RUNNING HEAD: Knowledge and Behavioral Tendency

Situational Judgment Tests, Knowledge, Behavioral Tendency, and Validity: A Meta-Analysis

Michael A. McDaniel, Nathan S. Hartman, and W. Lee Grubb III

Virginia Commonwealth University

Paper presented at the 18th Annual Conference of the Society for Industrial and Organizational Psychology. Orlando. April, 2003

Correspondence on this article may be addressed to Michael McDaniel
E-mail: mamcdani@vcu.edu

Abstract

Situational judgment tests are personnel selection instruments that present job applicants with work-related situations and possible responses to the situations. The response instructions are of two types: behavioral tendency and knowledge. Tests with behavioral tendency instructions ask respondents to identify how they would likely behave in the situation. Tests with knowledge instructions ask respondents to evaluate the effectiveness of possible responses to the situation. Results showed that the response instructions influenced the constructs measured by the tests. Tests with knowledge instructions had higher magnitude correlations with cognitive ability than tests with behavioral tendency instructions. Tests with behavioral tendency instructions showed higher correlations than those with knowledge instructions for the personality constructs conscientiousness, agreeableness and emotional stability. Tests with knowledge instructions showed higher criterion-related validity than tests with behavioral tendency instructions.

Research on situational judgment tests for employee selection has increased dramatically in recent years (e.g., Chan & Schmitt, 1997; Clevenger, Pereira, Wiechmann, Schmitt, & Schmidt Harvey, 2001; McDaniel, Morgeson, Finnegan, Campion, & Braverman, 2001; McDaniel & Nguyen, 2001; Motowidlo, Dunnette, & Carter, 1990; Olson-Buchanan et al., 1998; Smith & McDaniel, 1998). Situational judgment tests present applicants with work-related situations and possible responses to the situations. Researchers contend that situational judgment tests predict performance because they measure job knowledge (Motowidlo, Borman, & Schmit, 1997), practical intelligence (Sternberg, Wagner, & Okagaki, 1993), or general cognitive ability (McDaniel et al., 2001).

The criterion-related validity of situational judgment tests has been evaluated in several primary studies (Chan & Schmitt, 1997; Hanson & Borman, 1989; Motowidlo et al., 1990; Smith & McDaniel, 1998) and in a recent meta-analysis the typical criterion-related validity was estimated at .34 (McDaniel et al., 2001). The construct validity of situational judgment tests also has been assessed in several primary studies (Beaty Jr. & Howard, 2001; Leaman & Vasilopoulos, 1997; Vasilopoulos, Reilly, & Leaman, 2000; Weekly & Jones, 1997, 1999). Meta-analyses of the construct-validity data show that situational judgment tests primarily assess conscientiousness, emotional stability, agreeableness (McDaniel & Nguyen, 2001) and cognitive ability (McDaniel et al., 2001).

In a review of situational judgment tests, McDaniel and Nguyen (2001) hypothesized that response instructions affect the test's susceptibility to faking and the constructs measured. An applicant is faking to the extent that the applicant is intentionally presenting herself as being more qualified for the job than she actually is. McDaniel and Nguyen (2001) proposed that situational judgment tests that use knowledge instructions would be less fakable than situational judgment tests that use behavioral tendency instructions. Knowledge response instructions ask respondents to select the correct or best possible answer or to rate the effectiveness of available answers. Behavioral tendency response instructions ask the respondent to select the answer that represents what the respondent would likely do or what they would most and least likely do in a given situation. McDaniel and Nguyen (2001) hypothesized that the knowledge instructions would be resistant to faking because the correct answers would be the same for honest and faking respondents. They argued that tests with behavioral tendency instructions are susceptible to faking because job applicants may be motivated to select responses that appear desirable although they differ from their typical behaviors at work. Faking-resistant tests can be important for test validity because it is difficult for a motivated respondent to improve one's test score. Consistent with this argument, Reynolds, Sydell, Scott and Winter (2000) found higher validities for less fakable situational judgment tests.

The variety of instructions in the existing situational judgment literature also drew the interest of Ployhart and Ehrhart (in press). They used a grade-point criterion and focused on the distinction between knowledge and behavioral tendency instructions. Ployhart and Ehrhart found that tests with knowledge instructions have lower criterion-related and construct validity than tests with behavioral tendency instructions.

The present study examined response instructions as moderators of the construct and criterion-related validity of situational judgment tests. We evaluated several hypotheses concerning the effects of response instructions on validity.

Hypotheses

Consistent with McDaniel and Nguyen (2001), we contend that situational judgment tests themselves are not constructs, but a method for measuring different constructs. The type of response instruction is expected to affect the degree to which various constructs are assessed by the situational judgment test. Tests that ask respondents to select the best possible answer are designed to assess knowledge constructs. Tests that ask respondents to select a response that represents what they would likely do are designed to assess behavioral tendency constructs. Tests with behavior tendency instructions and tests with knowledge response instructions will therefore assess constructs differentially. Specifically, situational judgment tests that use knowledge response instructions will have higher magnitude positive correlations with general cognitive ability than tests with behavior tendency response instructions because knowledge is a function of one's cognitive ability and opportunity to learn (Schmidt, Hunter, & Outerbridge, 1986). In addition, Nguyen (2001) found a knowledge instruction situational judgment test to be more correlated with cognitive ability than the same situational judgment test with behavioral tendency instructions. Conversely, personality tests should have a stronger relationship with situational judgment tests using behavioral tendency instructions because personality measures are typically self-reports of behavioral tendencies. Based on this discussion, we offer two hypotheses:

Hypothesis 1: Situational judgment tests with knowledge instructions will have higher correlations with cognitive ability than situational judgment tests with behavioral tendency instructions.

