

## Evaluate Training and Performance Effectively, Quickly, and Inexpensively Using the Situational Judgment Test (SJT)

**Stephen Gunter, Ph.D., Ingrid Mellone, CPT**

**Camber Corporation**

**Orlando, Florida**

[jgunter@camber.com](mailto:jgunter@camber.com), [imellone@camber.com](mailto:imellone@camber.com)

**Kate Oakley, Carol Faben**

**Camber Corporation**

**Orlando, Florida**

[koakley@camber.com](mailto:koakley@camber.com), [cfaben@camber.com](mailto:cfaben@camber.com)

### ABSTRACT

Currently, governments and other organizations are operating in an environment where resources are limited for developing high-fidelity (realistic) training simulations and assessments. Yet training is still required and live training is even more expensive. Governments and businesses are being pressured to realize greater returns on their training investments, including more rigorously assessing job-related knowledge, skills, and abilities (KSAs). The pace and volume of change in job content further complicates this environment. A low-cost, reliable, and agile method of assessing the important KSAs required for job performance is needed, particularly as assessment methods also support the evaluation of training effectiveness.

This paper introduces and discusses the measurement qualities, benefits, and applications of a low-fidelity simulation, the situational judgment test (SJT). SJTs present—in a paper- or computer-based format—a range of scenarios that the individual is likely to experience on the job along with a set of possible options, and ask how the individual would respond (e.g., “How effective is each response?” or “Which response is the best/worst?”). Subject matter experts (SMEs) write the scenarios and response options, and determine the effectiveness level or quality (best/worst) for the options provided. Research findings report that SJTs have been used to measure training effectiveness, leadership skills, judgment, and interpersonal skills. SJTs are quickly developed and quickly adapted. The Veterans Benefits Administration uses SJT as a component of tests for three high-level job positions. Specifically, SJT items are used to measure core, job-related knowledge and skills via online assessments in a large-scale, Congressionally-mandated skills certification program. The item development process includes collaboration with SMEs. The SJT makes a significant contribution to measuring certification-level performance.

### ABOUT THE AUTHORS

**Stephen Gunter, Ph.D.** is an industrial/organizational psychologist with knowledge and experience in psychometrics and the design, development, and validation of employment and certification tests, especially the theory, measurement, and application of situational judgment tests. At Camber, he is responsible for the development of Skills Certification test blueprints and the accuracy and psychometric evaluation of all tests.

**Ingrid Mellone, MS, CPT** has over 20 years of experience designing, developing, and evaluating technology-based learning and performance solutions. She has a master’s in Instructional Design and Development, holds the Certified Performance Technologist (CPT) designation from the International Society for Performance Improvement (ISPI), and serves as Skills Certification Project Lead.

**Kate Oakley, MS** has 10 years of experience creating computer-based and instructor-led training courses, primarily for military contracts. She serves as a Skills Certification Test Lead and has a master’s in Instructional Design and Development.

**Carol Faben, MA, PMP** is the Program Manager for the VBA Skills Certification testing program. Before joining Camber, she spent nearly 20 years with Educational Testing Service (ETS). She is a certified project management professional (PMP) and a trained Six Sigma Black Belt. Carol also has a master’s in Teaching.

## **Evaluate Training and Performance Effectively, Quickly, and Inexpensively Using the Situational Judgment Test (SJT)**

**Stephen Gunter, Ph.D., Ingrid Mellone, CPT**

**Camber Corporation**

**Orlando, Florida**

[jgunter@camber.com](mailto:jgunter@camber.com), [imellone@camber.com](mailto:imellone@camber.com)

**Kate Oakley, Carol Faben**

**Camber Corporation**

**Orlando, Florida**

[koakley@camber.com](mailto:koakley@camber.com), [cfaben@camber.com](mailto:cfaben@camber.com)

### **INTRODUCTION**

Simulation is increasingly used in fields such as aviation, military, and medicine as an immersive method for training and assessing key knowledge and skills. High-fidelity, scenario-based simulations are used to perfect the skills necessary to perform effectively in typical, complex, and dynamic situations. These simulations present detailed environments, represent the target situations well, and provide realistic responses to employee actions. However, they are expensive, require significant resources to develop, have a long development cycle, and can quickly become obsolete due to changes in the job.

