Situational Judgment Tests
as Measures of Knowledge/Expertise

Mary Ann Hanson
Personnel Decisions Research Institutes

Kristen E. Horgen and Walter C. Borman
University of South Florida and Personnel Decisions Research Institutes

Overview

This paper discusses the Situational Judgment Test (SJT) methodology. The SJT approach to test development has become increasingly popular in recent years, and some likely reasons for this popularity are discussed. Evidence to date indicates that this approach can be used to develop valid predictors of performance, especially for managerial positions and other positions in which interpersonal interactions are important. Less is known about the nature of the construct(s) measured by these tests, why they are valid, when they are valid, and why they are sometimes not valid. This paper proposes that the situational judgment test methodology is best suited for measuring knowledge or expertise, and discusses available construct validity evidence consistent with this perspective. This perspective generates several testable hypotheses, and additional research is proposed. Finally, the implications of this perspective for the development of valid and useful SJTs are discussed.

Background

Situational Judgment as a Methodology

Situational judgment tests have been used as employee selection tools for several decades, but in recent years the situational judgment approach has become increasingly popular. These tests present realistic, job-related situations, usually described in writing. Examinees are asked to indicate, in a multiple choice format, what should be done to handle each situation effectively. These responses are often scored according to relative level of effectiveness, rather than simply right or wrong.

The most common use of situational judgment tests is for selecting managers and supervisors (e.g., Motowidlo, Dunnette, & Carter, 1990). However, situational judgment tests have also been developed to predict success in other types of jobs, including insurance agent, police, and sales positions. This sort of test has become increasingly popular for selecting employees for work in customer service positions as well (e.g., Motowidlo & Tippins, 1993). Although the actual content of situational judgment tests differs across the various applications, most of them

Paper presented in David A. Dubois (chair) Job Knowledge: Explorations of its Conception, Description, and Measurement. Symposium conducted at the Thirteenth Annual Conference of the Society for Industrial and Organizational Psychology, April, 1998, Dallas, Texas.
appear to focus on interpersonal and problem solving situations. Thus, a reasonable hypothesis is that these tests measure knowledge concerning how to effectively handle difficult interpersonal situations, especially situations in which there are no right or wrong answers, but rather more and less effective courses of action.

Based on the present definition (i.e., realistic situations with multiple choice responses that vary in terms of effectiveness), situational judgment is arguably more appropriately viewed as a measurement method than as targeting any particular individual differences constructs. The nature of the underlying constructs measured will differ according to the nature of the situations presented. Thus, SJTs could conceivably be developed to measure a wide variety of different individual differences traits, although they are probably best suited for developing measures involving judgment, decision-making and interpersonal skill. This also means that careful attention to the nature of the situations included is important both when developing these tests and in interpreting their construct validity. If two different situational judgment tests are not highly correlated with each other, this does not necessarily mean that one or the other is not a valid test; it might mean only that each one is measuring a different construct (Motowidlo, Hanson, & Crafts, 1997).

**Reasons for SJT Popularity**

In spite of our lack of understanding concerning what SJTs measure, their popularity continues to increase. The following are a few likely reasons why this is the case. One particularly compelling advantage of SJTs is the high face validity these tests typically possess. Presenting applicants with actual job situations and scoring their responses according to their effectiveness in handling these situations has a great deal more face validity than traditional cognitive ability measures. In fact, in 1961 Rosen argued that even if a situational judgment test did not add anything to the prediction of success beyond that obtained with intelligence tests and biodata “it can be argued that . . . the instrument’s high face validity makes it more desirable to use than some others” (p. 97).

In general, these tests also have good criterion-related validity (e.g., Forehand & Guetzkow, 1961; Tenopyr, 1969; Motowidlo et al., 1990). Situational judgment test scores have been shown to be related to measures of job performance, such as job performance ratings. Scores on situational judgment tests have also been shown to be related to other organizationally important criteria, such as salary, promotion rate and tenure.

Perhaps even more important is the consistent finding that these tests have less adverse impact than traditional ability tests. Several researchers have found that SJTs have about half the adverse impact against blacks as traditional cognitive ability tests (e.g., Motowidlo et al., 1990; Hanson & Borman, 1995). Valid predictors with relatively low adverse impact are difficult to find, and the search for such alternative predictors is increasingly important in most applied settings.

**Importance of Construct Validity**

One might argue that as long as SJTs work, it’s not particularly important what they measure. There are several reasons, however, that an understanding of the construct(s) measured is
important. While these tests are generally valid, they are not always valid. That is, SJTs sometimes fail to correlate with the intended aspects of job performance. An understanding of the nature of the construct(s) measured by these tests is likely to help in determining why SJTs sometimes do not work.

