instance, an organization that identifies writing ability as a key competency could work with HR to develop a writing exercise that is rated by SMEs using standard benchmarks. This could require few resources and little expertise. On the other hand, an organization that wants to measure management potential through an assessment center will require extensive resource commitments in terms of obtaining the expertise to develop and validate the different exercises, training the assessors and providing them the time they need to carry out their responsibilities, and setting up the necessary logistics, materials, and equipment.

MSPB's studies have noted that many agencies do not have the expertise and resources necessary to carry out high-quality assessment programs. OPM does provide assistance, but because of its own budget cutbacks, its services are available mostly on a reimbursable

basis. MSPB's reports have long recommended that Congress provide OPM the funding necessary to assist agencies in developing assessment tools that create the best return on investment for the Government. Even if OPM were to focus on Governmentwide occupations, high-volume occupations, and/or severe shortage occupations, the Government would still benefit from the economies of scale achieved.

When making decisions on how to allocate resources, agencies need to weigh the fact that it may be more costly in the long run to make poor hiring decisions than to spend the money to make good ones. As mentioned earlier, research indicates that the actual cost of hiring the wrong person for a job—including wasted salary, benefits, training costs, and hiring time—can be up to three times the employee's salary.[42] Poor selections can also result in lost productivity because of the time it takes a manager to carry out an adverse action.[43] In addition, poor selection can lead to increased turnover and absenteeism, as well as resentment among employees who often must complete the work which someone else has stopped doing.[44]

On the other hand, different lines of research suggest that using good assessment practices can help hire better people, which can lead to increased productivity and revenue gains. In terms of hiring better people by using better assessment methods, Taylor and Russell conducted a notable study in I/O psychology that demonstrates that increasing the predictive validity of an assessment will increase the percent of new hires who will perform satisfactorily on the job.[45]

Studies have shown that making selections based on the candidate's ability to do the work can lead to higher organizational performance and increased financial benefits.[46] For instance, utility analysis has been used to estimate the cost benefits of changing a selection procedure. In Schmidt and Hunter's study on validity and utility, they looked at the variability of employee performance and found that a "superior skilled worker" produces 32 percent more output than an average worker, and a superior manager or professional produces 48 percent more output.[47]

---

[42] Corporate Leadership Council, Literature Review, "Employee Selection Tests," Catalog No. 070-198-213, Washington, DC, Mar. 1998, p. 2.

[43] Laurence Karsh, "The Hidden Costs of Poor People Management," *Inc.com*, Dec. 2004, downloaded on June 14, 2005, from http://pf.inc.com/articles/2004/12/karsh.html.

[44] The Partnership for Public Service, *Asking the Wrong Questions: A Look at How the Federal Government Assesses and Select Its Workforce*, Washington, DC, October 2004, p. 3.

[45] Taylor and Russell, "The Relationship of Validity Coefficients to the Practical Effectiveness of Tests in Selection," pp. 565-578. The Taylor-Russell tables calculate the increase in performance using the base rate of success, the level of assessment validity, and the selection ratio.

[46] Terpstra and Rozell, "The Relationship of Staffing Practices to Organizational Level Measures of Performance," 1993, and Watson Wyatt Worldwide, *The Human Capital Index: Linking Human Capital and Shareholder Value*, 1999.

[47] Schmidt and Hunter, "The Validity and Utility of Selection Methods in Personnel Psychology," p. 263.

Research in Economic Value-Added also suggests that employing a superior, rather than average, performer will result in huge economic and productivity gains.[48] Furthermore, Schmidt and Hunter's analysis of selection utility literature found that improving hiring methods typically leads to large gains in economic value and the gains are directly proportional to the size of the increase in predictive validity.[49] They went so far as to say that "in economic terms, the gains from increasing the validity of hiring methods can amount over time to literally millions of dollars. However…by using selection methods with low validity, an organization can lose millions of dollars in reduced production."[50]

To reduce the costs associated with bad selections and improve the organization's ability to carry out its mission, it is important to employ good assessment strategies that help identify the best candidates for the job. While research indicates that using good assessment practices may increase the bottom line, the agency's leadership ultimately decides how important assessment tools are compared to other priorities. To help ensure that agency leaders understand the impact of good assessment practices on agency results, HR staffs need to build the business case for making assessment a priority and investing the necessary resources to it, even when the best choices may be the more costly ones.

Required skills. Some assessments are better than others for testing specific skills. For instance, because work samples often require that applicants already possess some of the knowledge, skills, or abilities necessary for the job, they may not be appropriate for measuring potential but may be good for a later phase of testing. A situational judgment test is an effective measure of social skills but is less likely to be effective at evaluating technical or motor skills. An assessment center can be used to test multiple competencies, but the resource requirements make them a better choice for mission-critical and/or high-volume occupations rather than for single vacancies. So, when choosing assessment methods, the organization must identify what it wants to measure and determine which assessment best evaluates that competency, knowledge, skill, or ability. Some of the time, a job simulation will be appropriate; other times, a non-simulation test will be best.

Assessment quality. As discussed earlier in this report, there are several factors that serve as the foundation for high-quality assessments. The assessment should be:

- Reliable— produce consistent results;
- Valid— measure what it was intended to measure; and
- Fair— be appropriate for the targeted applicant pool, be viewed by the applicant as unbiased, and reduce adverse impact.

---

[48] Spencer, "The Economic Value of Emotional Intelligence Competencies and EIC-Based HR Programs," 2001.

[49] Schmidt and Hunter, "The Validity and Utility of Selection Methods in Personnel Psychology," p. 263.

[50] Schmidt and Hunter, "The Validity and Utility of Selection Methods in Personnel Psychology," p. 273.

When developed and administered properly by qualified testing professionals, job simulations can meet all of these criteria.

Number and placement of assessments. There are no rules about how many assessments an organization should use in the hiring process. It depends on the job, the competencies necessary to carry out the job, and the resources available.

MSPB's reports have long advocated a multiple hurdle approach to assessment. This type of approach can help avoid poor selection by using a set of relatively valid assessment procedures successively to manage the candidate pool and narrow the field of qualified candidates. Research has demonstrated that using good assessment procedures in succession can improve the ability of the assessments to predict job performance, adding to the quality and cost effectiveness of the process.[51]

Each assessment tool should be used to complement, not replicate, one another. They should be implemented in a timely manner so that good applicants do not drop out of the process. Ideally, assessment procedures should be selected and sequenced based on cost and benefit. For instance, tools that are less costly to administer should be used in the beginning stage of the assessment process when the candidate pool is largest. A situational judgment test could serve well as an early hurdle because it is effective at narrowing the candidate field while requiring fewer resources to administer and score. Performance-based simulations, such as assessment centers or work samples, might be more effectively used later in the assessment process because they can take more resources to administer to large groups of applicants though they make better distinctions among applicant qualifications. This type of multiple hurdle approach can help agencies effectively manage their resources while increasing the quality of their selection process.

Number of expected applicants. To get the best return on investment for the assessment, the organization should consider how many people it expects will apply for the position. As mentioned above, we recommend selecting and sequencing assessment procedures based on cost and benefit. If the organization anticipates a large applicant pool, then a low-cost screening assessment would be appropriate to narrow the applicant field. However, if it is a hard-to-fill position that generally receives few applicants, then it makes sense to go straight to the higher cost but more precise assessment tool—saving time, money, and applicant burden by using fewer assessments.

Applicant burden. Two of the key complaints about the Federal hiring process are its complexity and its length. Assessment procedures can add to both of these issues. Therefore, it is important to ensure that the assessment methods the organization chooses are implemented in a thorough but timely manner. In terms of the multiple hurdle process, while successive hurdles can improve the quality of the hiring process, *excessive* hurdles can impede the effectiveness of the hiring process. Too many hurdles can result

---

[51] Schmidt and Hunter, "The Validity and Utility of Selection Methods in Personnel Psychology," p. 265.

in applicant attrition and do little to improve hiring decisions. An excessive number of assessments can also deter qualified applicants from applying because of the applicant burden.