Hypothesis 2: Situational judgment tests with knowledge instructions will have lower correlations with personality tests than situational judgment tests with behavioral tendency instructions.

The opportunity for individuals to learn is often a function of job experience. On average, individuals with more job experience have had more opportunities to acquire job knowledge. McDaniel, Schmidt, and Hunter (1988) found that job experience has a positive asymptotic relationship with job performance. Thus, we expect length of job experience to be more highly correlated with tests with knowledge instructions, than tests with behavioral tendency instructions. There, we offer:

Hypothesis 3: Situational judgment tests with knowledge instructions will have higher correlations with job experience than situational judgment tests with behavioral tendency instructions.

Ployhart and Ehrhart (Ployhart & Ehrhart, in press) found situational judgment tests using behavioral tendency response instructions to be better predictors of

performance than knowledge response instructions. This was consistent with Motowidlo et. al. (1990) classic situational judgment test study where the seminal personnel selection thesis was invoked: past performance is the best predictor of future performance. Using this logic, one would argue that situational judgment tests using behavioral tendency response instructions would have a higher correlation with job performance than such tests with a knowledge response instruction.

However, there are several reasons why situational judgment tests with knowledge response instructions would yield higher validities for predicting job performance than situational judgment tests with behavioral tendency response instructions. First, knowledge measures are excellent predictors of job performance (Dye, Reck, & McDaniel, 1993). Second, if our construct hypotheses are correct, a knowledge instruction measure will be more highly correlated with cognitive ability. The validity of measures that tap multiple constructs, such as interviews (Huffcutt & McDaniel, 1996), are likely to increase as their cognitive load increases. Third, knowledge instruction situational judgment tests are more faking resistant than behavioral tendency situational judgment tests (Nguyen, 2001). We therefore offer two opposing hypotheses:

Hypothesis 4a: Situational judgment tests with behavioral tendency instructions will have higher criterion-related validity than situational judgment tests with knowledge instructions.

Hypothesis 4b: Situational judgment tests with knowledge instructions will have higher criterion-related validity than situational judgment tests with behavioral tendency instructions.

Method

Literature search

The data set from the McDaniel et al. (2001) meta-analysis provided this study the bulk of its criterion-related data, as well as its construct validity data concerning cognitive ability. The McDaniel and Nguyen (2001) data set examining construct-related validity of the Big 5 and job experience provided the majority of the data for the non-g construct validity analysis. The reference lists developed from these studies were updated to represent the most comprehensive list of studies using situational judgment tests. In order to supplement our reference list, we contacted several researchers working in this area and asked them to identify any journal articles, dissertations, conference papers, or technical reports missing from our list of references. Several of the contacted researchers provided additional studies that were added to the data set used in this meta-analysis. In addition, we reviewed recent journals and the programs of recent conferences.

Analysis method

The psychometric artifact-distribution meta-analytic method was used in this study (Hunter & Schmidt, 1990). For the construct validity analyses, the estimated

population mean correlation is corrected for measurement error in both the situational judgment tests and the construct. The estimated population variance is corrected for sampling error and differences across studies in measurement error in both variables. No range restriction corrections were conducted.

For the criterion-related validity analyses, the estimated population mean correlation is corrected for measurement error in the criterion. The estimated population variance is corrected for sampling error and differences across studies in measurement error in the situational judgment tests and in the criterion. No range restriction corrections were conducted.

Reliability

The reliability artifact distributions for the situational judgment tests and the job performance criterion were drawn from McDaniel et. al. (2001). The reliability distribution for the personality constructs were drawn from studies in the meta-analysis and are presented in Table 1. The reliability of the length of experience measure was assumed to be one.

Response instruction taxonomy

A taxonomy of response instruction types was developed and is presented in Table 2. The taxonomy is hierarchical. At the highest level of the taxonomy, we distinguish between knowledge and behavioral tendency instructions. At the lowest level of the taxonomy, we show variation in knowledge instructions and variations in behavioral tendency instructions.

Insert Table 1 through 5 here

Decision rules

Analysis of the criterion related validity data generally followed the four main decision rules used in the original McDaniel et al. (2001) meta-analysis. First, only studies whose participants were employees or applicants were used**.** Second, situational judgment tests were required to be in the paper-and-pencil format. Other formats such as interview (e.g., Latham, Saari, Pursell, & Campion, 1980) and video-based situational judgment tests (e.g., Dalessio, 1994; Jones & DeCotiis, 1986) were not included. The third rule gave priority to supervisor ratings of job performance over other available measures. The fourth and final rule defined boundary conditions on the job performance measure. Surrogate performance measures such as years of school completed, hierarchical level in the organization, number of employees supervised, years of management experience, ratings of potential, and salary were not used. The rules concerning construct-related data generally followed the decision rules used in McDaniel and Nguyen (2001).

