

A Tradition of Stewardship  
A Commitment to Service

## Situational Judgment Tests: A Testing Format That Is Almost Too Good To Be True!

PTC/SC  
May 26, 2010

Michael A. Willihnganz

SJT Presentation



## 2010 IPAC Conference


---



July 18<sup>th</sup> through the 21<sup>st</sup>  
Hyatt Regency Newport Beach


Visit [IPACweb.org](http://IPACweb.org)  
For program and registration  
information

---




## Presentation Agenda

1. Define "situational judgment test (SJT)"
2. Trace the history of SJTs
3. Discuss advantages of written SJTs
4. Summarize the validity evidence
5. Explain the various scoring approaches
6. Present a 5-step developmental process




## What Exactly is a Situational Judgment Test?

- Any paper-and-pencil test designed to measure judgment in work settings (McDaniel, Morgeson, Finnegan, Campion, & Braveman, 2000).
- A testing format that presents applicants with hypothetical job-related scenarios and asks them to identify an appropriate response from a list of alternatives (Peeters & Lievens, 2005).



## What Exactly is a Situational Judgment Test?

- Simulations based on the assumption that one can predict how well an individual may perform on a job based on how the individual performs on a *simulation* of the job (McDaniel & Nguyen, 2000).
- A hybrid selection procedure that takes on some of the characteristics of job knowledge tests as well as some of the characteristics of work sample tests (Heneman & Judge, 2006).



## What Exactly is a Situational Judgment Test?

- A measurement method typically comprised of a series of job-related situations or scenarios that describe a dilemma or problem requiring the application of relevant knowledge, skills, and abilities (KSAs) to solve (Christian, 2008).
- SJTs are also referred to as low-fidelity simulations (Motowidlo, Dunnette, & Carter, 1990).



## Fidelity of Simulations

- Simulations vary in the *fidelity* with which they present a task stimulus and elicit a response.
- High-fidelity simulations show a very close correspondence with job tasks.
- High-fidelity simulations are excellent predictors of job performance.



## Examples of High-Fidelity Simulations

- In-Basket Exercise
- Role Play Exercise
- Data Entry Test



## Decrease in Fidelity

- Fidelity decreases as stimulus materials and responses become less and less exact approximations of actual job stimuli and responses.
- At the lower end of the fidelity continuum are simulations that simply present a candidate with a *hypothetical situation* to which the candidate indicates how he/she *would respond*, rather than carrying out the intended action.
- SJTs are low-fidelity simulations.



## Low-Fidelity Task Stimulus

- Interview (i.e., *situational interview questions*)
- DVD or Video
- Written (paper-and-pencil or CBT)



## Low-Fidelity Simulations

- Responses can follow either an open-ended format, in which the candidate describes how he/she would handle the problem situation in his/her own words, or responses can follow a multiple-choice format.
- As a written low-fidelity simulation, SJTs simply present a written description of a hypothetical work situation which asks the candidate to indicate how he/she *would* handle the situation.



## History of SJTs

- The use of SJTs dates back to the 1920s.
- One of the first SJTs called the George Washington Social Intelligence Test measured judgment in social situations.
- During World War II, army psychologists assessed the judgment of soldiers using the SJT model.
- Starting in the 1940s, a number of SJTs were developed to measure supervisory potential (e.g., *Practical Judgment Test*, *How Supervise?*, and *Supervisory Practices Test*).



## History of SJTs

- In the late 1950s and early 1960s SJTs were used by large organizations as part of selection test batteries designed to predict managerial success.
- Standard Oil Company of New Jersey developed and used an SJT called the Management Judgment Test.
- Renewed interest in SJTs followed the publication of an article by Motowidlo et al. (1990) entitled *An Alternative Selection Procedure: The Low-Fidelity Simulation*.



## Sample Item

*You are at the hors d'oeuvre table placing hot chicken wings onto your cocktail plate when you accidentally drop a chicken wing into the martini of another conference attendee in the hors d'oeuvre line. The attendee does not notice that there is now a chicken wing floating next to the green olive in her martini.*

**You would . . .**

- Walk away from the hors d'oeuvre table before the attendee notices the floating chicken wing.
- Go directly to the bar and purchase another martini for the attendee before she notices the chicken wing.
- Apologize for your accident and offer to buy the attendee another martini.
- Intentionally spill the attendee's martini before she notices the chicken wing and then offer to buy her another drink.



