



# So You Want to Learn How to Develop a Situational Judgment Test (SJT)?

Tutorial Presented at IPAC 2018 Conference

**Presenters:** Taylor Sullivan, PhD, HumRRO  
Will Taylor, PhD, HumRRO  
Angela Lee, PhD, HumRRO

July 30, 2018

# Agenda

---

- What is a Situational Judgment Test (SJT)?
- SJT development guidelines
  - Scenarios
  - Response Options
  - Scripted SJTs
- Practice drafting SJT items, review in pairs with coaching
- Scoring SJTs

You are a flight attendant on a plane. You have just started telling the passengers the safety procedures. One of the passengers says, in a loud voice to his traveling companion, that people who listen to the safety instructions are wasting their time because plane crashes are so rare. He then continues to talk loudly to his friend and ignores you.

- A. Explain to the passenger that although plane crashes are rare, it is important to be prepared.
- B. Ask the passenger to be quiet or he/she will be removed from the plane.
- C. Talk over the passenger in a louder voice.
- D. Whistle loudly to get everyone's attention. Then tell everyone to be quiet while you are giving the safety instructions

# What is a Situational Judgment Test (SJT)?

---

- A measure of a specific construct? A test format?
  - It's a bit of both.
- Best suited for measuring constructs related to making judgments in challenging situations.
  - The more that an SJT measures knowledge, the *more* it will correlate with general intelligence; the more that it measures personality, the *less* it will correlate with intelligence.

# What is a Situational Judgment Test?

- **Scenario** or stem describes the situation
- Several possible **actions**.
  - These actions are also called the *response options* or *options*.

You are a flight attendant on a plane. You have just started telling the passengers the safety procedures. One of the passengers says, in a loud voice to his traveling companion, that people who listen to the safety instructions are wasting their time because plane crashes are so rare. He then continues to talk loudly to his friend and ignores you. What should you do?

- A. Explain to the passenger that although plane crashes are rare, it is important to be prepared.
- B. Ask the passenger to be quiet or he/she will be removed from the plane.
- C. Talk over the passenger in a louder voice.
- D. Whistle loudly to get everyone's attention. Then tell everyone to be quiet while you are giving the safety instructions

# What is a Situational Judgment Test?

---

- SJTs tend to be multidimensional. Why?
  1. Challenging situations tend to involve more than one consideration.
  2. Different response options within an item are typically related to different constructs.
  3. Even the same response can be measuring different things for different examinees.

# What is a Situational Judgment Test?

---

- This has implications for the test blueprint. A few options...
  1. Create a blueprint based on types of challenging situations rather than competencies/KSAOs
  2. Develop a standard construct-based blueprint but use it only when writing items
  3. Forego having a blueprint