To improve the current study, we made a few changes to the decision rules used by McDaniel et al. (2001) and McDaniel & Nguyen (2001). First, we did not include studies using the *How Supervise?* test (File, 1943) in either the criterion-related and construct validity analysis. This was done because the items in *How Supervise?* were substantially different than those found in other situational judgment tests. The *How Supervise?* test presents respondents with single statements about supervisory practices and asks whether they agree with the statement. As an example, we present a paraphrase of an item: "Employees should be told about the financial status of the company." Note that this item lacks a stem containing a situation, and asks respondents only to rate a single statement as desirable or undesirable. Thus, *How Supervise?* items lack both a situation and a series of responses to the situations. Given these differences from other tests, we chose not to include *How Supervise?* in the collection of situational judgment tests examined. We also excluded studies where the situational judgment response instructions could not be coded due to lack of information or due to the response instructions not fitting into the taxonomy (Jones, Dwight, & Nouryan, 1999).

In contrast to the decision rules of McDaniel and Nguyen (2001), an additional rule was added for studies containing data with multiple measures of the same construct in a single sample. We reported one coefficient per construct, per sample. For example, if a single study reported two correlations for conscientiousness from the same sample, we only included one coefficient which was the mean of the two coefficients. An exception to this rule was made for the Thumin and Page (1966) study because their two correlations for length of experience were from two separate situational judgment tests administered in the same study. Another exception was Weekly and Jones (1997). Although, their two correlations for experience were from the same sample, the two correlations were included because one coefficient was from the empirically-keyed version of the test and the other coefficient was from the rationally-keyed version of the test. Because the correlation between the two tests was not very high (r = .53), we considered the empirically and rationally keyed tests to be different.

Results

Construct-validity results

Tables 3 and 4 present the construct validities. In both tables, the first column identifies the distribution of validities analyzed. Total sample size across studies, the total number of correlation coefficients, and the mean and standard deviation of the observed distribution are presented in columns 2 through 5. Columns 6 and 7 contain estimates of the population mean correlations and standard deviations. The percentage of variance in the observed distribution corrected for sampling error and reliability differences across studies and the 80% credibility interval for the true validities are presented in the remaining two columns. Whereas no corrections for range restriction were conducted, all reported populations correlations are likely to be downwardly biased (underestimates of the actual population value).

Table 3 presents the results for the correlation between situational judgment tests and general cognitive ability means. The estimated population correlation is (.39) which

is consistent with the McDaniel et al. (2001) results. However, the removal of one large sample study (Pereira & Schmidt Harvey, 1999, study 2) raised the correlation to (.45). Situational judgment tests with knowledge instructions had almost twice the correlation with cognitive ability (.43) than situational judgment tests with behavioral tendency instructions (.23). Situational judgment tests with "pick the best" instructions had a substantially higher correlation with cognitive ability (.71) than tests with "pick the best and worst" instructions (.51). Any interpretation of this difference should be tempered by the knowledge that the "best and worst" distribution contained only five coefficients.

Table 4 presents the results for the correlations between situational judgment tests with Big 5 measures and length of job experience. The estimated mean population correlations between situational judgment tests and the Big 5 were (.33) for agreeableness, (.37) for conscientiousness, (.41) for emotional stability, (.20) for extroversion, and (.12) for openness to experience. For three of the Big 5, the correlations between the situational judgment test and the personality trait were substantially higher for the behavioral tendency than for the knowledge instruction set: agreeableness (.53 vs .20), conscientiousness (.51 vs. 33), and emotional stability (.51 vs .11). The correlation between situational judgment tests and length of job experience was (.04). Almost all the data for the length of experience distribution were for situational judgment tests with knowledge instructions and thus it is not surprising that for tests with knowledge instructions, the correlation with job experience was .03. The correlation between experience and situational judgment tests with behavioral tendency instructions was .17. We note that some of the distributions have a relatively small number of coefficients.

Table 5 presents the criterion-related validity results. The first column identifies the distribution of validities analyzed. Total sample size across studies, the total number of correlation coefficients, and the mean and standard deviation of the observed distribution are presented in columns 2 through 5. Columns 6 and 7 contain estimates of the population mean correlations and standard deviations. The percentage of variance in the observed distribution corrected for sampling error and reliability differences across studies and the 80% credibility interval for the true validities are presented in the remaining two columns. Whereas no corrections for range restriction were conducted, all reported populations correlations are likely to be downwardly biased (underestimates of the actual population value).

The estimated population correlation of .32 is comparable to the .34 coefficient reported in McDaniel et al. (2001). Situational tests using behavioral tendency instructions have lower validity than tests using knowledge instructions (.27 vs. .33). The knowledge response instruction tests provided the bulk of the data and we were able to examine the validity of two sub-category tests that varied by the type of knowledge instructions. Tests that required the respondent to pick the best answer had an estimated mean population correlation of .40. Twenty-one of the 32 correlation coefficients in this distribution were from Greenberg (1963). If those tests are removed from the distribution, the estimated population mean drops to .36. Tests that required the respondent to pick both the best and the worst response had an estimated mean population correlation of .27.