## Sample Item

*You supervise an employee who is a chronic complainer. This employee frequently expresses his displeasure with work procedures. The employee's complaints have become disruptive to the work unit you supervise.*

**You would . . .**

- Counsel the employee on seeking a position in the organization that is more to his liking.
- Assign work to the employee that he will find enjoyable and rewarding.
- Talk to the employee about the issues causing his displeasure and your expectations for acceptable behavior.
- Be tolerant of the employee's complaints since it is not possible to change someone's personality.



## Sample Item

*You approach a suspect at his residence to serve an arrest warrant. As the suspect sees you coming toward him, he becomes increasingly agitated and verbally abusive. Which one of the following actions should you take FIRST to effect the arrest?*

- Draw your firearm and aim it at the suspect.
- Place a control hold on the suspect.
- Spray pepper spray in the suspect's face.
- Command the suspect to raise his hands above his head.



## Sample Item

*You are processing paperwork for a customer who is applying for a license. While you are completing the paperwork, the customer becomes impatient and tells you that you are working too slowly.*

**You would . . .**

- Apologize to the customer and work as quickly as possible to complete the paperwork.
- Suggest to the customer that she come back when she has more time.
- Ask the customer to please refrain from making rude comments to you.
- Inquire as to whether the customer would like to speak to your supervisor for quicker service.



## Popularity of SJTs

SJTs are becoming increasingly popular for several reasons:

- Large-scale studies have shown that SJTs have significant criterion-related validity (McDaniel et al., 2001).
- SJTs possess incremental validity over and above cognitive ability and personality tests (e.g., Chan & Schmitt, 2002).



## Popularity of SJTs

- Applicants respond enthusiastically to SJTs because they perceive SJTs to be related to the target jobs for which they are applying (e.g., Ployhart & Ryan, 1998).
- SJTs show less adverse impact on minorities than traditional cognitive ability tests (Clevenger et al., 2001).



## Advantages of SJTs

Unlike higher fidelity simulations (e.g., work samples, assessment centers), SJTs offer the convenience of mass administration, increased objectivity, reliability, standardization, face validity, and lower cost of administration (e.g., Motowidlo et al., 1990; Weekley & Jones, 1999).



## Advantages of Written SJTs

- Inexpensive to develop
- Transportable
- Ease of administration
- High candidate acceptance
- Relatively strong validity



## Validity Evidence

- Motowidlo et al. (1990) reported validity coefficients ranging from .28 to .37, with supervisory ratings serving as the criterion measure.
- Motowidlo, Hanson, and Crafts (1997, p. 248) have stated that researchers generally report significant relationships between written SJTs and ratings of job performance with correlations ranging from .20 to about .50.



## Validity Evidence

- A meta-analytic review of SJT validity studies found the mean validity of SJTs to be .34 (McDaniel et al., 2000).
- Hypothetical work behavior can predict performance without the expense of props, role players, and equipment typically needed by high-fidelity simulation tests (i.e., work sample tests), or the high-tech gadgetry of other SJT formats (e.g., DVD-based tests).



## SJT Assumptions

- Intentions are related to actual behavior.
- A person's behavior in certain kinds of situations in the past can predict how he/she is likely to respond to similar situations in the future.



## Faking and SJTs

- Faking on a selection measure can be defined as an individual's conscious distortion of responses to score favorably (e.g., McFarland & Ryan, 2000).
- Haas and McDaniel (1999) found that fakers improved their SJT scores by one-half standard deviation where as Juraska and Drasgow (2001) concluded that SJTs were not fakable.



## Faking and SJTs

- In a recent study that examined the fakability of an SJT of college students' performance, Peeters and Lievens (2005) found that faking negatively affected the criterion-related validity of the SJT.
- These results suggest that faking might be a possible threat to the use of SJTs in high-stakes testing programs.



## What Can SJTs Measure?

- The commonly used SJT format lends itself especially well to measuring various forms of job knowledge.
- SJTs may also be used to measure specific personality or ability variables (Motowidlo et al., 1997, p. 246).



## Scoring Approaches

### Dichotomous Scoring

- The alternative chosen as the **Best** response or the **Most Likely** response is scored as 1 if it is the best response, and scored as 0 if it is one of the other incorrect alternatives.
- This is the traditional dichotomous scoring approach used with most multiple-choice exams.



## Scoring Approaches

### Dichotomous Scoring with Most Likely & Least Likely

- Candidate responds twice to each presented scenario.
- The alternative chosen as the **Most Likely** response is scored as 1 if it is the best response, and scored as 0 if it is the worst response or one of the other alternatives.
- The alternative chosen as the **Least Likely** response is scored as 1 if it is the worst response, and scored as 0 if it is the best response or one of the other alternatives.