Applicant burden is an issue to consider when determining whether job simulations are appropriate. Job simulations, as well as other assessment tools such as in-person interviews, can present a significant burden to applicants if, for example, they require physical attendance at the test administration site. Applicants may then have to use their time and money to get to the location. Having a number of sites available or paying for applicant travel could help decrease the applicant burden but it would increase the agency's administration costs.

There are tools available to help organizations reduce applicant burden, as well as agency cost, when using job simulations. Two such tools are described below:

Video/Digital Technology. The use of video or digital technology has been increasing over the past decade. Video or digital technology can be used in test administration to present the applicant with the test instructions, contextual information, practice scenarios, and/or the actual test scenarios, questions, or items. Video and digital technology can also be used to capture the applicants' answers. Assessors can then evaluate candidates' responses at a later, more convenient time and location or computers can be used to score multiple-choice items. The advantages to this approach are greater standardization of the assessment process, a more efficient and cost-effective administration and evaluation process, and a potentially higher degree of fidelity of the assessment itself. The downsides to the approach include less interpersonal contact with the applicant, inconclusive research on the validity of video/digital administration versus traditional paper and pencil administration, and the potential for the technology to fail during the test.

Unproctored Internet Testing. Unproctored Internet testing (UIT) is a more controversial procedure in assessment circles. Generally, when administering a job simulation (as well as other tests) in proctored testing, the applicant appears at a location, checks in with the administrator, presents identification, and is observed while completing the test. These procedures help assure the security and consistency of the testing procedures. In UIT, the test is placed (securely or non-securely) on an Internet site, and applicants complete the test in the setting of their choice (at home, at work, at a local library). Organizations in the private and Federal sectors have become more interested in this approach because it can save considerable resources in terms of the time, money, logistics, and staff needed for establishing testing sites.[52] In addition, UIT can help reach and test a larger applicant pool, reduces applicant

---

[52] MSPB did not ask about the use of UIT on the interrogatory sent to agencies, and therefore has no specific data on the extent to which it is used in the Federal Government.

travel costs, standardizes the information provided to applicants, limits human administrator errors, and provides a cutting edge image to the applicants. A recent survey found that nearly 60 percent of U.S. companies surveyed permit the use of remote testing (for example, from their home).[53]

Nonetheless, unproctored Internet testing has serious drawbacks, including:

- Inability to establish that a third party is not taking the test for the applicant;
- Potential for the applicant to use materials not permitted by the testing instructions;
- Test content exposure;
- Nonstandardized testing conditions;
- Potential for technical problems; and,
- Some researchers' questioning of the validity of UITs.[54]

Users of UIT have proposed that there are ways to negate many of these drawbacks. For instance, the organization can provide protective procedures, to include using warnings, the right to re-test, applicant identifiers and tracking systems, and a multiple hurdle process that validates the UIT scores. As UIT matures, assessments are likely to use more sophisticated computer adaptive tests and high fidelity simulations that make it more difficult to cheat.

When deciding whether to use unproctored Internet tests, organizations should weigh the advantages and disadvantages thoroughly. While UITs may make the hiring process cheaper and faster, they may undermine the quality of the selections made if sufficient protections are not put into place.

(3) Develop Testing Procedures

Once the organization knows what competencies need to be tested and what assessments it will use to test those competencies, it needs to develop the testing procedures.[55] The first step is to identify who will develop them. Those chosen to develop the assessments should be qualified test developers who have experience not only developing tests, but also validating and defending them legally when necessary. It is not a good idea, therefore, to have selecting officials developing the tests on their own. Instead, as

---

[53] Fallaw and Solomonson of Previsor Talent Measurement, *2009 Global Assessment Trends Report*, p. 12.

[54] For instance, see how several panelists differ in their views of unproctored Internet testing, its utility, and its validity: Nancy T. Tippins, James Beaty, Fritz Drasgow, Wade M. Gibson, Kenneth Pearlman, Daniel O. Segall, and William Shepherd, "Unproctored Internet Testing in Employment Settings," *Personnel Psychology*, Spring 2006, Vol. 59, No. 1, p. 189.

[55] See 5 C.F.R § 300, Subpart A—Employment Practices, for the principles governing Federal employment practices that affect the recruitment, measurement, ranking, and selection of individuals for initial appointment.

reported by some of our questionnaire respondents, a consultative relationship among the selecting official, test developer(s), and HR staff does appear to be a better approach to developing nontraditional types of assessments such as job simulations. For that reason, we encourage collaboration throughout the entire test development process.

Some organizations will have the necessary in-house expertise, but many will not. The latter organizations may want to consider procuring the expertise from an outside source, such as OPM or a private contractor. Keep in mind, however, that the agency still must devote the necessary resources to the effort, even if it hires a contractor. The test developers will need assistance from SMEs in the program area to gain an understanding of the position and the competencies necessary to successfully carry out its duties, and to develop high-quality tests that draw from real life scenarios or situations.

Test development should include, among other things:

- A test blueprint to ensure that the content of the assessment matches the job and the relative importance of each content area;
- Standard procedures and instructions to ensure consistency;
- Review by SMEs for content;
- The elimination of contaminating factors, such as unnecessary jargon, equipment, or other testing elements;
- Review for content that may be offensive to some applicants;
- Field tests to ensure the reliability and validity of the instrument(s);
- Field tests to ensure that applicants understand the instructions and content;
- Clear and understandable criteria for scoring the test(s); and
- A consistent process for scoring and documentation.

(4) Select Qualified Assessors

A key factor in administering an effective assessment is having qualified assessors. Unqualified assessors who do not understand the job, the assessment, or how to effectively score the candidates will undermine the quality of the assessment itself. Thus, when choosing assessors, there are some key points to keep in mind.

First, assessors who know applicants being assessed may not be a good choice because they might find it difficult to be objective.[56] Second, hiring organizations should choose assessors who want to participate in the process. If the assessor does not value the process or feels it is a burden on his or her time, the assessor may not commit the necessary effort. Furthermore, the hiring organization should ensure that assessors have sufficient

---

[56] Cam Caldwell, George C. Thornton III, and Melissa L. Gruys, "Ten Classic Assessment Center Errors: Challenges to Selection Validity," *Public Personnel Management*, Vol. 32, No. 1, Spring 2003, p. 80.

time to devote to their duties. Because serving as an assessor is generally a collateral duty assignment, some assessors may find that other competing priorities prohibit them from efficiently carrying out their duties. Third, consider the diversity of the assessors. This diversity should take into consideration not only traditional diversity factors such as race, ethnicity, age, and gender, but also organizational level and functional work area.[57]

Finally, agencies should consider using managers and SMEs as assessors because they know the job. Some research indicates that if they are trained in proper assessment techniques, managers and SMEs may outperform even assessment professionals in evaluating candidates.[58] At a minimum, assessors should have some type of subject matter expertise, insight into the organization's needs and values, or unique skills in behavioral assessment.[59] The four agencies highlighted in Appendix A of this report use agency managers or SMEs to assess applicants using simulations, and one agency even uses customers on its evaluation panels.