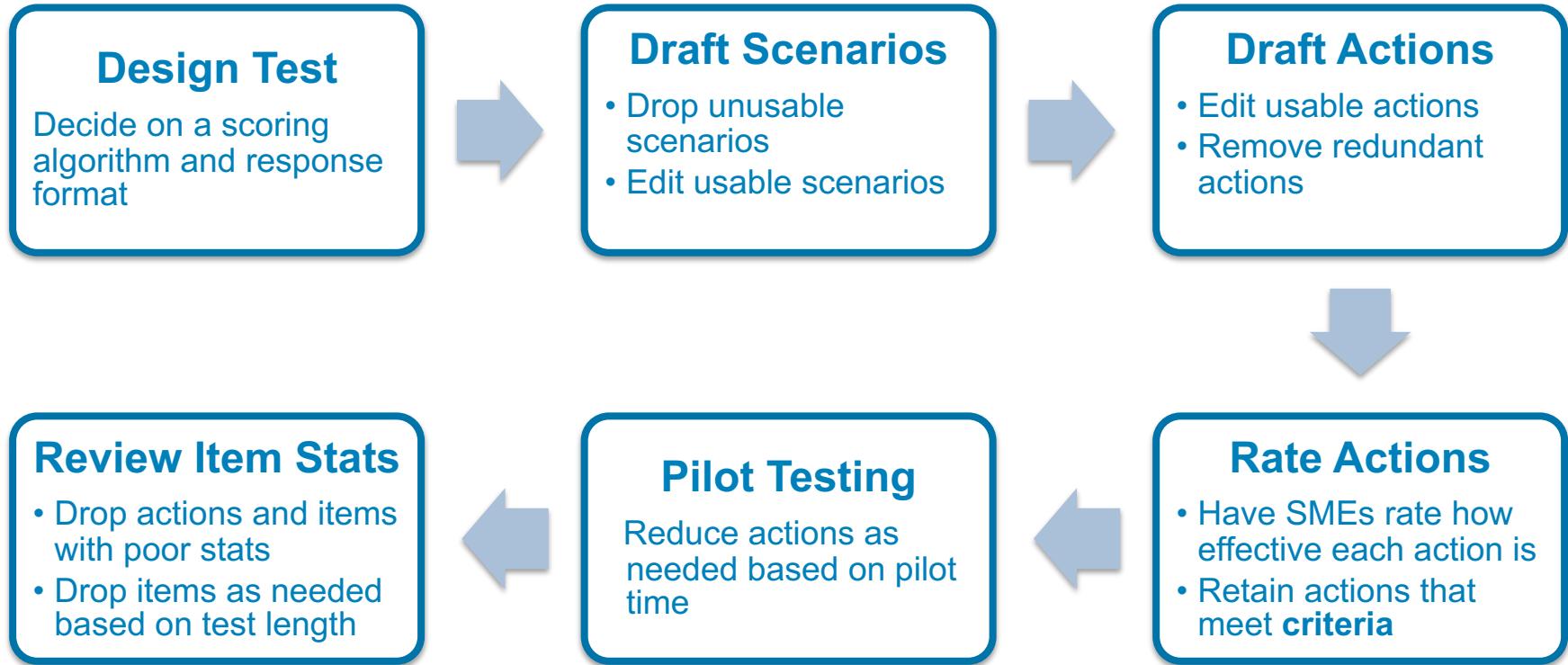
Other options possible...



# SJT Development Guidelines



# Typical SJT Development Process



# HumRRO SJT Item Guidelines – Scenarios

---

- Guideline 1. Try to focus on a specific challenging situation that requires judgment.
  - Keep scenario descriptions as concise as possible (4-6 sentences).
- Guideline 2. Ensure the scenario description is clear and realistic.
  - Situation and terminology should be understood by test-takers.
  - Test-taker does not have to make assumptions about the situation.
  - Responding to the scenario does not require knowledge of a specific department, etc.

# HumRRO SJT Item Guidelines – Scenarios

---

- Guideline 3. Scenario should lend itself to numerous possible actions or responses that vary in effectiveness.
  - Best action should not be obvious.
  - Response format has an impact here.
- Guideline 4. Avoid scenarios where “get more information” (and variations on this theme) is the best/most effective action.
  - Unless it’s a poor action to take.
- Guideline 5. Be consistent with the tense and the actor.
  - Usually in the midst of the challenging situation.
  - Usually second person, although some programs use third person.

# HumRRO SJT Item Guidelines – Scenarios

---

- Guideline 6. Use “would do” and “should do” appropriately.
  - Use *would do* when:
    - Aiming to measure personality constructs
    - The SJT is being used for developmental purposes
    - Faking is not a concern
  - Use *should do* when:
    - Aiming to measure application of knowledge in challenging situations
    - The SJT is being used for selection/credentialing purposes
    - Faking is a concern

# Practice – Draft a Scenario

---

- Start with a target competency
  - Ground yourself in the definition; keep checking to stay on track
  - A few example competencies to choose from (see handout):
    - Relating to Others
    - Customer Service
    - Problem Solving
- Think of a **work** situation that requires the competency
  - Situations you have encountered
  - Situations you have witnessed, particularly ones that involve people who demonstrated a low or a high level on the competency

# Practice – Draft a Scenario

---

- Spend 10 minutes writing a stem that includes basic facts about the scenario, in ways that clearly reflect the target competency
  - Remember:
    - Keep it short: about 4–6 lines.
    - Make sure it is clear and realistic.
    - Make sure it is complex enough to have several response actions.
- Switch with a partner, review, and discuss feedback.
- Our team is here to provide feedback.