Some research has also demonstrated the usefulness of this technique for developing criterion measures of job performance (e.g., Hanson & Borman, 1995). It is somewhat counterintuitive that a technique useful as a predictor measure can also be useful for developing criterion measures. In addition, with only a few exceptions (e.g., Dalessio, 1992; Jones & Youngblood, 1993), all of the available research on SJT validity has involved concurrent validation designs. Some researchers have argued that, in general, results obtained using concurrent and predictive validation studies do not differ systematically in the overall level of validity obtained (e.g., Barrett, Phillips, & Alexander, 1981), but there is reason to expect that this may not hold true for SJTs. These tests generally present job-related situations for the target job, and it seems likely that incumbents will have encountered similar experiences on the job. Applicants may or may not have had experience in similar situations. Thus, it seems possible that predictive and concurrent validities for these tests might differ systematically. A better understanding of what these tests are measuring may clarify the extent to which concurrent validities can be expected to approximate longitudinal validities for a given situational judgment test.

Finally, a better understanding of what is measured by SJTs and the reasons for their correlations with other measures would likely be useful for improving tests developed using this approach. Understanding why these measures often correlate with performance could be useful for improving the intended correlations. In addition, this understanding might suggest for which situations an SJT will be a valid predictor and for which purposes such an approach is unlikely to work. More information concerning the construct or constructs measured by this type of test would be extremely useful.

Knowledge as the Construct Measured

This paper proposes that all of the information available to date is consistent with the interpretation of situational judgment tests as measures of job-relevant knowledge or expertise. This has been suggested previously (e.g., Schmidt & Hunter, 1993). The purpose of this paper is to discuss available construct validity evidence in light of this perspective and the implications for further research and future SJT development efforts.

Available data and theory concerning the general construct of job knowledge provides information concerning the expected correlates of a knowledge measure. Hunter (1983) conducted a meta-analysis of the relationships between ability, knowledge, work sample and job performance. He concluded that ability only affects performance through it's effect on knowledge and skill. Similarly, Campbell, Gasser, and Oswald (1996) propose that the three direct determinants of performance are declarative knowledge, procedural knowledge and skill, and motivation. Further, individual differences only affect performance through their effect on these variables. Experience has been shown to lead to higher levels of performance through the acquisition of job-relevant knowledge (Schmidt, Hunter, Outerbridge, 1986). If SJTs are generally measures of job-related knowledge, we would expect this same pattern of relationships to hold for SJT scores as well.
The following sections first review available research in the context of the construct of job knowledge and propose additional research based on, as yet, untested expectations that proceed from these models. We propose that individual differences in job-relevant knowledge, and thus individual differences in SJT scores, will be directly affected by two antecedent variables: (1) relevant experience, and (2) ability to learn from this experience.

**Evidence for Construct Validity**

Available literature concerning situational judgment tests provides a fair amount of information concerning the correlates of these tests, which can provide the basis for beginning to form a nomological net (Cronbach & Meehl, 1955) to better understand the construct(s) measured. The focus of this section is on research assessing relationships between SJT scores and other important variables, such as personality, cognitive ability, and amount of training and experience. Research results are discussed in terms of their implications for understanding the nature of the construct measured, especially as they are consistent with the interpretation of situational judgment tests as measures of knowledge or expertise.

**SJT Reliability/Dimensionality**

Most researchers report that situational judgment tests have at least moderate reliability. Internal consistency reliabilities are generally moderate, with most in the 60s and 70s (Mowry, 1957; Mandell, 1950; Motowidlo et al., 1990; Bruce & Learner, 1958; Richardson et al., 1981). It is important to note that some of these tests have been designed to measure multiple constructs, so high internal consistency reliability is not always expected. Test-retest reliabilities are not as often reported. The few that have been reported are somewhat higher than the internal consistencies, generally in the 80s (Bruce & Learner, 1958; Richardson et al., 1981). Attempts to identify several underlying factors in SJTs have generally not been successful (e.g., Houston & Schneider, in press, Hanson & Borman, 1995; Motowidlo et al., 1990), even when these tests attempt to measure more than one underlying construct.