Anyone responsible for either administering the test or rating the candidates should be well trained in how to carry out their responsibilities. The Assessment Center Guidelines, while providing guidance for one particular assessment type, contain some good points to consider when training assessors in general. First, agencies should ensure that the content of the training is appropriate. At a minimum, the training should help assessors understand the job, the dimensions being measured, the assessment techniques being used, and the evaluation procedures. The training length should be appropriate to: (1) the trainer and instructional design used; (2) the skill and knowledge levels of the assessors; and (3) the content and complexity of the job and assessment technique(s) being used.[60]

Agencies should keep in mind that when considering developing a new assessment strategy, such as using job simulations, they do not have to do an immediate overhaul of their entire program. That approach can be overwhelming. A formal restructuring of the assessment program can require a large commitment of resources to build and manage the program. This kind of support and commitment from agency leadership is not always available.

Instead, selecting officials and HR staffs can collaboratively review the current process and identify small, informal steps they can take to improve the hiring process. For instance, they may find that the job analysis is out of date and needs to be updated before

---

[57] Assessment Center Guidelines, *Guidelines and Ethical Considerations for Assessment Center Operations*, p. 4.

[58] Caldwell, Thornton, and Gruys, "Ten Classic Assessment Center Errors: Challenges to Selection Validity," p. 80.

[59] Caldwell, Thornton, and Gruys, "Ten Classic Assessment Center Errors: Challenges to Selection Validity," p. 80.

[60] Assessment Center Guidelines, *Guidelines and Ethical Considerations for Assessment Center Operations*, pp. 6-7.

the next hiring cycle. A review of the existing assessments may show that they are using too many or too few, or that they are just asking the wrong questions. Agencies may also want to consult with OPM, whose Products and Services group works with agencies on a reimbursable basis and may be able to help the agency assess its current situation and focus its efforts for little to no cost.

(5) Measure the Success of the Assessment Procedures

As mentioned in the previous section, most of the organizations that reported using job simulations in the past year believed that the simulations increased the quality of the agency's selections, but few reported actually measuring the impact. Such measurement is a critical step in the selection process. Measuring the costs, benefits, and results allows an organization to determine which assessments identify the best candidates and, therefore, where assessment funding and resources should be targeted. Further, such evaluations can help justify resource allocation to agency leadership, a critical step in this time of limited resources.

The organizations that did measure the effectiveness of these assessments generally conducted validation studies, used surveys of HR and selecting officials, and measured the rate of successful probationary employees. These are valuable steps and indicate that there are different levels at which selection procedures should be measured.

First, as discussed previously, a good assessment is reliable, valid, and fair. Each of these factors can, and should, be statistically measured to ensure that the assessments are meeting these criteria and are legally defensible if litigation is brought. In particular, the research shows that as the assessment's predictive validity increases, so too does the percentage of the persons hired who become successful employees. Although, content valid assessments may be easier to develop and still legally defensible, criterion-related validity achieved through formal validity studies will help increase the organization's return on investment for the assessment.

Second, assessments can be measured in terms of participants' perceptions. Good assessments should result in positive applicant reactions, so it may be beneficial to obtain applicant feedback. Selecting officials can provide feedback regarding their satisfaction with the process and the resulting new hires. Those involved in the selection process—e.g., the test developers, administrators, and assessors—can provide input on what went well, what did not, and what improvements should be considered.

Finally, there are HR indicators that can demonstrate the success of the assessment process. Examining employee retention and the rate of successful probationary employees can tell the organization if the assessment resulted in a good person-to-job fit. Ultimately, the quality of the workforce and how well it is prepared to carry out

the agency's mission is the ultimate outcome of any human resources process, including assessment. Workforce quality is not an easy concept to measure, but demonstrating that the assessments resulted in a workforce that can efficiently and effectively perform the mission will solidify the business case for committing resources to assessment.

Conclusion

Job simulation assessments can be an effective tool to evaluate applicant qualifications. They have many advantages, including relatively high levels of validity, better person-to-job fit because of the realistic job preview, a greater degree of fairness, and generally positive applicant perceptions. They do have their drawbacks, though, particularly the potentially high resource costs to develop and administer more advanced simulation exercises. However, the evidence does suggest that agencies could achieve a higher return on investment if they use more predictive assessments during the selection process, making job simulations a good option.

Nonetheless, job simulations may not work in every situation. That is why it is important for agencies to have a good grasp of the job for which they are hiring, the competencies needed for that job, and knowledge about which assessments would best fit their specific needs and which would not help them. We have brought up a number of factors for agencies to consider when making decisions about their assessment process. They are by no means exhaustive. However, the five steps outlined in this assessment strategy should help get agencies on the right path to developing and implementing an assessment program that assists them in selecting employees on the basis of relative ability, knowledge, and skills, as set forth in the merit system principles.

Below, we provide information on some typical job simulations that organizations can add to their assessment tool kit. This is not meant to be a comprehensive discussion of the individual assessments, but an effort to provide readers with some context, generalized research findings, and usage examples so that they can become familiar with the advantages and disadvantages of these assessments and make educated decisions on what is appropriate to pursue for their organization. This appendix addresses work samples, situational judgment tests, assessment centers, and job tryout procedures, presenting for each a definition, appropriate uses, examples, scoring issues, validity ratings, and a discussion of fairness and adverse impact.

Work Samples

Definition. A work sample evaluates applicants' job-related skills by having them perform actual activities or tasks that are physically or mentally similar to the duties they would perform on the job. Work samples are generally high fidelity assessments in that they use hands-on simulations with realistic materials and equipment administered under controlled circumstances. Work samples are classified as motor or verbal. Motor work samples involve physically manipulating something, where verbal work samples deal with people or language-oriented problems.[61]

Appropriate Uses. Because applicants are asked to perform a task that is similar to what is performed on the job, work samples assume that applicants already possess some of the knowledge, skills, or abilities necessary for the job. Therefore, they are more appropriate when hiring for experienced or skilled workers. They only measure the competencies needed for the specific activity carried out during the test, making them less able than some other assessments to measure aptitude or future potential. This indicates that they may not be the best choice for entry-level positions. They are useful for tasks that can be completed in a short period of time, but less so for tasks that take longer to complete. Work samples can be time consuming and expensive to administer, so they are best used on smaller groups of applicants, such as those who have already scored well on other

---

[61] James J. Asher and James A. Sciarrino, "Realistic Work Sample Tests: A Review," *Personnel Psychology*, 1974, Vol, 27, p. 519.

selection instruments. In fact, work samples could be used to bolster assessments that ask for applicants' evaluations of themselves. While an applicant may report having a certain level of training or experience, a work sample will help measure the quality of that training or experience. Because they are limited in scope, work samples are best when used in conjunction with a multiple hurdle approach to assessment—successively using good assessment procedures that measure a variety of competencies.

Examples. Work samples can vary in complexity and design. A work sample assessment for a welder may include an onsite welding exercise. This could be a fairly simple work sample to design and administer. The applicants will most likely already have the basic knowledge to perform the task. Therefore, developers just need to make sure the equipment is available and develop measurements for the products produced. Of course, there may be applicants who do not have the appropriate knowledge level to operate the equipment, so the organization should ensure that safety measures exist to deal with these situations.

Work samples can get more complicated. For example, applicants for a customer service job might perform a role-play between the applicant and an "angry customer" (the assessor). This type of role-play requires much more preparation. Employers need to decide what equipment and materials the applicants will need to replicate the situation and ensure that they can control these supplies during the test. For example, they would probably need a work station, a working telephone, a computer with access to necessary records or the Internet, the agency's standard operating procedures, and any other information necessary to be able to respond to the scenario. For each scenario used, the test developers and subject matter experts need to anticipate what potential responses applicants may provide and design benchmarks for what are acceptable responses and what are not. Also, the assessors playing the "angry customer" will need extensive training in administering and scoring the role-play.