The incremental validity of situational judgment tests over measures of general mental ability has been a topic of several studies (Clevenger et al., 2001; Chan & Schmitt, 2002; Weekly & Jones, 1997, 1999). McDaniel et al. (2001) used the results of their meta-analysis to estimate the correlation among general cognitive ability, situational judgment tests, and job performance and then provided an estimate of the incremental validity of situational judgment tests over that of general cognitive ability. Based on the data in Table 5, we replicated that analysis using .25 as the observed validity for situational judgment tests with knowledge instructions and .21 as the observed validity for situational judgment tests with behavioral tendency instructions. We need to identify the correlation between situational judgment tests by instruction type and general cognitive ability. For all samples, the estimated correlation between general mental ability and situational judgment tests with knowledge instructions was (.34), and if we drop the Pereira and Harvey (1999, Study 2) sample, the correlation rises to (.43). Given this large disparity, we ran the incremental validity analysis twice for situational judgment tests with knowledge. The correlation between general cognitive ability and situational judgment tests with behavioral tendency instructions was (.18). The correlations between general cognitive ability and job performance was assumed to be .25.

On the basis of these values, the estimated validity of a composite of a situational judgment test with knowledge instructions and a general cognitive ability measure was (.30) or (.31) depending on whether the correlation with general ability is set at (.43) or (.34). The resulting value for situational judgment tests with behavioral tendency instructions is (.31). Based on an assumed criterion reliability of .60, the estimated population correlations would be .39 (for r = .30) and .40 (for r = .31).

Discussion

This study rests on the assumption that response instructions affect the construct and criterion-related validity of situational judgment tests. We drew a distinction between response instructions that asked for a knowledge response and instructions that asked for a behavioral tendency response. We evaluated our assumption concerning the affect of this response instruction distinction by conducting analyses to examine four hypotheses.

Evaluation of hypotheses

Hypothesis 1 argued that situational judgment tests with knowledge instructions would be more highly correlated with general cognitive ability than situational judgment tests with behavioral tendency instructions. This hypothesis received strong support in that tests with knowledge response instructions yielded higher validities than tests with behavioral tendency response instructions (.43 vs. .23).

Hypothesis 2 held that situational judgment measurers with behavioral tendency instructions would be more highly correlated with personality measures than situational judgment tests with knowledge instructions. This hypothesis received strong support for the constructs of agreeableness (.53 vs. .20), conscientiousness (.51 vs. .33), emotional

stability (.51 vs .11). Hypothesis 2 was not supported for extroversion (.11 vs. .21) and there were too little data to evaluate the hypothesis for openness to experience. We note that many of the comparisons are based on a relatively small number of coefficients and should be re-evaluated as more data are accumulated.

Hypothesis 3 posited that situational judgment tests with knowledge instructions would be more highly correlated with length of job experience measures than situational judgment measurers with behavioral tendency instructions. This hypothesis received no support in that tests with knowledge instructions yielded lower-magnitude correlations than situational judgment tests with behavioral tendency instructions (.03 vs .17). We do not find that test of the hypothesis to be compelling because there were only four coefficients in the distribution of correlations for behavioral tendency instruction tests with job experience measures. Thus, conclusions regarding this hypothesis should await the accumulation of additional data.

Hypothesis 4a stated that the validity of situational judgment tests would be higher for those with behavioral tendency instructions than for those with knowledge instructions. Hypothesis 4b held the opposite. Hypothesis 4b was supported in that situational judgment tests with knowledge instructions had greater validity (.33) than situational judgment tests with behavioral tendency instructions (.27).

Some believe that our findings are provocative because it appears that one can alter the construct and criterion-related validity of a situational judgment test simply by changing the instructions. However, there is an alternative hypothesis. This hypothesis asserts that there is something beyond response instructions that distinguishes tests with knowledge instructions from tests with behavioral tendency instructions. This yet to be discovered "something" might relate to the content of the scenarios in the items. We offer six lines of evidence refuting this alternative hypothesis. First, we examined items from tests with both types of instructions and discovered no systematic differences in the content areas covered by the tests. Second, in our consulting work in this area, the items are often developed before the client makes a decision concerning which type of instructions to use. Third, when subject matter experts are used to generate critical incidents as part of the question stem development process, they are not usually informed of the instructions to be used with the final test. Fourth, Nguyen (2001) administered the same situational judgment test to a sample of college students twice, once with knowledge instructions and once with behavioral tendency instructions. Nguyen (2001) found the test with knowledge instructions was more highly correlated with general cognitive ability than the same test administered with behavioral tendency instructions. However, the correlations with personality scales did not show a clear distinction between the two instruction sets. Fifth, Ployhart and Ehrhart (in press) administered the same set of situational judgment items under various instruction sets that could be classified as either knowledge or behavioral tendency. All analyses showed superior results for the tests with behavioral tendency instructions. Although Ployhart and Erhart's study in the educational domain provided different results than our results in the employment domain, they demonstrated that simply changing the instructions produced large effects for the validity of the test. Sixth, Griffith, Frei, Snell, Hamill, and Wheeler

(1997) demonstrated that the factor structure of a non-cognitive test battery changed substantially as a function of whether applicants were warned against faking. Based on these six lines of evidence, we conclude that simply changing the instructions on a situational judgment test can substantially alter its criterion-related and construct validity. However, additional research on this issue is encouraged. Specifically, we encourage additional studies similar to Nguyen (2001) and Ployhart and Erhart (in press) where the same test items are administered under different response instructions.