**SJT Criterion-Related Validity Evidence**

Researchers have investigated relationships between situational judgment test scores and important organizational criteria, including performance ratings by supervisors or peers, salary, promotion rate, and tenure. The vast majority of this research has used job performance ratings as the criterion. A few studies have failed to obtain significant correlations (e.g., Hedge, Hanson, Bruskiewicz, Logan & Borman, 1996; Smiderle et al., 1994) but in general the results have been positive. McDaniel, Finnegan, Morgeson, and Campion (1997) conducted a meta-analysis of this research, and concluded that the average correlation between SJTs and performance ratings, based on 95 correlations across a total sample of 15,234, is .27 with a standard deviation of .12 across studies. Since the time of this review, other studies have also reported validities in this same range. For example, Phillips (1993) obtained correlations ranging from .41 to .45 (p < .001) between scores on her Negotiation Skills Inventory and supervisory performance ratings. Pulakos and Schmitt (1996) obtained correlations of .20 and .14 between their situational judgment test and two supervisory rating composites (p < .05).

Scores on situational judgment tests have also been shown to be related to other important organizational criteria such as salary, promotion rate, and tenure. Tenopyr (1969) found that
scores on a managerial SJT correlated .36 (p < .01) with salary corrected for age and length of service, but scores on this test were not significantly correlated with seniority. Mandell (1950) reported that correlations between Administrative Judgment Test scores and pay grade ranged from .28 (ns) to .56 (p < .01) across the three samples studied, with a median of .52 (p < .01). For two other samples, Mandell (1956) reported that correlations between Administrative Judgment Test scores and salary level were .48 (p < .01) and .54 (p < .01) respectively, and in yet another sample, Administrative Judgment Test scores correlated .32 (p < .01) with the number of promotions received in a four-year period. Dalessio (1992) found that the three empirical keys developed to predict attrition based on a video situational judgment test had low but significant correlations with attrition in cross-validation samples (average r = .12; p < .05).

In general, results are similar if the SJT situations are presented in video rather than paper-and-pencil format, although there is not yet enough research available to make systematic comparisons. Weekly and Jones (1997) developed a video SJT to predict performance in nursing home jobs and obtained correlations ranging from .14 to .33 with performance ratings. Jones and DeCotiis (1986) conducted a concurrent validity study in developing a selection system for hotel employees (n = 362) and reported a correlation of .38 between a video SJT and supervisory ratings. In a predictive study of bank tellers and customer service representatives, Jones & Youngblood (1993) found that scores on a video-based test of financial services skills correlated significantly (r = .20) with supervisor ratings. Conflict resolution skills scores were significantly related to job-related conflict management performance (cross-validated r = .26, p <.01; Olson-Buchanan et al., 1994). Scores on a video-based situational judgment test for corrections officers were significantly correlated with both performance and attitudinal criteria (Pine, 1994). However, Smiderle et al. (1994) reported that scores on their video-based transit operator test were significantly related to only one of three criterion measures, and that correlation was of limited practical significance.

Correlations with Mental Ability

Situational judgment tests generally have substantial correlations with measures of general cognitive ability, although a few researchers have not obtained significant correlations (e.g., Motowidlo et al., 1990). McDaniel et al.'s (1997) meta-analysis examined correlations between SJTs and general cognitive ability. They conclude that the average across 54 correlations with a total sample of 6580 was .41 with a standard deviation of .24.

In research that has included both types of predictors, situational judgment tests generally predict job performance better than does cognitive ability (e.g., Tenopyr, 1969; Hanson & Borman, 1995; Mandell, 1950; Forehand & Guetzkow, 1961). Although more research is needed before any strong conclusions can be drawn regarding the relative validity of general cognitive ability and situational judgment measures, these results are consistent with previous research on job knowledge tests. Job knowledge tests, overall, tend to be slightly more valid than ability tests (Hunter & Hunter, 1984), and research has demonstrated that job knowledge appears to mediate the relationship between ability and job performance (e.g., Schmidt et al., 1986).

There are at least three potential reasons for the correlations between SJTs and measures of cognitive ability. First, people who are higher in cognitive ability may have had more opportunities to obtain relevant experience. For example, higher ability individuals are probably more likely to be placed in supervisory or other challenging situations. In this case,
ability would be seen as having an indirect affect on SJT scores, through experience. Second, and probably more importantly, higher ability people can be expected to learn more from relevant experiences. This is especially true if the knowledge to be obtained is difficult or complicated. Finally, higher ability people may simply be better able to figure out the answers to SJTs. This third potential reason is not necessarily consistent with viewing SJTs as job knowledge measure, but might rather be viewed as a source of unwanted variance. Regardless of the reasons for ability-SJT correlations, research to date is consistent with the notion that ability to learn from relevant experience is one of the direct antecedents of SJT performance.