Scoring. Work samples can be evaluated by scoring the product or the process (or a combination of the two). Product scoring involves scoring the applicant on the quality of the end product that results from the test, such as an edited document. In process scoring, the applicant is scored on his or her actions during the work sample, such as the customer service representative's response to the angry customer.[62]

Validity. Historically, work samples have been viewed by researchers as one of the most valid and predictive employee assessments available. A landmark meta-analysis of assessment validity, published by Schmidt and Hunter in 1998, found the work sample validity coefficient to be .54, making it the most predictive assessment of the 19 evaluated

---

[62] Robert D. Gatewood and Hubert S. Feild, "A Personnel Selection Process for Small Business," *Journal of Small Business Management*, October 1987, Vol. 25, No. 4, p. 23.

in the study.[63] However, more recent work in this area has pointed to the need for more research. A 2005 meta-analysis by Roth, Bobko, and McFarland pointed out limitations in the way work samples have been studied. Correcting for these perceived flaws, their study found the validity coefficient to be closer to .33, which is still better than many other assessment methods.[64]

Fairness. Assessment research generally considers work samples to have fairly high face validity, meaning that applicants view them as fair because they can see the relationship between the task performed in the work sample and the tasks to be performed on the job. One study found that work samples and structured interviews were viewed more favorably by applicants than other assessments.[65]

Typically, work samples have been viewed as having very low adverse impact, which means members of particular ethnic, racial, or gender groups are not disadvantaged by the nature of the assessment. However, this finding was recently challenged by a 2008 study which found that when controlling for perceived flaws in previous work sample studies, the work sample did result in higher black-white differences than had been previously thought.[66] Again, more research in the area of work sample assessments could help resolve differing points of view in the research community.

Though the jury is still out on the level of adverse impact, a recent study did show that work samples have a lower degree of exposure to discrimination lawsuits that are based on the selection procedure used. A review of Federal court cases involving hiring or selection discrimination from 1978 to 2000 found that work sample tests were less likely to be challenged in court than some other assessments (such as cognitive ability tests or unstructured interviews) and survived almost all of the litigation taken against them.[67]

Other Advantages. Because work samples ask applicants to actually perform a specific task that is necessary to do the job, applicants are less able to fake proficiency with the task and/or competencies required than other non-cognitive forms of assessment (e.g., personality tests). In addition, the work sample provides a realistic job preview of what the applicant will be doing on the job. If an applicant is able to determine if the job is a good fit for them before accepting the position, this could reduce turnover while increasing productivity and job satisfaction.

---

[63] Schmidt and Hunter, "The Validity and Utility of Selection Methods in Personnel Psychology," p. 265.

[64] Philip L. Roth, Philip Bobko, and Lynn A. McFarland, "A Meta-Analysis of Work Sample Test Validity: Updating and Integrating Some Classic Literature," *Personnel Psychology*, Winter 2005, Vol. 58, No. 4, p. 1009.

[65] Hausknecht, Day, and Thomas, "Applicant Reactions to Selection Procedures."

[66] Philip Roth, Philip Bobko, Lynn McFarland, and Maury Buster, "Work Sample Tests in Personnel Selection: A Meta-Analysis of Black-White Differences in Overall and Exercise Scores," *Personnel Psychology*, Autumn 2008, Vol. 61, No. 3, p. 637.

[67] David E. Terpestra and R. Bryan Kethley, "Organizations' Relative Degree of Exposure to Selection Discrimination Litigation," *Public Personnel Management*, Fall 2002, Vol. 31, No. 3, pp. 288-289.

Other Disadvantages. While some work samples can be developed and administered quite easily, more complex tests can take a lot of time and resources to develop and administer. Often, they can only be administered to one applicant at a time or to a small group of applicants. Technology can help to reduce some of these requirements, but may only be able to provide limited relief.

Work Samples in Action

The Internal Revenue Service's (IRS) Contact Representative (Customer Service Representative) is a mission-critical, entry-level position with no specific education or selective factor requirements. Contact Representatives are the people who answer tax questions for the public on the 1-800 number. This job is critical in ensuring that the public obtains the information it needs to correctly meet tax code requirements. To help identify high-quality applicants for this position at nationwide locations, IRS developed the Telephone Assessment Program (TAP).

During the TAP, an applicant is seated at a desk with a telephone and given 30 minutes to review a packet of information that describes the job and explains the procedures used for the assessment. During this 30-minute period, the applicant can ask questions of the HR staff administering the assessment to make sure he or she understands the process.

The applicant then receives four calls that are 5 to 7 minutes in length. Each call presents a scenario in which the applicant must carry out the duties of the position. The applicant can use the information packet to find the answers to each "taxpayer's" questions. The "taxpayer" is a trained evaluator who is rating the applicant on seven competencies that were identified through a job analysis and working with subject matter experts and managers. For each competency, there is a set of benchmarks the evaluator uses to measure the applicant's performance. The first call is a practice call, and the evaluator can come out of his role to assist the applicant if necessary. The three other role-plays are rated against the benchmarks. The evaluator records observations/ratings on a computer screen as they go through the call. The total time of the assessment is about 1 hour.

The assessment was developed by IRS senior specialists and I/O psychologists with the assistance of a contractor. They conducted a validation study that showed a strong relationship between the assessment and success on the job. They have looked at adverse impact and found virtually none. The assessment has been updated periodically for content.

The IRS had top management support and dedicated resources for improving assessments from the beginning. In 1998, Congress passed legislation to modernize IRS. As part of that effort, the Commissioner wanted a more professional workforce, more reliance on technology, and to reduce the burden on managers. He also wanted to explore new and innovative ways to carry out the mission.

Both applicant and supervisor reactions have been positive. The IRS has found that the work sample assessment provides a realistic job preview that gives applicants the opportunity to experience both the positive and negative aspects of the job before accepting a job offer. Anecdotal evidence suggests that the assessment has helped reduce turnover. The key drawbacks to the assessment are its expense and resource requirements. The IRS is currently piloting a new process that integrates online assessment with the TAP to help reduce resource and applicant burden.

Situational Judgment Tests

Definition. A situational judgment test presents applicants with a description of a work-related problem scenario and asks them to exercise their judgment by choosing or evaluating alternative courses of action to the situation. The tests are almost always multi-dimensional and test different skills and abilities. Historically, SJTs have been paper and pencil tests, making them a low fidelity simulation. However, there is a growing trend in video-based testing which uses video technology to present the scenarios and even to record applicants' responses. The technology provides applicants with a more realistic feel and greater job preview.

Appropriate Uses. SJTs have been found to be effective measures of social functioning dimensions such as conflict management, interpersonal skills, problem solving, negotiating, and teamwork. They have also been found to be particularly useful for assessing managerial and leadership competencies.[68] They are widely used by public organizations, particularly for law enforcement and firefighter positions as well as other occupations that require high levels of independent judgment.

Examples. Typically, SJT prompts are either behavioral or knowledge-based.[69] Behavioral tests ask respondents how they would personally react to a given situation, as shown in the following example of a multiple-choice question:

> You have noticed that there is a lack of cooperation and trust
> among the employees in the organization you manage. This
> counter-productive behavior is adversely affecting both the quality
> and the quantity of the work produced by the employees. You
> would most likely—

---

[68] Office of Personnel Management, "Section III: Situational Judgment Tests," *Assessment Decision Guide*, http://apps.opm.gov/ADT/.

