



## Situational judgment test research: Informing the debate on practical intelligence theory

Michael A. McDaniel<sup>a,\*</sup>, Deborah L. Whetzel<sup>b</sup>

<sup>a</sup>*Virginia Commonwealth University, VA, United States*

<sup>b</sup>*Work Skills First, Inc., VA, United States*

Received 30 December 2004; received in revised form 18 January 2005; accepted 5 February 2005

Available online 9 April 2005

---

### Abstract

[Gottfredson, L.S. (2003). Dissecting practical intelligence theory: Its claims and evidence. *Intelligence*, 31, 343–397.] provided a detailed critique of Sternberg's [Sternberg, R.J., Fotsythe, G.B., Hedlund, J., Horvath, J.A., Wagner, R.K., Williams, W.M., Snook, S.A., Grigorenko, E.L. (2000). *Practical intelligence in everyday life*. New York: Cambridge University Press.] practical intelligence theory. The current paper seeks to supplement Gottfredson's extensive critique using research and practice in the situational judgment literature in industrial/organizational psychology to inform the debate concerning practical intelligence. This paper makes four major points. First, there is a research and practice tradition in personnel selection that has used the item type that Sternberg uses to assess what he calls tacit knowledge. This item type is called situational judgment in the personnel selection field. There is an extensive literature on situational judgment tests that can inform debate concerning practical intelligence. Second, we present evidence that situational judgment tests do not measure a general factor, whether labeled practical intelligence, or something else. Both the individual test items and the tests are factorially complex and measure multiple known constructs. Third, since the measures assess multiple constructs (*g* and personality), they are best viewed as measurement methods. Fourth, we compare the validity of the measures with *g* for the prediction of job performance and examine the incremental validity of situational judgment tests. We conclude that this research can help guide a revision to practical intelligence theory.

© 2005 Elsevier Inc. All rights reserved.

*Keywords:* Practical intelligence; Situational judgment test (SJT); Tacit Knowledge Inventory for Managers (TKIM)

---

\* Corresponding author. Tel.: +1 804 827 0209.

E-mail address: [mamcdani@vcu.edu](mailto:mamcdani@vcu.edu) (M.A. McDaniel).

## 1. Introduction

In their book, *Practical Intelligence in Everyday Life*, Sternberg, Forsythe, Hedlund, Horvath, Wagner, Williams, et al. (2000) make two broad assertions regarding the nature of practical or tacit intelligence. First, they claim that there exists a general factor of practical or tacit intelligence that is substantively distinct from *g*. Second, they claim that this new construct, practical intelligence, predicts success in various domains as well as or better than *g*. In this paper, we describe research on situational judgment tests that may inform debate on both of these issues.

## 2. Sources used for informing this debate

The purpose of this paper is to inform the debate on practical intelligence by drawing on research from industrial/organizational psychology, particularly personnel selection. This critique originates from several sources. First, there is the literature on the use and validity of *g* in personnel selection (Schmidt & Hunter, 1998). A second source of literature for this paper is evidence on the validity of situational judgment tests (SJTs) (Chan & Schmitt, 2002; Clevenger, Pereira, Wiechmann, Schmitt, & Schmidt Harvey et al., 2001; McDaniel, Morgeson, Finnegan, Campion, & Braverman, 2001; McDaniel & Nguyen, 2001; McDaniel, Hartman, & Grubb, 2003; O'Connell, McDaniel, Grubb, Hartman, & Lawrence, 2002; Weekly & Jones, 1997, 1999). This research addresses the extent to which SJTs assess various constructs, and the extent to which SJTs predict job performance and provide incremental prediction over *g*.

## 3. Questions addressed in this critique

This paper seeks to answer several questions concerning the measurement of tacit knowledge and the use of SJTs: a) Is Sternberg's description of tacit knowledge tests the same as other measures? b) Is there any evidence of a general factor in situational judgment tests? c) Are SJTs correlated with *g* and, if so, are SJTs a construct measure or a measurement method? d) Is the validity of SJTs lower than the validity of *g* for predicting job performance? Each of these questions is addressed in turn.