# HumRRO SJT Item Guidelines – Options

---

- Guideline 7. Response options should be clear and concise.
  - 1 sentence is usually sufficient (2 max)
- Guideline 8. Don't list several actions in one response option.
  - State what should be done in general or what should be done first
  - Do not write double-barreled items (e.g., Do this and then do that)

# HumRRO SJT Item Guidelines – Options

- Guideline 9. Include an appropriate amount of detail.
  - Options should be similar in length, structure, and specificity

## Too General

Ask my manager.

## Too Specific

Look up my manager's phone number, dial it, and tell her about the problem.

## Just Right

Ask my supervisor to provide additional information about the assignment.

# HumRRO SJT Item Guidelines – Options

- Guideline 10. Avoid options that are clearly not the best way to react (or are clearly the worst way to react).

## Clearly Ineffective

Punch my coworker in the nose.

## Clearly Ineffective

Quit and find another job.

- Guideline 11. Avoid options that are tantamount to “get more information” (unless ineffective).
- Guideline 12. Avoid options that are tantamount to “do nothing” (unless effective).

# HumRRO SJT Item Guidelines – Options

---

- Guideline 13. Consolidate response options thoughtfully.

Some responses will be almost identical (in meaning) to or overlapping with other responses.

- At some point, the reviewing must eliminate the redundancies.
- Identify sets of completely redundant responses. Pick the best response and drop the others.
- Identify sets of partially overlapping responses. Determine the independent concepts in the set. Write one response for each concept. Drop the redundant responses.

# Practice – Draft 7 Response Options

---

- Think about possible ways to act in the situation you drafted
  - Are there multiple, clearly different ways to act?
  - Are some ways of acting clearly more effective than others?
  - Do the actions also involve the target competency?
  - Are the actions mostly independent of each other?
  - Would a person low on the competency be able to easily figure out which is the most socially desirable response?
  - Does the most effective way to behave depend on how a test-taker interprets the scenario? (In other words, is the scenario so ambiguous that different people would see entirely different things in it?)
- May need to tweak scenarios to answer yes, yes, yes, yes, no, no to these questions!

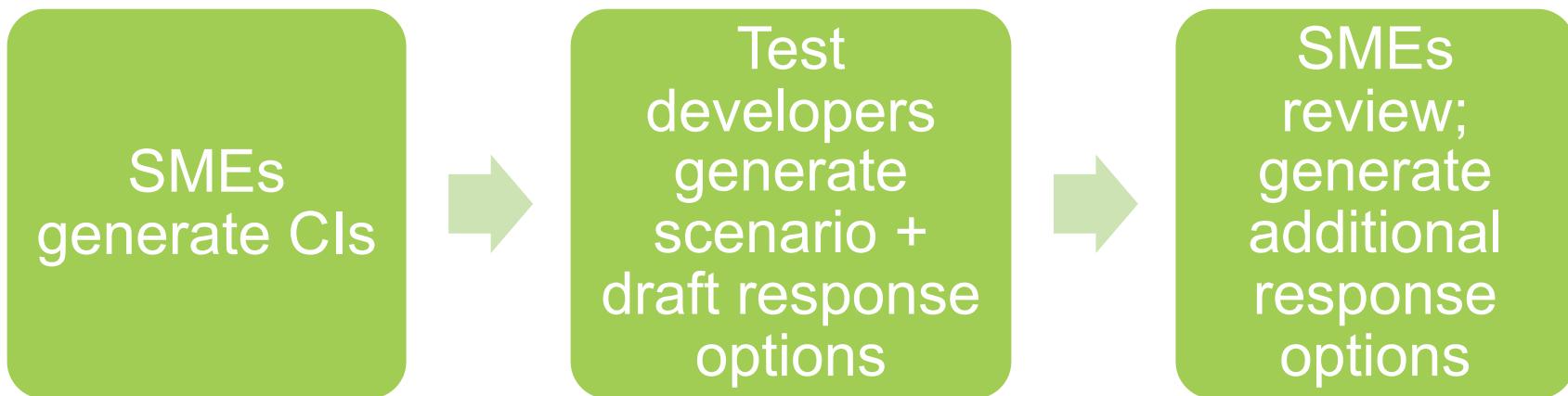
# Practice – Draft 7 Response Options

---

- At least one very effective (6 or 7 on 7-point scale) and one ineffective (1 or 2 on 7-point scale)
  - Try not to write them such that they stand out as much better or much worse than the other responses
- 5 that vary in effectiveness from low to high
  - Think “best response minus some key element” or “okay but not the best thing to do”
  - Think about how people with moderate levels of the competency would behave
- Switch with a partner, review, and discuss feedback.
- Our team is here to provide feedback.