Incremental validity

The incremental validity of a predictor is a function of the validity of each predictor and the redundancy of the criterion space explained by the predictors. The redundancy of the criterion space explained is influenced by the correlation among the predictors. Our data show that tests with knowledge instructions have larger correlations with general cognitive ability and larger criterion-related validity than do tests with behavioral tendency instructions. If one's aim is to maximize the criterion-related validity of the composite of a situational judgment test and a cognitive ability test, the distinction between knowledge and behavioral tendency instructions is not important because either composite yields about the same level of validity (about .31). We caution the reader about and over interpretation of this result. A given situational judgment test may have difference correlations with general cognitive ability and job performance than the mean values used in this analysis. Likewise, the validity of cognitive ability tests vary by job complexity and will not always be .25 as assumed in our analysis. Future research should examine whether the race differences in composite test scores vary as a function of the response instructions for the situational judgment test. One might expect a composite containing tests with behavioral tendency instructions to have smaller race differences due to the lower correlation with cognitive ability and higher correlations with personality variables.

Boundary conditions for validity generalization

This meta-analysis reviewed validity evidence for situational judgment tests, a measurement method. The draft *Principles for the Validation and Use of Personnel Selection Procedures* (SIOP, 2002) assert that meta-analyses of methods involve more interpretational difficulties than meta-analyses of single construct measures. The Principles use the example of a meta-analysis of an interview to illustrate issues and conclude that "a cumulative database on interviews where content, structure, and scoring are coded could support generalization to an interview meeting the same specification" (page 47). Therefore, with respect to situational judgment tests, the *Principles* require that the boundary conditions of the meta-analysis be specified with respect to the content, structure, and scoring of the tests within the meta-analysis. For this meta-analysis, we specify these boundary conditions as follows.

With respect to content, all situational judgment tests included in this meta-analysis were designed to have job-related content. Further, the criterion-related validity evidence was collected for the jobs for which the tests were developed. If a situational

judgment test is designed by someone without knowledge of the job or who did not rely on the knowledge of those familiar with the job (e.g., subject matter experts), the validity of the test could not be inferred from the results of this meta-analysis. Likewise, if a situational judgment test was built to be job-related for a given job or class of jobs (e.g., supervisors), but was administered to another job or class of jobs (e.g., non-supervisors), the validity of the test could not be inferred from the results of this meta-analysis.

With respect to structure, all the situational judgment tests included in this meta-analysis were written multiple choice tests where the stem of the item was a situation and the respondents were given either knowledge or behavioral tendency response instructions. Thus, the validity of a video-based situational judgment test could not necessarily be inferred from this meta-analysis. If a test, such as *How Supervise?*, did not include a situation in the stem, one could not infer its validity from the results of this meta-analysis. If the response instruction could not be categorized as a knowledge or a behavioral tendency instruction (Jones et al., 1999), one cannot infer the test's validity from this meta-analysis.

With respect to scoring, all situational judgment tests in this meta-analysis were objectively scored (e.g., there are a limited number of response options and some are more correct than others). Thus, one would not generalize these results to an open-ended situational judgment test in which one is to generate a response to a given situation. Some scoring keys were rationally-developed by individuals with job knowledge and others were empirically-keyed. Thus, the validity of a test that is rationally-keyed by someone without job knowledge could not be inferred from the results of this meta-analysis.

In brief, the validity of a situational judgment test could be inferred from this meta-analysis if the test had job-related content, is used for the job or job family for which it was designed, was a paper-and-pencil formatted test with either knowledge or behavioral tendency instructions, contained items with a situation as the stem and possible responses to the situation as item options, and where the test is scored objectively using either a rationally-developed or an empirically-developed key. If a test falls within these boundaries, this meta-analysis would be a very useful source of information for validity generalization.

Conclusion

The use of situational judgment tests in personnel selection has gained increasing interest recently. This study builds upon earlier meta-analyses of this literature by examining the moderating effect of response instructions. This response instruction moderator is shown to explain differences across studies in both construct validity and criterion-related validity. Although these analyses should be repeated as additional data cumulate, our results are currently the best estimates of the validity of situational judgment tests. These results can guide the development of future situational judgment tests and provide evidence for the validity of tests that fall in the boundary of the studies examined in this research.

References

*Beaty Jr., J. C., & Howard, M. J. (2001). *Constructs measured in situational judgment tests designed to assess management potential and customer service potential.* Paper presented at the 16th Annual Conference of the Society for Industrial and Organizational Psychology, San Diego, CA.

*Bosshardt, M., & Cochran. (1996). *Proprietary document.* Personal communication to Michael McDaniel.

Bruce, M. M. (1953a). Examiner's manual for the business judgement test. New Rochelle, New York: Martin M. Bruce.

*Bruce, M. M. (1953b). The prediction of effectiveness as a factory foreman. *Psychological Monographs: General and Applied, 67*, 1-17.

*Bruce, M. M. (1965a). Business judgment test (pp. 1-3).

*Bruce, M. M. (1965b). Examiner's manual, business judgment test, revised. 1-12.

*Bruce, M. M., & Friesen, E. P. (1956). Validity information exchange. *Personnel Psychology, 9*, 380.

*Bruce, M. M., & Learner, D. B. (1958). A supervisory practices test. *Personnel Psychology, 11*, 207-216.

Cardall, A. J. (1942). *Preliminary manual for the test of practical judgment*. Chicago, Illinois: Science Research Associates Inc.

*Carrington, D. H. (1949). Note on the Cardall practical judgment test. *Journal of Applied Psychology, 33*, 29-30.

*Chan, D. (2002). Situational judgment and job performance. *Human Performance, 15*(3), 233-254.

*Chan, D. (2002). *Situational judgment effectiveness X proactive personality interaction on job performance.* Paper presented at the 17th Annual Conference of the Society for Industrial and Organizational Psychology, Toronto, Canada.