### 3.1. *Is Sternberg's description of tacit knowledge tests the same as other measures?*

The Tacit Knowledge Inventory for Managers (TKIM) (Wagner & Sternberg, 1991) was offered as test to identify individuals whose tacit knowledge indicates the potential for successful performance in managerial or executive careers. Although not labeled as a situational judgment test, the similarities between SJTs and the TKIM are clearly shown in the examples given below. The first item is an example of a situational judgment test in which a scenario is described and the respondent must identify an appropriate response from a list of alternatives or rate the effectiveness of each alternative. The item is from an Army situational judgment test and is illustrative of SJTs (Guilford & Lacey, 1947, page 131, as cited in Northrop, 1989).

A man on a very urgent mission during battle finds he must cross a stream about 40 ft wide. A blizzard has been blowing and the stream has frozen over. However, because of the snow, he does not know

how thick the ice is. He sees two planks about 10 ft long near the point where he wishes to cross. He also knows where there is a bridge about 2 miles downstream. Under the circumstances, he should:

- A. Walk to the bridge and cross it.
- B. Run rapidly across the ice.
- C. Break a hole in the ice near the edge of the stream and see how deep the stream is.
- D. Cross with the aid of the planks, pushing one ahead of the other and walking on them.
- E. Creep slowly across the ice.

The second item is an example item from the instructions for the TKIM provided in Appendix A of *Practical Intelligence in Everyday Life*.

Your immediate supervisor has asked you for your opinion on a new promotional campaign that she has developed. You think the promotional campaign is terrible and that using it would be a big mistake. You have noticed previously that your supervisor does not take criticism well, and you suspect she is looking more for reassurance than for an honest opinion.

Given the present situation, rate the quality of each of the following reactions on this 1- to 7-point scale.

1	2	3	4	5	6	7
extremely bad			neither good nor bad			extremely good

- a. Tell her you think the campaign is great.
- b. Tell her you like the work but have some reservations about whether it is the right campaign for this client.

The actual items from the TKIM have more responses than are presented here. This example was taken from the instructions and is provided here for illustrative purposes. Clearly, there is a great deal of similarity between these two types of example items.

Such items have been in use for over 70 years. [McDaniel et al. \(2001\)](#) compared the content of the items in the TKIM to other SJTs to demonstrate that the TKIM item content is not novel, unique, or a new concept in measurement. They note that the TKIM and SJTs present situations that are hypothetical work-related scenarios that require either a judgment about the appropriateness of an action or a choice of the best action from among a range of options. [McDaniel et al. \(2001\)](#) described two specific items in the TKIM that are very similar to items on the Supervisory Practices Test ([Bruce, 1974](#)) and the Teamwork-KSA Test ([Stevens & Campion, 1999](#)). For example, item 4 of the TKIM asks the respondent to rate the effectiveness of ten strategies for handling the day-to-day work of a business manager. These strategies include thinking in terms of tasks instead of hours worked, being in charge of all phases of every task or project, and spending time planning the best way to do a task. Item 41 of the Supervisory Practices Test asks how the examinee would do daily work. Response options include taking care of the details while asking for planning from the boss, delegating details but planning with the boss, or handling details by oneself. Item 21 of the Teamwork-KSA Test asks the examinee to indicate which of four response options would most

likely help the team to do its planning and coordinating. The options include examining past practices as a guide, considering priorities, pacing, sequencing of tasks and activities, etc. As these examples illustrate, there are obvious similarities between these purportedly different measures. This is not to say that the measures are identical, but that they measure similar content using similar methodology.

The earliest judgment test is the Social Situations subtest of the George Washington Social Intelligence Test (Moss, 1926). During World War II, there were several applications of judgment tests in the U.S. Army (Guilford & Lacey, 1947; Northrop, 1989). Other examples of situational judgment tests include the Practical Intelligence Test (Cardall, 1942), the Business Intelligence Test (Bruce, 1965), the Supervisor Judgment Test (Greenberg, 1963), the Management Judgment Test used in the early 1950s and 1960s a part of the Early Identification of Management Potential program of Standard Oil (Campbell, Dunnette, Lawler, & Weick, 1970), and the U.S. Office of Personnel Management's Test 905 (Corts, 1980) of human relations skills in blue-collar supervisors. Motowidlo, Dunnette, and Carter (1990) examined the use of a situational judgment test, which they referred to as a low-fidelity simulation, for selecting entry-level managers. This measure, like other measures, presented applicants with verbal descriptions of work situations and several alternative courses of action for each situation.