## Sample Item

*You are at the hors d'oeuvre table placing hot chicken wings onto your cocktail plate when you accidentally drop a chicken wing into the martini of another conference attendee in the hors d'oeuvre line. The attendee does not notice that there is now a chicken wing floating next to the green olive in her martini.*

*You would . . .*

- Walk away from the hors d'oeuvre table before the attendee notices the floating chicken wing.
- Go directly to the bar and purchase another martini for the attendee before she notices the chicken wing.
- Apologize for your accident and offer to buy the attendee another martini.
- Intentionally spill the attendee's martini before she notices the chicken wing and then offer to buy her another drink.

1. *Most Likely* \_\_\_\_\_ 2. *Least Likely* \_\_\_\_\_



## Scoring Approaches

### Weighted-Response With Most Likely & Least Likely

- The alternative chosen as the **Most Likely** response is scored as 1 if it is the best response, -1 if it is the worst response, and 0 if it is one of the other alternatives.
- The alternative chosen as the **Least Likely** response is scored as 1 if it is the worst response, -1 if it is the best response, and 0 if it is one of the other alternatives.



## Scoring Approaches

### Weighted-Response Scoring Example

- A score of -2 means that the candidate would **Most Likely** respond to the situation by choosing the **worst** alternative, and would **Least Likely** choose the **best** alternative.
- A score of +2 means that a candidate would **Most Likely** respond to the situation by choosing the **best** alternative, and **Least Likely** choose the **worst** alternative.



## Developmental Process

- Conduct a job analysis
- Step I -- Develop critical incidents
- Step II -- Edit/refine the problem situations
- Step III -- Develop response alternatives
- Step IV -- Develop scoring key
- Step V -- Establish a pass point



## Step I Develop Critical Incidents

- Start with the job analysis to identify appropriate content to assess.
- Through the observation of incumbents, the collection of work samples, and/or direct input from SMEs, develop the critical incidents within the domain being assessed.
- The critical incidents or problem situations should represent events that actually occur on the job.



## Domain to be Assessed: Supervision

- Attendance
- Discipline
- Performance
- Time management
- Implementation of new policy or procedure
- Training and development
- Workload prioritization
- Organizational change
- Other \_\_\_\_\_



## Step I Develop Critical Incidents

- Start with the job analysis to identify appropriate content to assess.
- Through the observation of incumbents, the collection of work samples, and/or direct input from SMEs, develop the critical incidents within the domain being assessed.
- The critical incidents or problem situations should represent events that actually occur on the job.



### Step I Develop Critical Incidents

- The critical incidents should represent problems or issues that incumbents must handle effectively or their job performance will suffer.
- The critical incidents should be complex enough to allow for meaningful differences in how they can be handled.
- The critical incidents should be described in enough detail to provide the cues necessary to distinguish more effective from less effective approaches to dealing with them.



### Step I Develop Critical Incidents

**Assemble a group of SMEs (i.e., incumbents and supervisors) to develop the critical incidents. Each critical incident should include the following elements:**

- Background information and details of a situation or problem encountered by a job incumbent.
- A description of effective action to address the situation or problem.
- A description of ineffective or inappropriate action to address the situation or problem.



### Step II Edit/Refine the Problem Situations

- From the critical incidents developed by the SMEs, select a representative sample which will adequately assess the domain.
- Ensure that the final inventory of problem situations covers *important* problems likely to be encountered on the job.
- Edit/refine each problem situation into a standard format that describes the problem clearly and concisely in just a few sentences.



### Step III Develop Response Alternatives

- Response alternatives can be developed from the descriptions of effective and ineffective actions that were identified when the critical incidents were prepared.
- Response alternatives should represent differing *plausible* strategies for the handling the problem situation.
- The more correct alternatives should be more attractive to candidates with the best potential for job success.



### Step III Develop Response Alternatives

**Each response alternative should meet the following criteria:**

- Focus on a single action or response
- Be stated in a straight-forward, understandable manner
- Be a *plausible* and *reasonable* response to the problem situation
- Discriminate between the better qualified and less qualified candidates



### Step IV Develop Scoring Key

- A scoring key is developed by collecting judgments from SMEs about the effectiveness of the alternative response options for handling each problem situation.
- *Weighted-response scoring* – SMEs identify the best response and the worst response.
- *Dichotomous scoring* – SMEs identify the best or most appropriate response.



## Step V *Establish a Pass Point*

- SMEs review and evaluate each problem situation using a modified Angoff approach to establish the minimal acceptable competency (MAC) level for the test.
  - The MAC level becomes the starting point for establishing the pass point for the SJT.
- 



# Questions??

---