[69] Michael A. McDaniel, Nathan S. Hartman, Deborah L. Whetzel, and W. Lee Grubb III, "Situational Judgment Tests, Response Instructions, and Validity: A Meta-Analysis," *Personnel Psychology*, Spring 2007, Vol. 60, No. 1, p. 63.

    a. Inform the employees that failure to cooperate and work harmoniously will result in disciplinary action.

    b. Bring in someone to do some team building with the employees.

    c. Meet with the employees to discuss your observations and identify the cause of the interpersonal problems.

    d. Change some of the work assignments in order to attain a more cohesive and cooperative work group.[70]

Knowledge-based tests ask applicants to evaluate the effectiveness of possible responses to the given situation, as shown below in another multiple-choice question:

> You are assigned as a manager of a complex project coordinating the work of eight different teams. Four of the teams are on schedule to meet their deadlines. The other four have fallen behind and are in danger of missing their deadlines due to cutbacks in staffing levels. It is important that the work products from the eight teams stay on schedule. How effective is each of the following actions you could take?
>
>   a. Reassign the team leaders to help the project get back on track.
>
>   b. Talk to the teams that have fallen behind and find out what I can do to help.
>
>   c. Request overtime for the four teams that have fallen behind.
>
>   d. Inform upper management that the staffing cutbacks are causing setbacks in my project.[71]

The key to developing SJTs is to create items that are not obviously right or wrong. If correct answers are too obvious, the SJT can become a test of the applicants' subject matter knowledge rather than a test of their judgment in a complex situation, making it more difficult to make true distinctions among the applicants' qualifications. For instance, the following multiple-choice example will help identify who has basic computer knowledge, but not much more:

> You receive an email at work from a former college roommate whom you have not heard from in months and whom you generally communicate with through your personal email account. It is titled "Remind you of anyone?" and has an

---

[70] Example taken from Dennis A. Joiner, "Assessment Center Trends," Presentation at IPMA-HR Assessment Council Conference on Personnel Assessment, June 23, 2004.

[71] Example taken from FBI, "Situational Judgment Test Instructions," at http://www.fbijobs.gov/11215.asp.

attachment with a .jpg extension. The text of the message
contains several misspelled words. The attached file is 350 KB.
The best thing to do is:

a. Forward the message to your office's IT contact.
b. Save the attachment to read after work.
c. Open the attachment.

Applicants typically receive a few practice questions to help them get used to the format. Generally, tests that have more detailed questions have been found to have higher correlations with job performance. Good SJT scenarios contain a rich set of details, only some of which help determine the correct answer. When there is insufficient detail, the question does not fully test an applicant's ability to focus on what is most important.

Scoring. Most SJTs can be easily administered and scored. For instance, a multiple-choice SJT can be administered and scored like any other multiple-choice test and works equally well on paper and on the Web. Scoring is objective and can be done by test administrators, through computer scoring of a test answer sheet, or immediately by an online test administration program. When using video or digital technology to record applicant answers, assessors can evaluate candidates' responses at a later, more convenient time and location.

Test security is important to prevent unauthorized access to answer keys and answer sharing by test takers. But SJTs have a built-in security advantage. Because an SJT question presents a work scenario that includes relevant and irrelevant details, the narrative explanation is longer than a typical multiple-choice question found on other types of "tests," such as a job knowledge test. The narrative is therefore harder for test takers to remember completely and share accurately after the testing session.

Validity. Situational judgment tests have been found to be moderately predictive of future job performance. One meta-analysis found them to have an estimated validity of .34, though some tests had higher validity than others.[72]

Fairness. Because SJTs present applicants with realistic situations that they may experience on the job, they have a relatively high level of face validity. Applicants can see the direct relationship between the test and the job. The research community agrees that SJTs have lower levels of adverse impact than cognitive ability tests, but some studies have found that adverse impact levels are still moderate. Interestingly, studies have found that the method of testing affects the level of adverse impact, with video-based SJTs

---

[72] Michael A. McDaniel, Frederick P. Morgeson, Elizabeth Bruhn Finnegan, Michael A. Campion, and Eric P. Braverman, "Use of Situational Judgment Tests to Predict Job Performance: A Clarification of the Literature," *Journal of Applied Psychology*, 2001, Vol. 86, No. 4, p. 734.

showing lower levels of black-white subgroup differences than the paper and pencil SJT.[73] Video-based tests may be more expensive to develop than pencil and paper tests, but they are often less expensive to develop than high fidelity simulations, such as work samples and assessment centers. Also, they can still be used for a large number of applicants in a single session, provide a realistic job preview, and have low levels of adverse impact.

Situational Judgment Tests in Action

The U.S. Customs and Border Protection (CBP) uses a Video-Based Test (VBT) for CBP officers and agricultural specialists that could be considered an advanced version of a situational judgment test that increases the fidelity of the assessment, provides a more realistic job preview, and reduces the burden on managers assessing the candidates' responses. Through several critical incident job analyses, the Personnel Research Assessment Division (PRAD) identified a set of key tasks, competencies, and KSAs necessary to successfully perform the job. Specifically, the assessment concentrates on interpersonal, oral, and decision-making skills, which are competencies for which SJTs are an effective measurement. The I/O psychologist staff worked with subject matter experts to identify critical incidents to use in video-based scenarios.

Applicants report to one of 50 VBT sites across the country. During the VBT, applicants are given instructions by the administrator as part of the narration in the video. There are eight scenarios that are each a few minutes long. The scenarios go from least difficult to most difficult. Applicants watch a job-related situation on a television monitor in a specially equipped testing room. Each scenario is followed by a 45-second response period in which applicants act out what they would do in response to the scenario presented. The role-play response is captured on video. To get comfortable with this interactive type of role-play, CBP allows two test scenarios. In total, the VBT lasts about 30 minutes per applicant.

The videos can then be sent to over 300 ports of entry to be rated by trained raters—generally supervisory CBP officers who were chosen based on a list of criteria and who receive 1.5 days of training in the assessment and how to evaluate responses. When conducting the ratings, the supervisory CBP officers can fast-forward through the instructions and scenarios and focus on viewing the applicant responses and providing ratings, reducing the amount of time it takes to assess an applicant down to about 10 minutes. In addition, the tape-recorded applicant responses are saved and used as documentation, which reduces the amount of note-taking that raters must complete. Furthermore, the VBT can be rated at any time, such as during non-peak hours, reducing the operational burden.

---

[73] David Chan and Neal Schmitt, "Video-Based Versus Paper-and-Pencil Method of Assessment in Situation Judgment Tests: Subgroup Differences in Test Performance and Face Validity Perceptions," *Journal of Applied Psychology*, 1997, Vol. 82, No. 1, p. 143.

The VBT is used as part of a multiple hurdle process. Applicants are given the opportunity to voice their opinion about the process through a short survey given upon completion of the testing session, and responses indicate that the assessment is well received. It has also been well received by selecting officials. The VBT has been so successful for CBP hiring that its use has been extended to the agriculture specialist position. PRAD received the 2007 International Public Management Association for Human Resources Assessment Council (IPMAAC) Innovations in Assessment Award in recognition of its work in developing and implementing the Video-Based Test.

Assessment Centers

Definition. Assessment centers evaluate applicants on their job-related KSAs using multiple, standardized exercises. Each exercise is developed to measure clearly defined dimensions of behavior. Dimensions are clusters of activities important to the job that are observable, are specific, and consist of tasks related to the job. Trained assessors use predetermined criteria to systematically score the applicants' performance on each exercise. The scores on each of the assessments are statistically integrated so that each exercise contributes to the applicant's overall score (as opposed to a multiple hurdle process in which each exercise is used to screen applicants out of the successive exercises). There is no set number of exercises used during this process, but the International Task Force on Assessment Center Guidelines stipulates that at least one of the assessments must be a simulation, defined by the Task Force as "an exercise or technique designed to elicit behaviors related to dimensions of performance on the job requiring the participants to respond behaviorally to situational stimuli."[74]

Assessment centers were first used by military organizations and intelligence services to select officers and undercover agents during World War II. The skills necessary for these roles were rather complex and not easily measured by existing tools. Then, in the 1950s, Douglas Bray of AT&T began applying the technique to the business environment. This resulted in an extremely detailed analysis of assessment centers and their capability to measure success on the job. AT&T found a strong correlation between the assessment center scores of more than 400 entry-level employees and the different measures of managerial progress. A follow up study was conducted in the 1970s. Ultimately, these groundbreaking studies found that management potential is highly predictable using assessment center techniques.