Another way to assess skills is to observe the actual behaviors in live role plays, assessment centers, or on the job. Yet, as with high-fidelity simulations, such assessments are time-consuming and expensive to develop and maintain.

If key knowledge and skills are trained using scenario-based multimedia simulations or live role plays, the measurement of that learning should also use a scenario-based method. This is especially true when the knowledge and skills trained (e.g., interpersonal skills and leadership) have no hard rules for expected behaviors, but rather have general rules, concepts, and principles that must be applied in a variety of situations (Baldwin & Ford, 1988).

Currently, governments and organizations are operating in an environment where resources are limited for developing high-fidelity training simulations and assessments. Yet training is still required and live training is even more expensive. Governments and businesses are being pressured to realize greater returns on their training investments, including more rigorously assessing job-related knowledge, skills, and abilities (KSAs). The pace and volume of change in job content further complicates this environment.

A low-cost, reliable, and agile method of assessing the important KSAs required for job performance is needed, particularly because such assessment methods also support the evaluation of training effectiveness. This paper describes the situational judgment test (SJT), a low-fidelity (paper- or computer-based format), scenario-based method for measuring key, job-related knowledge and skills when high-fidelity assessment is not feasible or desired. SJT questions (hereafter referred to as “items”) may be developed and modified quickly for a wide range of content areas and have several attractive statistical properties. The Veterans Benefits Administration (VBA) uses the SJT across three job occupations in a large-scale, Congressionally-mandated Skills Certification Program.

### **DEFINING THE SJT**

The SJT presents job-related situations that an employee or trainee (hereafter referred to as “employee”) is likely to face on the job, along with a set of possible responses. The SJT requires the employee to make a decision, based on the provided responses, about how to handle a given situation. The SJT portrays job-related situations and responses as closely as possible to real situations, with responses that the individual would or should make. Therefore, SJTs are often called “low-fidelity” simulations (Motowidlo, Dunnette, & Carter, 1990). Although SJTs can be developed to include audio and video, this paper largely focuses on the low-fidelity SJT that presents written descriptions of job-related situations and responses in a paper- or computer-based format.

## Components

To ensure alignment between job requirements and SJT items, job analysis data forms the foundation for developing SJT items. Analysts capture critical incidents encountered by employees, including a description of the incident, the behaviors exhibited by the employee or others in the incident, any tools or equipment the employee used, and the outcomes of the behaviors. Subject matter experts (SMEs) or professional item writers translate the incidents into SJT items, each comprised of a situation and several realistic possible responses. Situations range from difficult dilemmas or problems with no obvious solution, to more typical situations that employees frequently face. Situations often represent those that an employee has been trained to handle, but may also represent knowledge or a skill that is expected to be informally acquired.

After the description of the situation and before the responses, a response instruction asks an employee how to handle the situation. Typical response instructions include “Pick the best response,” “What should/would you do?”, “What are you most/least likely to do?”, “What is the best/worst response to this situation?”, and “Rate the effectiveness of each response.” The choice of which response instruction to use may be based on research, professional judgment, or employee reactions.

See Figure 1 for an example of an SJT item from Weekley, Ployhart, and Holtz (2006). See Scoring, below, for more information about the scale that would be used to rate the responses:

<p>One of the trainers assigned to your team informs you that he is scheduled for 7 days of medical leave starting next week. The trainer scheduled the medical leave months ago to allow recovery time for a minor surgical procedure. Rate the effectiveness of the following responses:</p> <ol style="list-style-type: none"><li>1. Ask your supervisor for another trainer to replace the individual going on leave.</li><li>2. Inform the trainer that he is responsible for finding a co-worker to fill in in his absence.</li><li>3. Inform your supervisor that you will not be able to train all recruiters before the season starts because you are short one trainer for an entire week.</li></ol>
--

**Figure 1. Sample SJT Situation, Response Instructions, and Responses**

## Scoring

The scoring of SJT items is based on the response instruction and on the design approach that is used. One common approach is the rational approach, where SMEs determine how responses will be scored. The “Pick the best,” “What would/should you do?”, “What are you most/least likely to do?”, and “What is the best/worst response to this situation?” response instruction designs are typically scored in a +1, 0, or -1 fashion (Weekley & Jones, 1999). For example, employees who choose the best response as rated by SMEs will receive +1 point. Those who choose a response that was not identified by the SMEs as the best or the worst will receive 0 points. Those who choose as the best response what SMEs keyed as the worst response will receive -1 point. The combinations for the worst response are scored in the same way.