# Who should write SJT items – Test Developers or SMEs?

- Depends on:
  - Competencies being targeting
  - Nature/complexity of the job
  - Presence/quality of critical incidents
- One common approach:





# Scoring SJTs



# SJT Scoring Overview

---

- SJTs are distinct from other assessments.
  - May not have an unambiguously “correct” answer
  - Instead, there are “more correct” and “less correct” answers
- How do you come up with the “right” answer?
  - It depends.
  - We learn the answers by asking people and getting a consensus about what the “right” answers are.
  - There won’t be perfect agreement in what people think are the “best” and worst answers, so we need our scoring system to account for those complexities

# Scoring Key Development

---

- **Empirical** – Key is based on the relationship between the incumbents' responses and a criterion such as their job performance ratings.
- **Theoretical** – Key is based on what a theory would say is the “best” answer.
- **Rational** – Key is defined by SMEs’ judgments.
  - We are going to focus on this method.

# Developing a Rational Scoring Key

---

- SMEs rate the effectiveness of each response option
  - Not much research on who the SMEs should be
  - Some of our personal rules of thumb:
    - Should have at 10-15 SMEs, but more is better
    - Should be individuals with operational experience in the SJT content and familiar with the target population
    - Diversity in perspectives is good
  - The effectiveness ratings will be used to determine the “rightness” and “wrongness” of the ratings
- Can sometimes use incumbents/novices (consensus-based)
  - In our experience, keys developed with incumbents are very similar to those developed using SMEs
  - Need larger numbers if using novices rather than experts

# Common Response Formats

---

- Pick-best or pick-best-and-worst
  - **Advantages:** Easier for examinees, scoring is simple.
  - **Disadvantages:** Lower reliability because there are only 1 or 2 responses per item. It's also harder to write options because they must vary considerably in effectiveness within each item.
- Rate each option's effectiveness
  - **Advantages:** Higher reliability because there is one response for each option. Flexibility in option effectiveness: They can be similar or different. Flexibility in scoring.
  - **Disadvantages:** Scoring is more complex. Scores unintuitive unless they are rescaled.

# Reviewing Item Statistics

---

- We often write more items and options than we need. So, we must decide which options to keep and which items to drop.
- Select the best set of response options for each item; then decide which items to keep.
  - Administer draft SJT to incumbents and/or SMEs
  - Compute diagnostic statistics (Mean, SD, ITCs, etc.)
  - Use statistics to make decisions about response options and items
  - Compare test-taker scores with the keyed score (e.g., distance score)
  - Use scores to compute diagnostic stats to evaluate options
  - Compute stats at item level to evaluate items



# Questions?



# SJT Bibliography

---

- Arthur Jr., W., Glaze, R.M., Jarrett, S.M., White, C.D., Schurig, I., & Taylor, J.E. (2014). Comparative evaluation of three situational judgment test response formats in terms of construct-related validity, subgroup differences, and susceptibility to response distortion. *Journal of Applied Psychology*, 99, 535-545.
- Arthur, W. Jr., & Villado, A.J. (2008). The importance of distinguishing between constructs and methods when comparing predictors in personnel selection research and practice. *Journal of Applied Psychology*, 93, 435-442.
- Bergman, M.E., Drasgow, F., Donovan, M.A., Henning, J.B., & Juraska, S.E. (2006). Scoring situational judgment tests: Once you get the data, your troubles begin. *International Journal of Selection and Assessment*, 14, 223-235.
- Chan, D., & Schmitt, N. (2002). Situational judgment and job performance. *Human Performance*, 15, 233-254.
- Christian, M.S., Edwards, B.D., & Bradley, J.C. (2010). Situational judgment tests: Constructs assessed and a meta-analysis of their criterion-related validities. *Personnel Psychology*, 63, 83-117.

