How to Develop and Score a Situational Judgment Test (SJT)

Gordon W. Waugh & Matthew Allen
Human Resources Research Organization (HumRO)

Annual IPAC Conference
Washington, DC
July 20, 2011

Sample SJT Item

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

Note. Some typical response instructions are:

1. *What is the best action*,

2. *Indicate the best action and the worst action*, or

3. *Rate the effectiveness of each action*, or

4. Which action would you take? or

5. Which action would you most likely take, and which action would you least likely take?

What is a Situational Judgment Test?

Is a situational judgment test a measure of a specific construct or is it a test format? It's a bit of both. It is a test format that is 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. An SJT item consists of two elements: a *scenario* which describes the situation and several possible *actions*. These actions are also called the *response options*—or just *options* for short.

Would-Do vs. Should-Do

An SJT item can ask either what you *should* do or what you *would* do in the situation. Should-do items tend to measure personality constructs better. However, would-do SJTs can be faked just

like other tests that measure personality constructs. Therefore, many organizations do not use would-do SJTs for selection or in other situations where there is motivation to fake. But they still use would-do SJTs for developmental purposes. On the other hand, faking is not an issue with should-do SJTs because examinees' responses measure whether they know what to do in a situation rather than what they would actually do.

What does a should-do SJT measure? It typically measures the ability to apply knowledge to challenging situations—whether the knowledge has been obtained through training, education, or experience. However, if an examinee has no knowledge related to the situation, then he might respond based on his personality. For example, consider an examinee who is aggressive. How would he answer the flight-attendant item we just looked at? If he remembers his training, then he will show respect for the passenger and he will judge the first action to be the best one: *Explain to the passenger that although plane crashes are rare, it is important to be prepared*. On the other hand, if an aggressive examinee has no knowledge relevant to this situation, he will tend to judge the *second* action to be best: *Ask the passenger to be quiet or he/she will be removed from the plane.*

Multidimensionality

SJTs tend to be multidimensional. Not only is it difficult to develop a set of SJT items that measures a single construct, it's difficult to write even a single item that is unidimensional. There are at least three reasons for this. First, challenging situations tend to involve more than one thing to consider. Second, the different response options within an item are typically related to different constructs. In the item we looked at, for example, the first option seems to be measuring respect for the customer whereas the second option seems to be measuring aggression. Third, even the same response can be measuring different things for different examinees. One examinee might pick an option as the best one based on knowledge that she gained in training whereas another examinee might pick it based on her personality.

Therefore, I rarely compute dimension scores for an SJT. The items in a dimension are probably not measuring that dimension very well. Although they might be tapping the intended dimension, they are probably tapping several other dimensions as well.

Summary of Steps for Developing an SJT

There are various ways to develop a situational judgment test. I'll be showing you the way that I typically do it. Here are the steps:

1. Decide which scoring algorithm and response format to use.
2. Write draft scenarios.
3. Drop the unusable scenarios. Edit the usable ones.
4. Write several draft actions for each retained scenario.
5. Edit the draft actions. Remove redundant ones.
6. Ask subject matter experts (SMEs) to rate the effectiveness of each action.
7. Drop actions where SMEs' ratings vary considerably.

8. In preparation for pilot testing, reduce the number of actions if administration time is limited (e.g., 7 actions per item).
9. Pilot/field test the items.
10. Review option and item statistics. Drop options and items with poor statistics.
11. Decide which options to retain for each item. For example, you might want each item to have only four options.
12. Decide which items to keep.

During the remainder of the workshop, we will work through the additional handouts.