Chan, D., & Schmitt, N. (1997). Video-based versus paper-and-pencil method of assessment in situational judgment tests: Subgroup differences in test performance and face validity perceptions. *Journal of Applied Psychology, 82*, 143-159.

*Chan, D., & Schmitt, N. (2002). Situational judgment and job performance. *Human Performance, 15*, 233-254.

*Clevenger, J., Jockin, T., Morris, S., & Anselmi, T. (1999). *A situational judgment test for engineers: Construct and criterion related validity of a less adverse alternative.* Paper presented at the 14th Annual Conference of the Society for Industrial and Organizational Psychology, Atlanta, Georgia.

*Clevenger, J., Pereira, G. M., Wiechmann, D., Schmitt, N., & Schmidt Harvey, V. (2001). Incremental validity of situational judgment tests. *Journal of Applied Psychology, 86*(3), 410-417.

*Clevenger, J. P., & Haaland, D. E. (2000). *The relationship between job knowledge and situational judgment test performance*. New Orleans: Aon Consulting.

*Corts, D. B. (1980). *Development of a procedure for examining trades and labor applicants for promotion to first-line supervisor*. Washington, D.C.: U. S. Office of Personnel Management, Personnel Research and Development Center, Research Branch.

Dalessio, A. T. (1994). Predicting insurance agent turnover using a video-based situational judgement test. *Journal of Business and Psychology, 9*, 23-32.

*Dulsky, S. G., & Krout, M. H. (1950). Predicting promotion potential on the basis of psychological tests. *Personnel Psychology, 3*, 345-351.

Dye, D. A., Reck, M., & McDaniel, M. A. (1993). Moderators of the validity of written job knowledge measures. *International Journal of Selection Assessment, 1*, 153-157.

File, Q. W. (1943). *How Supervise?* (Questionnaire Form B). New York, NY: The Psychological Corporation.

*Greenberg, S. H. (1963). *Supervisory judgment test manual* (Technical Series No. 35). Washington, D.C.: Personnel Measurement Research and Development Center, U.S. Civil Service Commission, Bureau of Programs and Standards, Standards Division.

Griffith, R.L., Frei, R.L., Snell, A.F., Hamill, L.S., & Wheeler, J.K. (1997, April). Warnings versus no-warnings: Differential effect of method bias. Paper presented at the 12th Annual Conference of the Socirty for Industrial and Organizational Psychology, Sty. Louis, MO.

Hanson, M. A. (1994). *Development and construct validation of a situational judgment test of supervisory effectiveness for first-line supervisors in the U.S. Army.* Unpublished Doctoral, University of Minnesota.

Hanson, M. A., & Borman, W. C. (1989). *Development and construct validation of a situational judgment test of supervisory effectiveness for first-line supervisors in*

the U.S. Army. Paper presented at the Symposium conducted at the 4th Annual Conference of the Society for Industrial and Organizational Psychology, Atlanta, Georgia.

*Hilton, A. C., Bolin, S. F., Parker Jr., J. W., Taylor, E. K., & Walker, W. B. (1955). The validity of personnel assessments by professional psychologists. *Journal of Applied Psychology, 39*(4), 287-293.

Huffcutt, J. M., & McDaniel, M. A. (1996). A meta-analytic investigation of cognitive ability in employment interview evaluations: Moderating characteristics and implications for incremental validity. *Journal of Applied Psychology, 81*, 459-473.

*Hunter, D. R. (2002). *Measuring general aviation pilot decision-making using a situational judgment technique*. Washington, DC: Federal Aviation Administration.

*Jagmin, N. (1985). *Individual differences in perceptual/cognitive constructions of job-relevant situations as a predictor of assessment center success.* University of Maryland, College Park, MD.

Jones, C., & DeCotiis, T. A. (1986). Video-assisted selection of hospitality employees. *The Cornell H.R.A. Quarterly*, 67-73.

Jones, M. W., Dwight, S. A., & Nouryan, T. R. (1999). *Exploration of the construct validity of a situational judgment test used for managerial assessment.* Paper presented at the 14th annual conference of the Society for Industrial and Organizational Psychology, Atlanta, Georgia.

Jurgensen, C. E. (1959). Supervisory Practices Test. In O. K. Buros (Ed.), *The Fifth Mental Measurements Yearbook - Test & Reviews: Vocations - Specific Vocations* (pp. 946-947). Highland Park, NJ: Gryphon Press.

Latham, G. P., Saari, L. M., Pursell, E. D., & Campion, M. A. (1980). The situational interview. *Journal of Applied Psychology, 65*, 422-427.

*Leaman, J. A., & Vasilopoulos, N. L. (1997). *Development and validation of the U.S. immigration officer applicant assessment.* Washington, D.C.: U.S. Immigration and Naturalization Service.

*Leaman, J. A., & Vasilopoulos, N. L. (1998). *Development and validation of the detention enforcement officer applicant assessment*. Washington, D.C.: U.S. Immigration and Naturalization Service.

*Leaman, J. A., Vasilopoulos, N. L., & Usala, P. D. (1996). *Beyond integrity testing: Screening border patrol applicants for counterproductive behaviors.* Paper

presented at the 104th Annual Convention of the American Psychological Association, as part of the symposium, Multidimensional Assessment for Selecting Border Patrol Agents., U.S. Immigration and Naturalization Service, Washington, DC.