# SJT Bibliography (cont.)

---

- Clevenger, J., Pereira, G.M., Weichmann, D., Schmitt, N., & Harvey, V.S. (2001). Incremental validity of situational judgment tests. *Journal of Applied Psychology*, 86, 410-417.
- Cullen, M. J., Sackett, P. R., & Lievens, F. (2006). Threats to the operational use of situational judgment tests in the college admission process. *International Journal of Selection and Assessment*, 14, 142-155..
- Lievens, F. & Motowidlo, S.J. (2016). Situational judgment tests: From measures of situational judgment to measures of general domain knowledge. *Industrial and Organizational Psychology: Perspectives on Science and Practice*, 9, 3-22.
- McDaniel, M.A., Hartman, N.S., Whetzel, D.L., Grubb, W.L. (2007). Situational judgment tests, response instructions, and validity: A meta-analysis. *Personnel Psychology*, 60, 63-91.
- McDaniel, M.A., List, S.K., & Kepes, S. The “hot mess” of situational judgment test construct validity and other issues. *Industrial and Organizational Psychology: Perspectives on Science and Practice*, 9, 47-51.

# SJT Bibliography (cont.)

---

- McDaniel, M.A., Morgeson, F.P., Finnegan, E.B., Campion, M.A., & Braverman, E.P. (2001). Use of situational judgment tests to predict job performance: A clarification of the literature. *Journal of Applied Psychology*, 86, 730-740.
- McDaniel, M.A., Psotka, J., & Legree, P.J. (2009, April). *Toward an understanding of situational judgment item validity*. Paper presented at the annual meeting of the Society for Industrial and Organizational Psychology, New Orleans, LA.
- 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.
- McDaniel, M.A., Whetzel, D.L., Hartman, N.S., Nguyen, N.T. & Grubb III, W. L. (2006). Situational judgment tests: Validity and an integrative model. In J.A. Weekley & R.E. Ployhart, (Eds). *Situational judgment tests: Theory, measurement, and application.*, (pp. 183-203). Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.
- 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.

# SJT Bibliography (cont.)

---

- Ployhart, R.E. (2006). The predictor response process model. . In J.A. Weekley & R.E. Ployhart, (Eds). *Situational judgment tests: Theory, measurement, and application.* (pp. 83-106). Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.
- Ployhart, R.E. & Ehrhart, M.G. (2002). Modeling the practical effects of applicant reactions: Subgroup differences in test-taking motivation, test performance, and selection rates. *International Journal of Selection and Assessment, 10,* 258-270.
- Ployhart, R.E., Ehrhart, M.G., (2003). Be careful what you ask for: Effects of response instructions on the construct validity and reliability of situational judgment tests. *International Journal of Selection and Assessment, 11,* 1-16.
- Ployhart, R.E. & MacKenzie Jr., W.I. (2011). Situational judgment tests: A critical review and agenda for the future. In S. Zedeck (Ed.). *APA handbook of industrial and organizational psychology, Vol 2: Selecting and developing members for the organization.* (pp. 237-252). Washington, DC, US: American Psychological Association.

# SJT Bibliography (cont.)

---

- Putka, D.J., & Waugh, G.W. (2007, April). *Gaining insight into situational judgment test functioning using spline regression*. Paper presented at the annual meeting of the Society for Industrial and Organizational Psychology, New York, NY.
- Waugh, G.W., Russell, T.L. (2006, April). *The effects of content and empirical parameters on the predictive validity of a situational judgment test*. Paper presented at the annual meeting of the Society for Industrial and Organizational Psychology, Dallas, TX.
- Waugh, G.W., Russell, T.L. (2005, April). *Comparison of situational judgment test formats, scoring key developers, and scoring algorithms*, . Paper presented at the annual meeting of the Society for Industrial and Organizational Psychology, Los Angeles, CA.
- Weekley, J.A., Ployhart, R.E., & Holtz, B.C. (2006). *Situational Judgment Tests: Theory Measurement and Application*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Whetzel, D.L., McDaniel, M.A., & Nguyen, N.T. (2008). Subgroup differences in situational judgment test performance: A meta-analysis. *Human Performance*, 21, 291-309.