Another response instruction design is the “Rate the effectiveness of each response” instruction, which is designed and developed, and therefore scored, differently from SJT items that use the response instruction strategies described above. “Rate the effectiveness” SJT items feature an effectiveness scale, commonly of 5 or 7 points, from highly ineffective to highly effective. During the item development process, SMEs review each response and rate how effective or ineffective they judge it to be using the effectiveness scale. SMEs may make their ratings independently or by consensus. If SMEs rate the responses independently, then the average of the SME ratings, rounded to the nearest whole number, is used as the keyed rating. If the consensus approach is used, SMEs discuss the effectiveness of each response and agree on a final rating that is used as the keyed rating. See Figure 2 for a sample VBA SJT item illustrating the “Rate the effectiveness” instruction.

Test Administration: | Logged in as: | [Instructions](#) [EXIT](#)

**UNITED STATES DEPARTMENT OF VETERANS AFFAIRS**  
**Veterans Benefits Administration**

Estimated Time Elapsed: **1:23:00**

**Session 1** | **Session 2** | **Background**

1 2 3 4 5 6 7 8 **[9]** 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33  
34 35 36 37 38 39 40 41 42 43 44 Review

**Situational Judgment: Rate the effectiveness of each response separately. You will not be ranking them but, rather, rating each independently of the others. This means that some responses may be equally effective or ineffective. [SJT Instructions](#)**

9. You are conducting a formal training session for the Appeals Team DROs on complications related to diabetes mellitus II. A newly promoted DRO follows up your presentation with an example that raises an area to which you don't know the answer. Rate the effectiveness of the following responses.

7 Very effective  Ask the new DRO to write up the example for your review, and state that you will research it and provide an answer with reference in an email to all the DROs.

3 Slightly ineffective  Have the new DRO research the example and have him provide his answer in an email to the other DROs in attendance.

5 Slightly effective  Open the example to the other DROs in attendance for any answers.

Ask the new DRO to provide the answer now, citing references for his example.

7 Very effective  
6 Moderately effective  
5 Slightly effective  
4 Neutral  
3 Slightly ineffective  
2 Moderately ineffective  
1 Very ineffective

Item Comments? Please enter any item comments below.

**Figure 2. Sample “Rate the Effectiveness of Each Response” SJT Item**

“Rate the effectiveness” SJT items are scored using a partial-credit scoring model, where the score is derived from how close the rating is to the keyed SME effectiveness rating. For example, if an employee rated a response a “5” and the SMEs also rated that response a “5,” the employee receives full credit (e.g., +1 point). If the employee rated the response a “4” and the SMEs rated that response a “5,” the employee receives partial credit (e.g., +0.5 point). The number of points/partial points an employee receives in the partial-credit scoring model may be based on professional judgment of how important the KSAs measured by the SJT are to the job. Points may also be awarded based on the level of expertise with KSAs that employees should have as measured by the SJT. For example, if SMEs determine that employees are expected to have a high level of interpersonal skills, SMEs might select a relatively restrictive scoring model that awards fewer points the farther a person rates a response from the SMEs’ keyed rating.

## **BENEFITS OF SJTS FOR ASSESSMENT**

A key benefit of SJTs is their short development cycle. Multimedia simulations and simulators are very time- and cost-intensive to develop and maintain (Dragow, Olson-Buchanan, & Moberg, 1999). For example, Dragow et al. (1999) report that the development of a video-intensive conflict resolution skills assessment took nearly 2 years because of the many iterations needed to fine tune the dialog in the scripts. As another example, Burroughs, Murray, Wesley, Medina, Penn, Gordon, and Catello (1999) state that their fully computerized assessment center, containing an in-basket simulation (in which a set of tasks is provided in a simulated “in-basket”) and a structured interview (in which multiple interviewees are asked the same questions in the same order), took 18 months to develop. In contrast, lower fidelity, written SJTs do not require scripts, intricate software development for animation or video

deployment, or intensive software testing. Therefore, SJTs can easily be developed and implemented within a matter of months. In addition, SJTs can work well in agile environments because SJTs may easily be modified to reflect changes in the job or training, because the developers need not create video, audio, or animated scenes. Thus, the quick development and revision cycles are beneficial in the measurement of job-related KSAs in industries where training needs to be rapidly deployed or changes are frequent.