**Correlations with Experience/Training**

Experience is hypothesized to be one of the direct antecedents of the knowledge measured by situational judgment tests, and training can be viewed as a special case of experience that more directly targets the relevant aspects of job knowledge. Thus, one especially important type of construct validation evidence concerns the relationship between situational judgment tests and amounts of relevant training and experience. Not a great deal of information is available on this topic, but some researchers have found significant relationships between measures of relevant training or experience and situational judgment test scores.

Bosshardt and Cochran (1996) obtained a small but significant correlation between scores on their SJT and tenure in the financial planner job. Hanson and Borman (1995) report significant correlations between time in a supervisor position, frequency of supervisory responsibility, number of supervisory training courses attended, and scores on a supervisory SJT. Weekly and Jones (1997) obtained small but significant correlations between their video situational judgment test and a 5-item measure of experience in several different samples.

It is worth noting that conceptualizing and assessing the experience relevant to a situational judgment test is not straightforward. The use of SJTs as predictors is arguably based on the assumption that people can pick up knowledge concerning how to handle difficult, job-related situations before even beginning a job. This is a reasonable assumption for the type of situations included in many SJTs. However, this sort of informal approach to obtaining relevant experience is difficult to assess. Time on a job or in a career field might be only weakly related to SJT scores, because much of the relevant knowledge could be picked up informally. In addition, two people with the same job title might encounter different interpersonal situations (such as those included on many SJTs) with different frequencies. Thus, most readily available experience measures could be viewed as incomplete for assessing experience relevant to SJT performance. The fact that significant correlations have been obtained at all could be viewed as fairly good evidence that experience is an important antecedent of the type of job knowledge assessed by SJTs.

**Correlations with Personality Variables**

The hypothesis that SJTs generally measure job-relevant knowledge does not preclude relationships with non-cognitive variables. In fact, one of the two proposed direct antecedents of knowledge, experience, is likely to have important relationships with non-cognitive variables, especially for interpersonal situations. In generating hypotheses concerning personality correlates of SJTs, it again becomes apparent that the nature of the situations included in each test is likely to be related to the construct(s) measured. If experience in
relevant situations is related to personality predictors, the nature of the personality predictors of SJT performance would be expected to vary systematically according to the situations included. For example, we could hypothesize that sociability would be related to obtaining experience in interpersonal situations, and that dominance would be related to obtaining experience in leadership situations. Thus, while SJT scores could be expected to correlate with personality, we would not necessarily expect every SJT to correlate with the same personality constructs. Personality is probably the most relevant for SJTs targeting interpersonal or leadership constructs, and since the majority of SJT research to date has focused on these general domains, significant correlations with at least some personality dimensions are expected.

Some research has, in fact, obtained significant correlations between personality measures and SJT scores. Hanson and Borman (1995) found that scores on an SJT targeting supervisory knowledge/skill correlated significantly with dominance, dependability, and work orientation. Bosshardt and Cochran (1996) developed an SJT to predict performance in financial planner jobs. They obtained significant correlations between their SJT and several personality scales, also developed to predict performance in this job, including communication/persuasiveness and service orientation. Houston and Schneider (in press) obtained significant correlations between an SJT designed to predict insurance agent performance and several personality scales, including people/service orientation, drive to achieve, flexibility and leadership. Interestingly, the largest correlation was with integrity \( r = .39; p < .01 \).

Although limited information is available concerning the personality correlates of SJTs, available data suggest that SJTs can correlate significantly with certain personality dimensions, especially the more interpersonal aspects of personality (e.g., dominance). Some would argue that these correlations demonstrate that situational judgment tests measure more than just knowledge. However, these data are also consistent with the present hypothesis that personality traits are related to the likelihood of obtaining experience in relevant interpersonal situations (e.g., leadership experiences), and that it is this experience that actually leads higher SJT scores.

**Correlations with Other Measures**

Several researchers have found that SJT scores correlate significantly with role play or simulation exercises targeting similar constructs (e.g., Hanson & Borman, 1995; Bosshardt & Cochran, 1996; Pulakos & Schmitt, 1996). This is consistent with Hunter’s (1983) meta-analysis, in which he found a direct path from knowledge to work sample performance.

Bosshardt & Cochran (1996) also found that SJT scores correlated significantly with social interests. Interests probably affect SJT scores via the same mechanism as personality measures, through obtaining and learning from experience in relevant situations.