*Lobsenz, R. E., & Morris, S. B. (1999). *Is tacit knowledge distinct from g, personality, and social knowledge?* Paper presented at the 14th Annual Society for Industrial and Organizational Psychology, Atlanta, Georgia.

*MacLane, C. N., Barton, M. G., Holloway-Lundy, A. E., & Nickels, B. J. (2001). *Keeping score: Empirical vs. expert weights on situational judgment responses.* Paper presented at the 16th Annual conference of the Society for Industrial and Organizational Psychology, San Diego, California.

McDaniel, M. A., Morgeson, F. P., Finnegan, E. B., Campion, M. A., & Braverman, E. P. (2001). Predicting job performance using situational judgment tests: A clarification of the literature. *Journal of Applied Psychology, 86*, 730-740.

McDaniel, M. A., & Nguyen, N. T. (2001). Situational judgment tests: A review of practice and constructs assessed. *International Journal of Selection and Assessment, 9*, 103-113.

McDaniel, M. A., Schmidt, F. L., & Hunter, J. E. (1988). Job experience correlates of job performance. *Journal of Applied Psychology, 73*, 327-330.

*Meyer, H. H. (1951). Factors related to success in the human relations aspect of work-group leadership. *Psychological Monographs: General and Applied, 65*, 1-29.

*Moss, F. A. (1926). Do you know how to get along with people? *Scientific American, 135*, 26-27.

*Motowidlo, S. (1991). *The situational inventory: A low fidelity for employee selection*. Gainesville: University of Florida.

Motowidlo, S., Borman, W., & Schmit, M. (1997). A theory of individual differences in task and contextual performance. *Human Performance, 10*, 71-83.

Motowidlo, S., Dunnette, M. D., & Carter, G. W. (1990). An alternative selection procedure: The low-fidelity simulation. *Journal of Applied Psychology, 75*, 640-647.

*Mowry, H. W. (1957). A measure of supervisory quality. *Journal of Applied Psychology, 41*, 405-408.

*Mullins, M. E., & Schmitt, N. (1998). *Situational judgment testing: Will the real constructs pleas present themselves.* Paper presented at the 13th Annual

Conference of the Society for Industrial and Organizational Psychology, Dallas, TX.

Nguyen, N. T. (2001). *Faking in situational judgment tests: An empirical investigation of the Work Judgment Survey.* Unpublished Doctoral Dissertation, Virginia Commonwealth University, Richmond.

*O'Connell, M. S., McDaniel, M. A., Grubb, W. L., III, Hartman, N. S., & Lawrence, A. (2002). *Incremental validity of situational judgment tests for task and contextual performance.* Paper presented at the 17th Annual Conference of the Society of Industrial Organizational Psychology, Toronto, Ontario.

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.

*Pereira, G. M., & Schmidt Harvey, V. (1999). *Situational judgment tests: Do they measure ability, personality or both.* Paper presented at the 14th Annual Conference of the Society for Industrial and Organizational Psychology, Atlanta, Georgia.

*Ployhart, R. E., & Ehrhart, M. G. (in press). 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*.

*Pulakos, E. D., & Schmitt, N. (1996). An evaluation of two strategies for reducing adverse impact and their effects on criterion-related validity. *Human Performance, 9*, 241-258.

Reynolds, D. H., Sydell, E. J., Scott, D. R., & Winter, J. L. (2000). *Factors affecting situational judgment test characteristics.* Paper presented at the 15th Annual Conference of the Society for Industrial and Organizational Psychology, New Orleans, LA.

*Richardson, B., Henry & CO., INC. (1949). Test of supervisory judgment (pp. 1-4).

*Richardson, B., Henry & CO., INC. (1985). *Manager profile record*: National Computer Systems, Inc.

*Rusmore, J. T. (1958). A note on the "test of practical judgment". *Personnel Psychology, 11*, 37.

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.

*Smith, K. C., & McDaniel, M. A. (1998). *Criterion and construct validity evidence for a situational judgment measure.* Paper presented at the 13th Annual Convention of the Society for Industrial and Organizational Psychology, Dallas.

*Spitzer, M. E., & McNamara, W. J. (1964). A managerial selection study. *Personnel Psychology*, 19-40.

*Stees, Y., & Turnage, J. J. (1992). *Video selection tests: Are they better?* Paper presented at the ??th Annual Conference of the Society for Industrial and Organizational Psychology, Montreal, Quebec.

Sternberg, R. J., Wagner, R. K., & Okagaki, L. (1993). *Practical intelligence: The nature and role of tacit knowledge in work and at school.* Yale University, New Haven, Ct.

*Stevens, M. J., & Campion, M. A. (1999). Staffing work teams: Development and validation of a selection test for teamwork settings. *Journal of Management, 25*, 207-228.

*Thumin, F. J., & Page, D. S. (1966). A comparative study of two test of supervisory knowledge. *Psychological Reports, 18*, 535-538.

*Vasilopoulos, N. L., Reilly, R. R., & Leaman, J. A. (2000). The influence of job familiarity and impression management on self-report measure scale scores and response latencies. *Journal of Applied Psychology, 85*, 50-64.

*Wagner, R. K., & Sternberg, R. J. (1985). Practical Intelligence in real-world pursuits: The role of tacit knowledge. *Journal of Personality and Social Psychology, 49*, 436-458.