The research base supporting SJTs is equally compelling, although the authors note a lack of recent data about relevant reliability studies. First, because SJTs are derived from job analysis, they can be demonstrated to measure job-related KSAs directly, which makes them valid. Second, they are reliable and predict performance (McDaniel, Morgeson, Finnegan, Campion, & Braverman, 2001). For example, McDaniel et al. (2001) performed a meta-analysis on the predictive ability of SJTs. They found an acceptable average internal consistency reliability estimate of .77 and an average, uncorrected correlation with job performance of .26. These findings are typical of those in the literature for other tests and compare favorably to individual studies of multimedia SJTs (Chan & Schmitt, 1997; Olson-Buchanan, Drasgow, Moberg, Mead, Keenan, & Donovan, 1998; Weekley & Jones, 1997; 1999), which have found reliability estimates ranging from .32 to .73 and correlations with job performance, on average, ranging from .20 to .30.

## **SJT APPLICATIONS**

In addition to the benefits of reasonable cost to develop, validity, and reliability, SJTs can be used for a range of purposes, jobs, and job-related KSAs because they are constructed based on the content of the job.

SJTs are typically used for selection purposes; for example, the Federal Bureau of Investigation (FBI) currently uses an SJT as part of its selection process for Special Agents (FBI, n.d.). In addition, SJTs can be used for training needs analysis, as training tools (Fritzsche, Stagl, Salas, & Burke, 2006), and to evaluate training effectiveness (Hauenstein, Findlay, & McDonald, 2010) and decision making (e.g., Military Decision Making Process (MDMP)).

Since the 1940s, commercially-published SJTs such as *How Supervise?* (File, 1945; File & Remmers, 1948) and the *Supervisory Practices Test* (Bruce & Learner, 1958) have been used for supervisory positions. More recently, SJTs have been used for a wide variety of jobs, including insurance agents (Dalessio, 1994), managers and marketing positions in the telecommunications industry (Motowidlo & Tippins, 1993; Olson-Buchanan et al., 1998), sales personnel (Phillips, 1992), employees in manufacturing (Stevens & Campion, 1999), and retail employees (Weekley & Jones, 1999).

The range of KSAs that SJTs have been used to measure is also quite broad. SJTs have been used to measure job knowledge (Clevenger, Pereira, Wiechmann, Schmitt, & Schmidt-Harvey, 2001), problem solving and teamwork (Banki & Latham, 2010), overall ability to adapt to work situations (Chan & Schmitt, 2002), interpersonal skills (O'Connell, Hartman, McDaniel, Grubb, & Lawrence, 2007; Weekley & Jones, 1997), and personality characteristics (Motowidlo, Hooper, & Jackson, 2006).

Next, we discuss selected examples of the applicability of SJTs and the KSAs they have been used to measure.

### **Evaluating Training**

Hauenstein, Findlay, and McDonald (2010) utilized an SJT to evaluate a cross-service training program for military Equal Opportunity Advisors (EOAs). The authors focused the content of the SJT on formal and informal complaint processing procedures and had SMEs develop the SJT, keeping in mind that test items had to measure the KSAs trained in the curriculum. They conducted the study with two sets of participants. The first set included 46 EOAs and the second included 55. The authors used a pretest-posttest design for the first set, wherein the SJT was administered online 16 weeks apart, before and after the EOA training. The second set of EOAs completed the SJT prior to beginning training, and twice more after the completion of two parts of the training curriculum: 7 weeks after the pretest, once they completed the core EOA training curriculum; and 10 weeks after the pretest, once they completed the service-specific (e.g., Army, Navy, Marine) training curriculum. With both EOA groups, posttest scores on the SJT were significantly higher than pretest scores.

## **Measuring Judgment**

SJTs have also been applied in environments that are dominated by high-fidelity simulations and flight simulators (Fritzsche, et al., 2006). Specifically, Hunter (2003) constructed an SJT to measure aviation judgment and predict flight safety. Items were constructed from critical incidents that were gathered from pilot and accident reports. Several themes emerged from these incidents, including weather phenomena, mechanical malfunctions, and air traffic control requests, and formed the content areas of the SJT. A volunteer sample of 467 pilots who had an average of 11 years of experience and predominantly flew private and commercial planes participated in the study. Hunter hypothesized that good judgment, as measured by the SJT, should be correlated with fewer experienced hazardous events and accidents. As hypothesized, SJT scores significantly correlated (-.215) with the number of hazardous events and accidents experienced by the convenience sample of participating pilots—higher SJT scores correlated with fewer accidents.