Appropriate Uses. Research has found that assessment centers can be effectively used for both employee selection and career development. They have been cited as being particularly helpful in selecting or developing supervisors, managers, and executives for promotional opportunities or development programs.

---

[74] Assessment Center Guidelines, *Guidelines and Ethical Considerations for Assessment Center Operations*, p. 3.

Assessment centers require considerable time and financial resources. They take time and expertise to develop and organize. They require multiple trained raters. They require space, equipment, and materials to administer. All of these resources amount to a fairly significant cost. Because of the resource requirements, this type of assessment is best used for mission-critical and/or high volume occupations. For that reason, large organizations tend to use them more than small organizations.

While assessment centers are resource intensive, they can be streamlined by using technology that can reduce the cost and resources needed to administer them. Some organizations have started using online exercises. Some organizations have also turned to video technology to administer and record simulation exercises.

Examples. Assessment centers always use more than one exercise to measure the different applicant dimensions, such as group, individual, written, analytical, and/or role-play exercises. While there are many different exercises that can be used, both simulation and non-simulation, some are more common than others. For instance, a study found that:[75]

- 83 percent of ACs used in-basket exercises, which assess an applicant's ability to perform managerial administrative duties. The applicant is instructed to take action on various issues or problems that have accumulated in a manager's in-basket. The in-basket may contain letters, reports, telephone messages, emails, or other communications from supervisors or staff that need to be addressed, and the applicant's actions are evaluated using predefined criteria.
- 78 percent used simulated coaching meetings. During this simulated exercise, applicants role-play situations in which they are resolving performance issues with a subordinate.
- 70 percent used leaderless group discussions (LGDs), a simulation exercise in which applicants are presented with an issue or problem and put into groups to develop solutions. For instance, the group may be instructed to develop a policy or to work out budget issues among their divisions. The applicants' roles may be defined or undefined and they may be put in situations of collaboration or competition. All applicants are separately rated on their participation and response to the situation.
- 70 percent used structured interviews. Structured interviews are interviews in which the applicants are asked the same questions in the same order and are rated on their responses using the same rating scale.

Guidelines recommend that at least two exercises be used to measure each competency dimension. Some organizations use off-the-shelf exercises for their assessment centers, but this is generally not recommended. The exercises should be directly tied to the job analysis and be tailored to the job for which the organization is hiring to be most effective.

---

[75] Warren Bobrow, "Assessment Centers," *Intervention Resources Guide: 50 Performance Improvement Tools*, Danny G. Langdon, Kathleen S. Whiteside, and Monica M. McKenna, eds., Jossey-Bass/Wiley: San Francisco, CA 1999.

All exercises should be standardized, including the instructions provided, the manner in which they are provided, the materials used to carry out the exercise, and the scoring criteria used to rate the applicants. To achieve this, developers recommend that all assessment center exercises be piloted to ensure the clarity of instructions, to test the time limits given, and to ensure administrators and assessors understand their roles.

Scoring. Each exercise in the assessment center has predetermined, standard rating criteria to be used for each applicant. Each exercise requires a set of trained assessors to rate each candidate. Some exercises require more assessors than others. For instance, the leaderless group discussion usually requires one assessor per applicant to carefully observe that applicant's participation in the exercise. On the other hand, a written exercise that can be rated after the exercise takes place can require fewer assessors for all of the applicants.

One of the most important factors to effective scoring is to ensure that assessors are well trained in systematically recording and evaluating behaviors. The length of the training will vary depending on design, assessor knowledge, and the assessments used, but a typical rule of thumb for training assessors who have no AC experience may be 2 days of training for each day of the assessment center administration.[76] AC guidelines indicate that training should provide: extensive knowledge of the organization; understanding of what the behavioral dimensions are, their relationship to job performance, and examples of effective performance; thorough knowledge of the assessment techniques being used, what dimensions they are measuring, and specific policies and practices of the organization regarding assessment; the ability to observe, record, and classify behavior; extensive understanding of the rating procedures; and the ability to give accurate feedback when necessary.[77]

Validity. Assessment centers have shown moderate validity levels for job performance, ranging from .36 to .43 in the studies we reviewed.[78]

Fairness. Assessment centers are generally viewed as one of the more fair and objective assessment methods available. They have shown little to no adverse impact. Studies have evaluated ACs in terms of age, race, and gender and found that they are equally valid for all candidates. Federal courts have supported this view by actually mandating assessment centers to overcome selection problems in other assessment methods.[79]

---

[76] Assessment Center Guidelines, *Guidelines and Ethical Considerations for Assessment Center Operations*, p. 7.

[77] Assessment Center Guidelines, *Guidelines and Ethical Considerations for Assessment Center Operations*, p. 6.

[78] For instance, see: John E. Hunter and Ronda F. Hunter, "Validity and Utility of Alternative Predictors of Job Performance," *Psychological Bulletin*, Vol. 96, No. 1, 1984; Barbara B. Gaugler, Douglas B. Rosenthal, George C. Thornton III, and Cynthia Bentson, "Meta-Analysis of Assessment Center Validity," *Journal of Applied Psychology*, Vol. 72, No. 3, 1987; Paul G. W. Jansen and Bert A. M. Stoop, "The Dynamics of Assessment Center Validity: Results of a 7-Year Study," *Journal of Applied Psychology*, Vol. 86, No. 4, 2001.

[79] Byham, "Section 2: Validity and Fairness," *What is an Assessment Center?*

Furthermore, ACs are viewed to have high face validity because of the emphasis on behavior in relation to the job, making them appear more fair to applicants. They can also provide applicants with a realistic job preview to better inform their decision on whether or not they want the job.

Assessment Centers in Action

The Department of Commerce's Foreign Commercial Service Officers are assigned to foreign and domestic field offices to promote the export of U.S. goods and services and defend U.S. commercial interests abroad. To hire entry-level Foreign Service Officers, Commerce uses an assessment center process that has been successful for many years.

The vacancy announcement is opened every 2 years, and OPM accepts and pre-screens the applications using KSA narratives. The AC uses a number of different assessments to further evaluate the candidates in terms of six critical competencies identified through the job analysis: written and oral communication, problem solving/decision making, personal/professional growth, working with and leading others, cultural skills, and achieving results.

The assessment center consists of a structured interview, a leaderless group exercise, an electronic in-basket, a writing exercise, and a Demarche exercise (a formal diplomatic representation of the official U.S. position on a given subject). Each exercise addresses different competencies and trade craft. About 12 candidates go through the center per day and are evaluated by trained assessors. The assessment center generally lasts 10 days and evaluates the top 100 candidates. Applicants pay for their own transportation and accommodations to attend the assessment center.

Assessors are mostly Commerce employees who volunteer for the responsibility or who have been recruited. The Foreign Commercial Service (FCS) also recruits private sector customers to assess candidates, thereby obtaining an external perspective about the candidates. The assessors receive three days of training in what their role is and how to effectively carry it out. Assessors are trained in their particular exercise and therefore stay with the exercise, not the candidate.

A contractor plays a large role in helping to plan, develop, and administer the assessment center. The contractor reviews the job analysis every few years to make sure the exercises are still valid, still address the job requirements, and are legally defensible. The contractor assists in pulling together the logistics for the center, such as reserving multiple rooms in a conference center to accommodate each individual assessment and making sure FCS has the equipment and assessors necessary for the exercises. This person trains the assessors and generally manages the flow of the process. At the end of each center, the contractor conducts an after-action report that details what worked, what did not, and suggests improvements.

