

19990423 035

AB-34D - Symposium

Situational Judgment: An Alternative Approach to Selection Test Development

Mary Ann Hanson
Personnel Decisions Research Institutes

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

This paper discusses the Situational Judgment Test (SJT) methodology for developing selection measures, and provides a brief review of some key research on this type of test. SJTs 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 typically 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 SJTs is for selecting managers and supervisors (e.g., Motowidlo, Dunnette, & Carter, 1990). However, SJTs 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). The military has used SJTs for years (e.g., Helme, 1968), and recently there seems to be an increase in military interest in this type of measure (e.g., Arad & Borman, in preparation; Hanson & Borman, 1995; Hedge, Hanson, Borman, Bruskiewicz, & Logan, 1997; Legree, 1995).

Reasons for SJT Popularity

There are several likely reasons for the popularity of this approach. 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 an SJT 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).

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 (e.g., Tenopir, 1969; Motowidlo et al., 1990). 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 African Americans 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. Finally, these measures often have significant correlations with personality variables and other non-cognitive measures (e.g., Bosshardt & Cochran, 1996). Some have even argued that this methodology can be used to develop "maximal" measures of personality, that avoid problems typically associated with traditional personality inventories (e.g., susceptibility to response distortion).

SJTs as a Measurement Method

Situational judgment is arguably most appropriately viewed as a measurement method, rather 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 SJTs 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 is measuring a different construct (Motowidlo, Hanson, & Crafts, 1997).

SJT Relationships with Other Measures

Available literature concerning SJTs 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.

Most researchers report that SJT scores have at least moderate reliability. Internal consistency reliabilities are generally moderate, with most in the 60s and 70s (Mowry, 1957; Motowidlo et al., 1990; Bruce & Learner, 1958). 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 (e.g., Bruce & Learner, 1958). 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.

Researchers have investigated relationships between SJT scores and important organizational criteria. 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., 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. Scores on SJTs have also been shown to be related to other important organizational criteria such as salary, promotion rate, tenure, and attrition (e.g., Tenopyr, 1969; Dalessio, 1992). In general, results are similar if the SJT situations are presented in video rather than paper-and-pencil format (e.g., Weekly & Jones, 1997; Olson-Buchanan et al., 1994), although there is not yet enough research available to make systematic comparisons of the different formats.

It is worth noting that, with only a few exceptions (e.g., Dalessio, 1992), all of the available research on SJT validity has involved concurrent validation designs. While 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), 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 SJT.

SJT scores generally have substantial correlations with measures of 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 concluded 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, SJTs generally predict job performance better than does cognitive ability (e.g., Tenopyr, 1969; Hanson & Borman, 1995; Forehand & Guetzkow, 1961). 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 more likely to be placed in supervisory or other challenging situations. In this case, ability would be seen as having an indirect effect 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 situations are difficult or complicated. Finally, higher ability people may simply be better able to figure out the answers to SJT questions.

Not a great deal of information is available on relationships between relevant training and experience and SJT scores, but some researchers have obtained significant relationships. 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 supervisory 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 SJT and a 5-item measure of experience in several different samples. It is worth noting that conceptualizing and assessing the experience relevant to an SJT is not straightforward. The use of SJTs as predictors is arguably based on the assumption that people learn 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 difficult 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.

Some research has also 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 in this latter study 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 SJTs measure more than just knowledge. However, these data are also consistent with the 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. Bosshardt and Cochran (1996) also found that SJT scores correlated significantly with social interests. Interests could be viewed as affecting SJT scores via the same mechanism as personality measures.

Construct(s) Measured by SJTs

All of the information available to date is consistent with the interpretation of SJTs as measures of job-relevant knowledge or expertise (Schmidt & Hunter, 1993). 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 its 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. Individual differences in job-relevant knowledge, and thus individual differences in SJT scores, are expected to be directly affected by two antecedent variables: (1) relevant experience, and (2) ability to learn from this experience. However, other variables (e.g., personality traits) can be expected to affect SJT scores through their affect on one of these two direct antecedents. All of the results discussed to this point are consistent

with this model. In addition, 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.

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 SJTs. 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 were worded more broadly. Perhaps this leads to a broader array of experiences which are 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 concurrent validation research designs to assess the validity of SJTs. 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.

Finally, some research has demonstrated the usefulness of the SJT 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. However, interpreting SJTs as measures of job-related knowledge that is sometimes obtained on the job and sometimes obtained through general life experiences is consistent with both uses.

Conclusions

This paper suggests that SJTs are best viewed as a measurement method, rather than measures of 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. 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 (Motowidlo 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