## **Measuring Adaptability of Leaders**

Hedlund, Forsythe, Horvath, Williams, Snook, and Sternberg (2003) developed an SJT to measure intrapersonal, interpersonal, and organizational knowledge of U.S. Army officers. Their SJT measured tacit knowledge—the ability to learn from and solve everyday problems in order to adapt—which is acquired through experiential learning that occurs largely outside of formal training (Sternberg, 1988; 1997). Hedlund et al. developed an instrument measuring tacit military leadership knowledge and administered it to 368 platoon leaders, 163 company commanders, and 31 battalion commanders.

These investigators compared scores on the SJT instrument with ratings of leadership effectiveness from peers or superiors. The authors found that the SJT scores correlated significantly with superior's ratings (.17) for platoon leaders, peer ratings (.18) for company commanders, and superior's ratings (.46) for battalion commanders.

## **Predicting Team Performance**

McClough and Rogelberg (2003) and Morgeson, Reider, and Campion (2005) examined the ability of an SJT to predict team-based performance. Both used Stevens and Campion's (1999) Teamwork Knowledge Skill and Ability SJT, which measures interpersonal and self-management skills. The McClough and Rogelberg study included 227 undergraduates, in teams of 3 to 5 people, and had them plan and develop a new, fictitious project. Morgeson et al. included 90 steel mill employees, in teams of 5 to 10 people, who rotated their team-based jobs and shared workload to complete several production-oriented tasks. In both studies, participants completed the teamwork SJT during class or working hours.

- McClough and Rogelberg trained raters to rate the performance of each team member during the business simulation on the dimensions of conflict resolution, collaborative problem solving, communication, performance management, and task coordination. They also had team members rate each of their peer's performance on selected dimensions such as participation and interpersonal skills. The authors found that performance on the teamwork SJT correlated significantly with teamwork ratings (.31) and peer ratings (.34).
- Morgeson et al. had team members' supervisors rate team members on tasks (e.g., efficiently performing one's duties and planning and organizing work) and contextual performance (e.g., interpersonal facilitation and helping and individual initiative). Like McClough and Rogelberg, these investigators also found that the teamwork SJT correlated significantly with task (.36) and contextual performance (.32).

## **Applications Summary**

Each SJT selected for discussion in this section was developed using methods that yielded a content valid instrument. In addition, each study demonstrated the SJT was predictive of knowledge attainment and important performance criteria such as overall leadership performance, team performance, or hazardous events experiences.

## **SJT IN PRACTICE**

VBA's Skills Certification Program is a large-scale system of legally defensible, criterion-referenced assessments of individual job knowledge/skill. "Legally defensible" means the tests are designed and developed in a way that would stand up to a court challenge. "Criterion-referenced" means the employee must demonstrate a defined level of knowledge of the job domain. This is contrasted with norm-referenced tests, which compare the performance of employees with one another. Thus, when VBA employees are certified, they have demonstrated they are qualified to perform the essential duties of their job occupation.

VBA began developing the Congressionally-mandated Skills Certification Program in 2003 to improve the organizational performance and professionalism of claims processing personnel. The program is comprised of over half a dozen tests for technical and managerial positions. Each test has a Design Team that implements VBA's governance policies. Teams of experienced subject matter experts (SMEs) write and review the items for each test in regularly scheduled, facilitated workshops. Each year over a dozen test administrations are offered online, worldwide. The number of employees tested ranges from approximately 175 to over 1,000 per test event, or administration. VBA uses the results of testing to provide insight into workforce capability, to support employee development, and as input into the strategic planning of training.

The use of SJTs within this certification program is appropriate for several reasons. First, employees must manage a wide range of interpersonal and workload management situations, and how to handle these situations is often not documented or prescribed. Second, the job occupations change often to reflect new regulations and organizational initiatives. Third, adequate money and time to develop high-fidelity simulations are not available. Fourth, the high-stakes nature of the program (a promotion or other action may be riding on the outcome) necessitates the items be reviewed and revised based on data from test administrations to ensure they possess high-functioning psychometric characteristics.