The result of the assessment center is a registry that is rank-ordered by the scores obtained during the assessment battery. Ultimately, FCS places about 40 of the top performers, for a total cost of about $12,000 per successful offer. Though the process is costly, FCS has been very happy with the quality of the applicants and believes it is getting a good return on its investment.

Job Tryout Procedures

Definition. In a job tryout, applicants are generally hired with minimal screening of their qualifications and are given an evaluation period. During the evaluation period, they receive the training they need to perform the duties of the job, and their performance is evaluated to determine if they meet the established levels of satisfactory performance. At the end of the evaluation period, they are either retained or terminated based on their performance.

There is not a significant amount of professional literature pertaining to job tryout procedures. This is probably because true job tryouts are not used very often. Most organizations would view them as impractical because of the high cost of terminating lightly screened, low performers.

Appropriate Uses. The job tryout procedure can be used with almost any kind of position, particularly entry-level jobs. The evaluation period may vary, depending on the complexity of the job. Typically, it is about 6 to 8 months.[80] An easy-to-learn position may have a short evaluation period, such as two weeks, whereas a more complex position could have a longer evaluation period. However, users of this procedure should keep in mind that the more training that is required for the position, the lower the return on investment will be because more resources will be put into an employee who was minimally screened and could easily fail. Also, job tryouts may be more useful for high-volume occupations with high turnover that need applicants continuously in the pipeline. The organization would get a higher return on investment by using its resources to train and assess a larger volume of people that can fill multiple vacancies than by using those same resources to train and assess applicants for a single position.

Examples. In the Federal Government, new employees are generally subject to a probationary or trial period. The purpose of the probationary or trial period is to provide the Government the opportunity to evaluate the employee on the job to determine if the appointment should become final at the end of the period. Until the probationary period has been completed, a probationer is still just an applicant for an appointment. If used correctly, the probationary or trial period could be considered a job tryout procedure.

---

[80] Schmidt and Hunter, "The Validity and Utility of Selection Methods in Personnel Psychology," p. 268.

Scoring. A true job tryout procedure will have established criteria defining what satisfactory performance on the job is. The criteria will be based on a job analysis that defines what the duties and tasks of the position are and what outcomes are expected from successful completion of those duties and tasks. The people hired into the program will be trained in the duties and tasks necessary to complete the job, will be told how they will be evaluated throughout the tryout, and will be evaluated by a trained assessor/supervisor against the pre-established criteria. Those who do not meet the expected performance by the end of the evaluation period will be terminated.

Validity. The two key studies that address job tryout procedures found them to be highly predictive of job performance, with a validity coefficient of .44. Some may be surprised that the coefficient is not higher, given that the "applicant" actually performs the job for a significant period of time. However, the study authors found that supervisors are reluctant to terminate marginal performers. Because this is an unpleasant action, supervisors are more likely to lower the standards by which they evaluate the applicants, thereby undermining the effectiveness of the assessment.[81] Our research on the Federal probationary period supports this finding, given that many supervisors we surveyed did not intend to remove probationers who they thought were not an asset to the organization.[82] Thus, the key to ensuring an effective, predictive job tryout is following through on terminating employees who do not meet the standards for satisfactory performance.

Fairness. We could not find any in-depth research concerning the fairness and/or adverse impact of job tryouts. It would seem that this type of simulation assessment would be considered fair and objective, simply because the applicants are actually hired into the position and given the opportunity to succeed. However, we know that Federal probationers do raise claims of discrimination, so we need to be cognizant that fairness issues may exist.

Job Tryouts in Action

Best Friends Animal Society is a nonprofit organization that operates the Nation's largest sanctuary for abused and neglected animals. The sanctuary is located in a small, remote town in Utah, and the work performed at the sanctuary can be both physically and emotionally grueling. Both of these factors have led to problems with employee turnover and job burnout. To ensure that new employees are able to adapt to the small town life and can endure the physical and emotional demands of the job, Best Friends uses a job tryout procedure to assess applicants for many of their entry-level positions, including animal caregivers and support staff.

---

[81] Schmidt and Hunter, "The Validity and Utility of Selection Methods in Personnel Psychology," p. 268; Hunter and Hunter, "Validity and Utility of Alternative Predictors of Job Performance," p. 82.

[82] MSPB, *The Probationary Period: A Critical Assessment Opportunity*, p. 7.

The hiring process starts with the applicant submitting an application on-line. The HR staff reviews the applications, conducts telephone interviews, and works with the hiring manager to determine who will be invited for the 2-week evaluation period. They extend the invitation to the applicant and schedule the evaluation period generally within 2 to 4 weeks of the notification.

Applicants use their own time and pay their own expenses (including accommodations, transportation, and meals) to come for the 2-week job tryout. Best Friends pays each applicant $10 per hour and does not guarantee employment at the end of the tryout. Because their mission is so appealing to people in the industry, applicants generally do not balk at the job tryout requirements.

Best Friends has established a cadre of trainers who provide applicants with the necessary training, give them their assignments, and evaluate them as they perform their duties. They have identified a set of core competencies for the occupations being evaluated and established legally defensible benchmarks against which applicants are measured. The trainers record their observations and feedback and provide these to the hiring manager. The hiring manager meets with the applicant at the beginning, middle, and end of the 2-week evaluation period to provide the applicant with feedback.

About 80 percent of those who make it to the job tryout are hired. Reaction from the applicants has been very good. They understand the need for the evaluation period and are supportive of it. There have only been a few instances in which applicants were not able to attend the tryout because of logistics, and these applicants were kept in the pipeline for future positions.

The 2-week paid job tryout has been in place for almost 10 years, and Best Friends reports that it has been successful in helping the organization hire a qualified workforce that can deal with the unique location and work of the sanctuary.

U.S. Merit Systems Protection Board Interrogatory
Agencies' Use of Job Simulations in Assessing Applicants

Introduction:

The U.S. Merit Systems Protection Board (MSPB) is conducting a study of how Federal agencies use job simulations to assess applicant qualifications. The study will help us determine whether this type of assessment can help agencies improve the quality of their personnel selections.

This interrogatory should be completed by an individual who has a good understanding of the personnel assessment tools your agency uses. If your agency has components or organizations with distinctive practices, policies, or concerns, we ask that your response include these or that you forward the interrogatory to the appropriate components for a separate response. Some of the questions will require the personal judgment of the subject matter expert.

Definition: For the purpose of this interrogatory, *a **job simulation** means presenting applicants with realistic, job-related situations and documenting their behaviors to help determine their qualifications for the job.* Job simulations could include but are not limited to work samples, situational judgment tests, assessment centers, or job tryout procedures.

Contact Information: You may email your response to laura.shugrue@mspb.gov; fax it to (202) 653-7211; or mail it to MSPB, Office of Policy and Evaluation, ATTN: Laura Shugrue, 1615 M Street, NW, Washington, DC 20036. If you have questions or would like to obtain an electronic version of the questionnaire, please contact Laura Shugrue at the email above or call (202) 653-6772 x1124. **Please respond by July 25, 2008.**

Questionnaire:

Agency/subagency: _____

Contact name/title: _____

Contact organization: _____

Contact telephone: _____

Contact email: _____

1. *Overall*, how would you rate the quality of the *applicants* for your agency/ subcomponent's vacancies?

___ Excellent ___ Very Good ___ Good ___ Fair ___ Poor

2. *Overall*, how would you rate the quality of the *new employees* selected by your agency/subcomponent?

___ Excellent ___ Very Good ___ Good ___ Fair ___ Poor

3. *Overall*, how knowledgeable would you say the following are about the assessment tools available for measuring applicant qualifications?