*Wagner, R. K., & Sternberg, R. J. (1991). *The common sense manager*. San Antonio: Psychological Corporation.

*Watley, D. J., & Martin, H. T. (1962). Prediction of academic success in a college of business administration. *Personnel and Guidance Journal*, 147-154.

*Weekly, J. A., & Jones, C. (1997). Video-based situational testing. *Personnel Psychology, 50*, 25-49.

*Weekly, J. A., & Jones, C. (1999). Further studies of situational tests. *Personnel Psychology, 52*, 679-700.

Table 1.
Reliability Distributions and Frequencies.

Reliability	Frequency
0.63	2
0.66	2
0.67	1
0.69	1
0.70	1
0.72	1
0.73	1
0.74	2
0.76	5
0.77	3
0.78	3
0.79	2
0.80	3
0.81	2
0.82	2
0.84	2
0.87	1

Table 2.
Response Instruction Taxonomy

Response Instruction	Illustrative Study
Behavioral Tendency Instructions	
1. Would Do	Bruce (1953a)
2. Most & Least Likely	Pulakos & Schmitt (1996)
3. Rate and rank what you would most likely do	Jagmin (1985)
Knowledge Instructions	
4. Should Do	Hanson (1994)
5. Best Response	Greenberg (1963)
6. Best & Worst	Clevenger & Haaland (2000)
7. Best & Second Best	Richardson, Bellows, Henry & Co. (undated)
8. Rate Effectiveness	Chan & Schmitt (1997)
9. Best, Second, & Third Best	Cardell (1942)
10. Level of Importance	Corts (1980)

Table 3
Meta-Analytic Results of Correlations Between Situational Judgment Tests and General Cognitive Ability Measures

	Observed distribution				Population distribution			
Distribution	N	No. of *r*s	Mean *r*	σ	ρ	σ_p	% of σ_p^2 due to artifacts	80% CI
All coefficients	22,553	62	.30	.19	.39	.23	8	.08 to .59
All coefficients Excluding Pereira & Harvey's (1999) Study 2	16,967	61	.35	.19	.45	.23	10	.16 to .74
Behavioral tendency responses	5,263	21	.18	.13	.23	.15	20	.03 to .42
Knowledge responses	17,290	41	.34	.18	.43	.23	7	.14 to .72
Knowledge responses All coefficients Excluding Pereira & Harvey's (1999) Study 2	11,704	40	.43	.15	.55	.17	16	.33 to .78
Knowledge responses – Pick the best	4,568	21	.55	.12	.71	.13	29	.54 to .87
Knowledge responses- Pick the best excluding Greenberg (1963)	2,079	4	.47	.07	.60	.06	59	.53 to .68
Knowledge responses – Pick the best and worst	3,017	5	.40	.04	.51	.00	100	.51 to .51

Table 4
Meta-Analytic Results of Correlations Between Situational Judgment Tests with Big 5 measures and Length of Job Experience

	Observed distribution				Population distribution			
Distribution	N	No. of *r*s	Mean *r*	σ	ρ	σ_p	% of σ_p^2 due to artifacts	80% CI
Agreeableness	14,131	16	.25	.18	.33	.23	5	-.04 to .62
Knowledge	8,303	5	.15	.09	.20	.12	8	.04 to .35
Behavioral tendency	5,828	11	.40	.16	.53	.20	11	.26 to .79
Conscientiousness	19,656	19	.28	.12	.37	.15	11	.18 to .56
Knowledge	13,754	8	.25	.08	.33	.10	15	.21 to .46
Behavioral tendency	5,902	11	.36	.15	.51	.19	11	.27 to .75
Emotional stability	7,718	14	.31	.19	.41	.24	6	.09 to .72
Knowledge	1,990	4	.08	.07	.11	.07	40	.01 to .20
Behavioral tendency	5,728	10	.39	.15	.51	.19	11	.27 to .75
Extroversion	12,607	10	.15	.07	.20	.08	20	.10 to .30
Knowledge	11,867	5	.16	.05	.21	.06	23	.14 to .28
Behavioral tendency	740	5	.08	.18	.11	.21	19	-.16 to .38
Openness to experience	874	5	.09	.08	.12	.02	97	.09 to .14
Knowledge	160	1	.19	-	.25	-	-	-
Behavioral tendency	714	4	.07	.07	.09	.00	100	.09 to.09
Length of experience correlations	12,742	21	.03	.12	.04	.13	11	-.13 to .21
Knowledge	12,003	17	.02	.12	.03	.13	9	-.14 to .20
Behavioral tendency	739	4	.15	.08	.17	.04	78	.11 to .23

Table 5.
Meta-Analytic Results of Criterion-Related Validity of Situational Judgment Tests

	Observed distribution				Population distribution			
Distribution	N	No. of *r*s	Mean *r*	σ	ρ	σ_p	% of σ^2 due to artifacts	90% CV
All correlations	11,809	84	.24	.13	.32	.13	40	.15
Behavioral tendency responses	2,610	21	.21	.13	.27	.13	42	.11
Knowledge responses	9,199	63	.25	.13	.33	.13	41	.16
Knowledge responses – Pick the best	4,231	32	.30	.11	.40	.08	65	.29
Knowledge responses – Pick the best without Greenberg (1963)	1,248	11	.27	.15	.36	.16	34	.15
Knowledge responses – Pick the best and worst	1,261	16	.21	.15	.27	.13	54	.10