### **SJT Audiences**

SJTs are used to measure interpersonal, leadership, self-management, and workload management skills in three rather disparate, advanced positions within VBA. The first position is a Decision Review Officer (DRO), who is a technical expert in claims processing and appeals. The DRO works with other employees to process Veterans' disability claims. The second position is a Coach (i.e., first-line supervisor), who manages the quality and productivity of claims processing teams. The third position is a Vocational Rehabilitation Counselor (VRC), who works directly with disabled Veterans to provide employment and independent living services. SJTs are administered to DROs and Coaches who have at least 2 years of experience in the position and VRCs who typically have 3 or more years of experience.

### **SJT Development/Revision Process**

Test Leads (Instructional Systems Designers (ISDs)) facilitate SJT item writing by VBA SMEs. Test items are written to support KSAs that appear on the test blueprint (test plan). These items undergo a rigorous review and validation process, receive a cut score (required passing standard), and are piloted prior to being included as a scored item on a live test (see Figure 3).

A large volume of data is collected and used to evaluate the effectiveness of the SJT items in VBA's program (Figure 3, "Item Performance Data Acceptable?" decision points). Data about item performance (e.g., difficulty, response patterns) is analyzed. Items that are not functioning well (i.e., at identifying those who have certification-level skills) are removed from the active item bank and subsequently reviewed and revised by SMEs. These items are then repiloted on another test administration to ensure the revisions actually improved the items. Such rigorous evaluation and revision would be very difficult and costly with a high-fidelity simulation.

Employee reactions provide another source of item evaluation data, and we have found the novelty of SJT items for employees could be an important issue. If unfamiliarity causes employees to answer the SJT items in unreliable ways, they may not perform well on SJTs even though they possess acceptable levels of skills. Indeed, in the case of the VBA SJTs, DROs, Coaches, and VRCs typically react by saying they think the items are very job-related, but that they are not familiar with the item type. However, the data reveals reliability estimates for these tests that is

consistent with those cited above by McDaniel et al. (2001). In addition, response profiles (i.e., the match between the average employee response compared to the SMEs' consensus effective rating) show that employees understand SJTs well despite their lack of experience with them.