- Arad, S. & Borman, W. C. (in preparation). A situational judgment test to predict adaptability in the U.S. Army. (Institute Report). Minneapolis: Personnel Decisions Research Institute.
- Barrett, G. V., Phillips, J. S., & Alexander, R. A. (1981). Concurrent and predictive validity designs: A critical reanalysis. *Journal of Applied Psychology*, 66, 1-6.
- Borman, W. C., Hanson, M. A., Oppler, S. H., Pulakos, E. D., & White, L. A. (1993). The role of early supervisory experience in supervisor performance. *Journal of Applied Psychology*, 78, 443-449.
- Bosshardt, M. J., & Cochran, C.C. (1996). *Development and validation of a selection system for financial advisors* (Technical Report No. 276). Minneapolis: Personnel Decisions Research Institutes, Inc.
- Bruce, M. M., & Learner, D. B. (1958). A supervisory practices test. *Personnel Psychology*, 11, 207-216.
- Campbell, J. P., Gasser, M. B., & Oswald, F. L. (1996). The substantive nature of job performance variability. In K. R. Murphy (Ed.), *Individual differences and behavior in organizations* (pp. 258-299), San Francisco: Jossey-Bass.
- Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 105, 317-327.
- Dallessio, A. T. (May, 1992). Predicting insurance agent turnover using a video-based situational judgment test. Paper presented at the seventh Annual Conference of the Society for Industrial and Organizational Psychology, Inc., Montreal, Quebec.
- Forehand G. A., & Guetzkow, H. (1961). The administrative judgment test as related to descriptions of executive judgment behaviors. *Journal of Applied Psychology*, 45 257-261.
- Hanson, M. A., & Borman, W. C. (1995). *Construct validation of a measure of supervisory job knowledge*. Poster presented at the 10th Annual Conference of the Society for Industrial and Organizational Psychology, Orlando, FL.
- Hedge, J. W., Hanson, M. A., Borman, W. C., Bruskiewicz, K. T., & Logan, K. K. (1996). *Predicting the crew resource management skills of Air Force pilots*. (Institute Report #283). Minneapolis: Personnel Decisions Research Institute.
- Helme, W. H. (1968). *Factor analysis of a situational leadership measure -- The speeded practical judgment test*. (Technical Research Note 202). U.S. Army Behavioral Science Research Laboratory.

- Houston, J. S. & Schneider, R. J. (in press). *Development and validation of a selection system for Nationwide insurance agents* (Technical Report No. 292). Minneapolis: Personnel Decisions Research Institutes, Inc.
- Hunter, J. E. (1983). A causal analysis of cognitive ability, job knowledge, job performance and supervisor ratings. In F. Landy, S. Zedeck & J. Cleveland (Eds.), *Performance measurement and theory* (pp. 257-266). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- Legree, P. J. (1995). Evidence for an oblique social intelligence factor established with a Likert-based testing procedure. *Intelligence*, 21, 247-266.
- McDaniel, M. A., Finnegan, E. B., Morgeson, F. P., & Campion, M. A. (1997). *Predicting job performance from common sense*. Paper presented the 12th annual Society for Industrial Organizational Psychology conference: St. Louis, MO.
- Motowidlo, S. J., Borman, W. C., & Schmit, M. J. (1997). A theory of individual differences in task and contextual performance. *Human Performance*, 10 (2), 71-84.
- Motowidlo, S. J., Dunnette, M. D., & Carter, G. W. (1990). An alternative selection procedures: The low fidelity simulation. *Journal of Applied Psychology*, 75, 640-647.
- Motowidlo, S. J., Hanson, M. A., & Crafts, J. L. (1997). Low-fidelity simulations. In D. L. Whetzel, G. R. Wheaton (Eds.), *Applied Measurement Methods in Industrial Psychology*. Palo Alto, CA: Consulting Psychologists Press.
- 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.
- Mowry, H. W. (1957). A measure of supervisory quality. *Journal of Applied Psychology*, 41, 405-408.
- Olson-Buchanan, J. B., Drasgow, F., Moberg, P. J., Mead, A. D., Keenan, P. A., & Donovan, M. A. (1994). Interactive video assessment of conflict resolution skills. *Personnel Psychology*, 51, 1-24
- Rosen, N. A. (1961). How Supervise? -- 1943-1960. *Personnel Psychology*, 14, 87-99.
- Schmidt, F. L., & Hunter, J. E. (1993). Tacit knowledge, practical intelligence, general mental ability, and job knowledge. *Current Directions in Psychological Science*. 8-9.
- Schmidt, F. L., Hunter, J. E., & Outerbridge, A. N. (1986). Impact of job experience and ability on job knowledge, work sample performance, and supervisory ratings of job performance. *Journal of Applied Psychology*, 71, 432-439.
- Smiderle, D., Perry, B. A., & Cronshaw, S. F. (1994). Evaluation of video-based assessment in transit operator selection. *Journal of Business and Psychology*, 1, 3-22.
- Tenopyr, M. L. (1969). The comparative validity of selected leadership scales relative to success in production management. *Personnel Psychology*, 22, 77-85.
- Weekly, J. A., & Jones, C. (1997). Video-based situational testing. *Personnel Psychology*, 50, 5-50.

INTERNET DOCUMENT INFORMATION FORM

A . Report Title: Situational Judgment: An Alternative Approach to Selection Test Development

B. DATE Report Downloaded From the Internet 4/21/99

C. Report's Point of Contact: (Name, Organization, Address, Office Symbol, & Ph #): Navy Education and Training
Professional Development and
Technology Center
Navy Advancement Center Dept
Dr. Grover Diesel, (850) 452-1815
6490 Saufley Field Road
Pensacola, FL 32509-5237

D. Currently Applicable Classification Level: Unclassified

E. Distribution Statement A: Approved for Public Release

F. The foregoing information was compiled and provided by:
DTIC-OCA, Initials: VM_ **Preparation Date:** 4/22/99__

The foregoing information should exactly correspond to the Title, Report Number, and the Date on the accompanying report document. If there are mismatches, or other questions, contact the above OCA Representative for resolution.