Agency Selecting Officials:
___ Very knowledgeable
___ Knowledgeable
___ Somewhat knowledgeable
___ Not very knowledgeable

Agency Human Resources Staff:
___ Very knowledgeable
___ Knowledgeable
___ Somewhat knowledgeable
___ Not very knowledgeable

4. Who is *typically* responsible for deciding what assessment tools will be used to measure applicant qualifications? (select all that apply)

___ Agency HR staff
___ Agency Industrial/Organizational (IO) Psychologist or other testing staff
___ Selecting official(s)
___ Subject matter expert(s) or line staff
___ Office of Personnel Management
___ Other (please specify)

5. If you marked the selecting official or subject matter expert above, how much assistance/advice does the HR office provide regarding the value of different assessment tools.

___ A Lot (e.g, HR consults regularly with selecting officials about their options)
___ Some (e.g., HR provides assistance when selecting official have questions)
___ Little (e.g., HR does not have the resources to provide regular consultation)

6. In the past year, which of the following assessment methods has your agency/ subcomponent used to assess applicant qualifications? (select all that apply)

___ **Structured interviews** that employ objective, pre-determined rules for eliciting, observing, and evaluating interview responses

___ **Unstructured interviews** in which questions asked may be unplanned and vary across interviews, and the results are analyzed and applied subjectively

___ **Cognitive ability tests** that estimate applicants' abilities involved in thinking (e.g., reasoning, perception, memory, verbal and math ability, and problem solving)

___ Ratings of education and experience as described through **occupational questionnaires** (e.g., QuickHire, AVUE, or USAStaffing multiple-choice questions)

___ Ratings of education and experience as described in **KSA narratives** (knowledge, skills, and abilities)

___ **Reference checks** that evaluate an applicant's past job performance using information collected from those who have prior knowledge of the applicant

___ **Biographical data** that includes questions about past events and behaviors reflecting personality attributes, attitudes, experiences, interests, skills and abilities validated as predictors of overall performance for a given occupation (e.g., ACWA Form C)

___ **Accomplishment records** in which applicants provide written descriptions of personal accomplishments that best illustrate their proficiency on critical job competencies, and a panel of trained raters use competency-based benchmarks to score the degree to which the behaviors and outcomes reflect the benchmark levels of proficiency

___ **Job knowledge tests** that measure the applicant's current knowledge of the field/job (e.g., a test that measures an applicant's knowledge of basic chemistry or accounting principles).

___ **Grade point average** (e.g., Outstanding Scholar requirement)

___ **Education level** (e.g., college graduate, graduate degree, Ph.D.)

7. In the past year, which of the following types of job simulations has your agency/ subcomponent used to assess applicant qualifications? (select all that apply)

___ **Work sample tests** that require the applicant to actually perform a critical job task to measure his/her current ability to perform a part of the job (e.g., a role-play between a customer service representative applicant and an angry customer or an x-ray technician reading an x-ray).

___ **Situational judgment tests** that measure how an applicant might react to a particular situation (e.g., showing the applicant a video of a specific situation and asking them a variety of questions that would demonstrate how they might react in that situation).

___ **Assessment centers** in which applicants' performance is observed in several job-related exercises, such as leaderless group discussions and business games (e.g., the assessment procedure for Presidential Management Fellows).

___ **Job try-out procedures** in which an applicant is hired with minimal screening, evaluated over a specified period of time to determine if he/she is a good fit for the job, and is *actively separated or retained on the basis of performance* (as opposed to a typical probationary or trial period in which finalizing the applicant's appointment is fairly standard procedure).

___ **We have not used job simulations in the past year (Go to Question 20)**

___ **Other** type(s) of job simulation. Please list.

For agencies that have used a job simulation in the past year, please answer the following questions in regard to the simulations administered in the past year. For agencies that have not used a job simulation in the past year, please go to Question 20.

8. For approximately what percentage of your organization's individual vacancies has a job simulation assessment been used in the past year (count open continuous vacancies by the number of times closed during the year and include Presidential Management Fellow selections)?

___ Less than 10% ___ 11-25% ___ 26-50% ___ 51-75% ___ More than 75%

9. For what types of positions does your organization use job simulations more frequently?

___ Merit Promotion ___ Delegated Examining ___ Both equally

10. For what occupations and grades have you used these job simulation tests?

Occupation   General Schedule Grade (or equivalent)

_____

_____

_____

_____

_____

_____

11. Please list the competencies/skills typically measured by the job simulations (e.g., writing ability, customer service, technical competencies).

_____

_____

_____

_____

_____

_____

_____

_____

12. For which purpose is the job simulation most used? (select only one)

___ To determine minimum qualifications

___ To identify relative qualifications, after minimum qualifications have been determined

___ To distinguish among the best qualified applicants, after relative qualifications have been determined

___ It is used at different times throughout the process, depending on the vacancy

___ Other (please specify)

13. The last time your organization used a job simulation, was it part of a multiple hurdle approach (i.e., using multiple assessments in a certain order, each designed to measure different skills to further narrow the applicant pool)?

___ Yes   ___ No

14. *Typically*, **who is responsible for developing the job simulation test (select all that apply)?**
___ Agency HR staff
___ Agency IO Psychologists or other professional testing staff
___ Selecting official(s)
___ Subject matter expert(s) or line staff
___ Office of Personnel Management
___ We use commercial-off-the-shelf program
___ We procure the services of a contractor to develop
___ Other _____

15. *Typically*, **who is responsible for administering the job simulation test (select all that apply)?**
___ Agency HR staff
___ Agency IO Psychologists or other professional testing staff
___ Selecting official(s)
___ Subject matter expert(s) or line staff
___ Office of Personnel Management
___ Contracted staff
___ Other _____

16. How much training is generally provided to those who administer the simulations?
___ 5 days or longer
___ 2-4 days
___ 1 day or less
___ No training
___ I don't know

17. In your opinion, have job simulation tests increased the quality of your agency's selections?

___ Yes   ___ No   ___ In some situations but not others

18. Does your agency measure how effectively job simulations increase the quality of your selections?

___ Yes   ___ No

If yes, briefly describe how:

19. Please describe briefly how your agency determines if it will use a job simulation to assess applicants.

20. MSPB would like to highlight best practices in the area of job simulations. Please describe any job simulation methods you use that you feel have been especially effective.

21. What obstacles have you faced in using job simulations OR what factors have kept you from using job simulations? (select all that apply)

___ The agency's jobs do not lend themselves well to this type of assessment
___ We use other assessments that we feel better differentiate among candidates
___ The agency does not have the resources to develop and/or administer
___ The agency does not have the expertise to develop and/or administer
___ The agency does not have the time required to develop assessments
___ The agency does not have the time required to train those who administer the assessments
___ The HR staff does not have enough knowledge about these types of assessments to determine if they would be beneficial
___ Selecting officials do not have enough knowledge about these types of assessments to determine if they would be beneficial
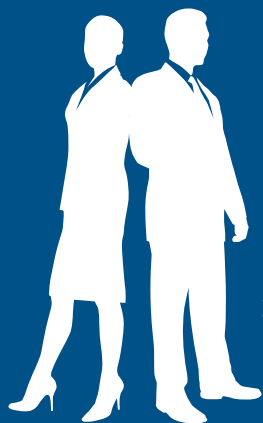___ Other (please list)

Thank you for your participation!
You may email your response to laura.shugrue@mspb.gov; fax it to (202) 653-7211; or mail it to U.S. Merit Systems Protection Board, Office of Policy and Evaluation, ATTN: Laura Shugrue, 1615 M Street, NW, Washington, DC 20036

JOB SIMULATIONS:
TRYING OUT FOR A FEDERAL JOB