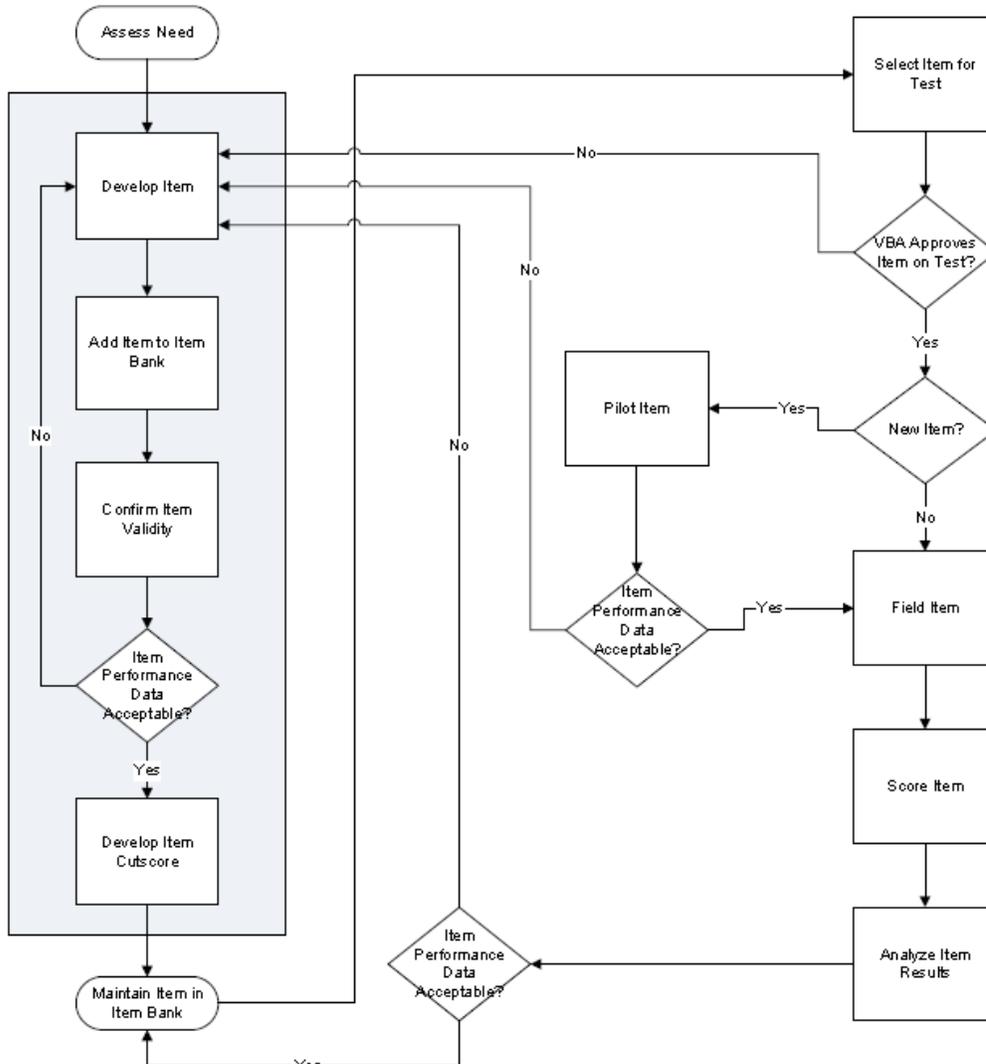


Figure 3. Test Item Development/Evaluation/Revision Cycle

## SUMMARY

SJTs provide comparable reliability to more expensive simulation solutions, but for an investment of less time and money. The VBA's use of the SJT in the Skills Certification Program highlights the benefits, adaptability, and data-driven nature of SJTs. These characteristics can be applied to any certification program, advanced position, or training intervention where job-related KSAs can be measured with scenario-based items. Also, the SJT can be used for many different purposes, such as selection, certification, or training evaluation. As the Federal government comes under increasing pressure to expend its limited training funds more efficiently, the SJT provides an important measurement mechanism.

## ACKNOWLEDGEMENTS

The authors wish to express their appreciation to the Veterans Benefits Administration for enabling them to share these findings with other organizations. The authors also thank Camber Corporation for its support in the preparation of this paper.