**Proposed Prerequisites for SJT Validity**

The research reviewed to this point is consistent with the hypothesis that most situational judgment tests are measuring knowledge or expertise. Although alternative explanations of the available data can not be ruled out, the view that these tests measure knowledge has important implications for test development.
The view that SJTs predict job performance because they assess job-related knowledge suggests two prerequisites for SJTs functioning as valid predictors of performance. First, if SJTs are to successfully measure job-related knowledge, the situations included must be similar to those encountered in the target job. This is supported by McDaniel et al.'s (1997) meta-analysis results. They demonstrated that SJTs developed based on careful job analysis were systematically more valid than those that were not (average correlation of .29 versus .21). This has important implications for the transportability of situational judgment tests. A test developed and validated in one setting (e.g., one organization or job) may not be a valid predictor in another setting (e.g., another organization or job). Careful attention to the similarity of the two settings is important.

The second prerequisite for SJT validity from this perspective is that the examinees must have experience in the target situations, or very similar situations, in order to have had an opportunity to pick up the relevant knowledge. As mentioned previously, relevant experience is difficult to assess. For many SJTs, it is likely that the relevant experience could be obtained informally. Interestingly, this latter prerequisite for SJT validity provides a possible explanation for the one unexpected finding in McDaniel et al.'s meta-analysis of SJT validities. They found that less detailed situations are actually more valid (although the number of studies with more detailed questions was relatively small). It seems reasonable to expect that these less detailed questions are worded more broadly. Perhaps this leads to a broader array of experiences being relevant to the situation presented. While this is highly speculative at this point, it is very consistent with the interpretation of SJTs as measures of the knowledges important for job success.

This also has implications for using a concurrent validation research design to assess the validity of situational judgment tests. Job incumbents have, by definition, had opportunities to obtain job-relevant experience. The same is not necessarily true of job applicants. If a test is validated based on job incumbents, and if applicants differ systematically from job incumbents in terms of relevant experience, it is not clear that the same level of SJT validity would be expected. One way to avoid this potential problem is to develop situations that are sufficiently general such that most applicants have a reasonable amount of relevant experience. If the present hypothesis that SJTs measure job-relevant knowledge is born out in future research, it may not always be appropriate to assume that concurrent validities are a good approximation of predictive validities for this type of test.

**Directions for Future Research**

One important direction for future research is longitudinal validation of situational judgment tests as predictors of job performance and other important organizational criteria.

The hypothesis that SJTs measure job knowledge also generates several additional testable hypotheses. For example, scores on an SJT would be expected to change as a result of experience or training, and research concerning the extent to which these scores change as a result of relevant experience/training would be very informative. It would also be useful to develop more targeted measures of amount of experience in relevant situations and assess correlations of these measures with SJT scores.
Some research is available suggesting that SJT scores can account for the relationship between ability and performance (e.g., Borman, Hanson, Oppler, Pulakos, & White, 1993), but additional research concerning the extent to which scores on these tests can account for relationships between individual differences and performance would be highly informative.

Conclusions

This paper proposes that situational judgment tests are best viewed as a measurement method, rather than a distinct individual differences construct. However, it is a measurement method well suited for measuring job-relevant knowledge, especially knowledge related to interpersonal situations. It is important to emphasize that this is only a hypothesis, and further research is needed to confirm or refute this perspective. Still, it is consistent with available data and might be our best guess at this point as to what these tests measure.

It is also important to note that viewing SJTs as measuring job-related knowledge does not necessarily make these measures any less interesting. In Campbell et al.‘s (1996) model of performance, two of the three direct determinants of performance involve knowledge/skill. If SJTs do, in fact, assess direct determinants of performance, have relationships with important personality and experience variables, and show less adverse impact than more traditional cognitive ability measures, one would be hard pressed to conceive of a more interesting and useful measure. SJTs have been used as predictors and as criterion measures, and their interpretation as knowledge measures is consistent with both uses. If SJTs measure job knowledge, they could also be very useful for training needs assessment and training evaluation.

The significant relationships often obtained between personality measures and SJT scores suggest that this methodology may be useful for assessing personality-related constructs. If the construct(s) measured by some SJTs (e.g., job related knowledge and skill) does, in fact, mediate the relationship between personality variables and job performance and there are theoretical reasons to suggest this is the case (Motiwidlo et al., 1997), this would make them particularly appropriate as personality-related performance predictors. Even if SJT scores do not account for the validity of personality measures, capitalizing on their correlations with personality constructs could be useful. A better understanding of the construct measured by these tests may be useful in developing approaches for increasing the personality-SJT correlations. For example, if the effect of personality on SJT scores is mediated by relevant experience, developing a theory concerning the types situations for which experience is most likely to be affected by personality could aid in these efforts.

Finally, if the construct validity of SJTs as knowledge measures holds up in future research, the proposed prerequisites for SJT validity may prove extremely useful to test developers.
References