## REFERENCES

- Baldwin, T. T., & Ford, J. K. (1988). Transfer of training: A review and directions for future research. *Personnel Psychology, 41*, 63-105.
- Banki, S., & Latham, G. P. (2010). The criterion-related validities and perceived fairness of the situational interview and the situational judgment test in an Iranian organization. *Applied Psychology: An International Review, 59*, 124-142.
- Bruce, M. M., & Learner, D. B. (1958). A supervisory practices test. *Personnel Psychology, 11*, 207-216.
- Burroughs, W. A., Murray, J., Wesley, S. S., Medina, D. R., Penn, S. L., Gordon, S. R., & Catello, M. (1999). Easing the implementation of behavioral testing through computerization. In F. Drasgow & J. B. Olson-Buchanan (Eds.), *Innovations in computerized assessment* (221-247). Mahwah, NJ: Lawrence Erlbaum.
- Chan, D. & Schmitt, N. (1997). Video-based versus paper-and-pencil method of assessment in situational judgment tests: Subgroup differences in test performance and face validity perceptions. *Journal of Applied Psychology, 82*, 143-159.
- Chan, D., & Schmitt, N. (2002). Situational judgment and job performance. *Human Performance, 15*, 233-254.
- Clevenger, J., Pereira, G. M., Wiechmann, D., Schmitt, N., & Harvey, V. S. (2001). Incremental validity of situational judgment tests. *Journal of Applied Psychology, 86*, 410-417.
- Dalessio, A. T. (1994). Predicting insurance agent turnover using a video-based situational judgment test. *Journal of Business and Psychology, 9*, 23-32.
- Drasgow, F., Olson-Buchanan, J. B., & Moberg, P. J. (1999). Development of an interactive video assessment: Trials and tribulations. In F. Drasgow & J. B. Olson-Buchanan (Eds.), *Innovations in computerized assessment* (177-196). Mahwah, NJ: Lawrence Erlbaum.
- Federal Bureau of Investigation. (n.d.). Phase I testing. Retrieved from <https://www.fbijobs.gov/1121.asp>
- File, Q. W. (1945). The measurement of supervisory quality in industry. *Journal of Applied Psychology, 29*, 381-387.
- File, Q. W., & Remmers, H. H. (1948). *How Supervise? manual 1948 revision*. Cleveland, OH: Psychology Corporation.
- Fritzsche, B. A., Stagl, K. C., Salas, E., & Burke, C. S. (2006). Enhancing the design, delivery, and evaluation of scenario-based training: Can situational judgment tests contribute? In J. A. Weekley & R. E. Ployhart (Eds.), *Situational judgment tests: Theory, measurement, and application* (301-318). Mahwah, NJ: Lawrence Erlbaum.
- Hauenstein, N. M. A., Findlay, R. A., McDonald, D. P. (2010). Using situational judgment tests to assess training effectiveness: Lessons learned evaluating military equal opportunity advisor trainees. *Military Psychology, 22*, 262-281.
- Hedlund, J., Forsythe, G. B., Horvath, J. A., Williams, W. M., Snook, S., & Sternberg, R. J. (2003). Identifying and assessing tacit knowledge: Understanding the practical intelligence of military leaders. *The Leadership Quarterly, 14*, 117-140.
- Hunter, D. R. (2003). Measuring general aviation pilot judgment using a situational judgment technique. *The International Journal of Aviation Psychology, 13*, 373-386.
- McClough, A. C., & Rogelberg, S. G. (2003). Selection in teams: An exploration of the teamwork knowledge, skills, and ability test. *International Journal of Selection and Assessment, 11*, 56-66.
- McDaniel, M. A., Morgeson, F. P., Finnegan, E. B., Campion, M. A., & Braverman, E. P. (2001). Use of situational judgment tests to predict job performance: A clarification of the literature. *Journal of Applied Psychology, 86*, 730-740.
- Morgeson, F. P., Reider, M. H., & Campion, M. A. (2005). Selecting individuals in team settings: The importance of social skills, personality characteristics, and teamwork knowledge. *Personnel Psychology, 58*, 583-611.
- Motowidlo, S. J., Dunnette, M. D. & Carter, G. W. (1990). An alternative selection procedure: The low-fidelity simulation. *Journal of Applied Psychology, 75*, 640-647.

- Motowidlo, S. J., Hooper, A. C., & Jackson, H. L. (2006). Implicit policies about relations between personality traits and behavioral effectiveness in situational judgment items. *Journal of Applied Psychology, 91*, 749-761.
- Motowidlo, S. J., & Tippins, N. (1993). Further studies of the low-fidelity simulation in the form of a situational inventory. *Journal of Occupational and Organizational Psychology, 66*, 337-344.
- O'Connell, M. S., Hartman, N. S., McDaniel, M. A., Grubb, W. L., & Lawrence, A. (2007). Incremental validity of situational judgment tests for task and contextual job performance. *International Journal of Selection and Assessment, 15*, 19-29.
- Olson-Buchanan, J. B., Drasgow, F., Moberg, P. J., Mead, A. D., Keenan, P. A., & Donovan, M. A. (1998). Interactive video assessment of conflict resolution skills. *Personnel Psychology, 51*, 1-24.
- Phillips, J. F. (1992). Predicting sales skills. *Journal of Business and Psychology, 7*, 151-160.
- Sternberg, R. J. (1988). *The triarchic mind: A new theory of human intelligence*. New York: Penguin Books.
- Sternberg, R. J. (1997). *Successful intelligence*. New York: Penguin Books.
- Stevens, M. J., & Campion, M. A. (1999). Staffing work teams: Development and validation of a selection test for teamwork settings. *Journal of Management, 25*, 207-228.
- Weekley, J. A., & Jones, C. (1997). Video-based situational testing. *Personnel Psychology, 50*, 25-49.
- Weekley, J. A., & Jones, C. (1999). Further studies of situational tests. *Personnel Psychology, 52*, 679-700.
- Weekley, J. A., Ployhart, R. E., & Holtz, B. C. (2006). On the development of situational judgment tests: Issues in item development, scaling, and scoring. In J. A. Weekley & Ployhart, R. E. (Eds.) *Situational judgment tests: Theory, measurement, and application* (157-182). San Francisco, CA: Jossey-Bass.