In sum, Sternberg's tacit knowledge item type is not a new item type, but is simply one application of situational judgment items. We do not fault Sternberg for being unaware of the situational judgment literature in personnel psychology. Science is replete with instances of good ideas being reinvented. Fortunately for all parties interested in the advancement of intelligence theory, there is a large amount of cumulated knowledge on situational judgment tests that can help inform debate on practical intelligence theory.

### *3.2. Is there any evidence of a general factor in situational judgment tests?*

Sternberg et al. (2000, p. 223) stated that, "tacit knowledge appears to reflect a single underlying ability, which we label practical intelligence." This assertion is at odds with much situational judgment literature. Northrop (1989) prepared a detailed psychometric history of selected ability factors that included a chapter on judgment. In that review, she summarized research on situational judgment tests developed for the Army during World War II, much of it drawn from Guilford (Guilford & Lacey, 1947). She presented several items from various Army judgment tests that are easily identifiable as situational judgment items. Initial analyses of tests made from these items claimed a judgment factor but items also had large loadings on other factors including verbal, deductions, and mechanical experience. Northrop wrote:

"the communalities . . . of the various tests of Judgment were generally not high (averaging .36), indicating either factorial complexity, or considerable variance specific to each individual test. The former was probably the case, since it has been established that the concept of judgment is multidimensional. Therefore, it may be impossible to construct factorially simple items to measure judgment." (p. 193).

Northrop noted that the Army research program concerning judgment was continued in Guilford's laboratory (e.g., Guilford, Hertzka, & Christensen, 1953; Nihira, Guilford, Hoepfner & Merrifield, 1964) where it was concluded that judgment "was a cognitive attribute with many facets" (p. 199). She also noted that Judgment was excluded as a factor in the French, Ekstrom, and Price (1963) list of established factors. Northrop concluded her chapter with the observation that "Any test of judgment will necessarily be contaminated with the many other abilities with which judgment, by virtue of its pervasiveness in cognitive functioning, is associated." (p.199). More recently, Clause, Mullins, Nee, Pulakos, and Schmitt (1998)

stated that in situational judgment tests, multidimensionality may occur within individual items and, in such cases, the interpretation of factor structures is rendered hazardous.

An example of multifaceted nature of SJT items is demonstrated in the following example. Fig. 1 shows two situational items that demonstrate the factorial complexity of SJT measures at the item level. These items were developed for professional positions in a major corporation and are presented here with permission. Each item presents a scenario and several response options. The respondents were asked to rate the effectiveness of each response option for resolving the problem depicted in the scenario. Thus, each response option can be individually correlated with other variables. Cognitive ability and personality test data were available for each respondent. In scenario A, response option 1 was judged effective by those higher in  $g$  ( $r=.10$ ) and lower in agreeableness ( $r=-.13$ ). Options 2 and 5 were judged effective by those high in  $g$  ( $r=.11$  and  $.13$ ). Option 3 was judged effective by those low in agreeableness ( $r=-.16$ ). Option 4 was judged effective by those low in  $g$  ( $r=-.17$ ). Other correlations were suggestive of relationships with the effectiveness ratings but were not statistically significant. In scenario B, options 1 and 2 were found effective by those higher in conscientiousness (both  $r=.11$ ). Options 3 and 4 were found effective by those higher in  $g$  ( $r=.12$  and  $.17$ ) with option 4 also found effective by those low in agreeableness ( $r=-.10$ ). The response options for these scenarios, like most SJT scenarios and response options, are construct heterogeneous. Tests made up of such items measure multiple constructs and have loadings on multiple factors.

Although construct heterogeneity of SJT items makes coherent factor analysis results difficult, we know of four studies that report factor analyses of SJT items. Lee and Kim (2002) derived five substantive factors and a method factor for a measure of tacit knowledge in high school. However, we note that only 33% of the items in the SJT loaded on any of the factors. Legree, Heffner, Pstotka, Martin, and Medsker (2003) identified three factors in one of two automobile driving related SJT tests. They also extracted three factors from the second driving test but reported that the factor results “were less clear” (p 20) for the second test than the first. Clause et al. (1998) reported a two-factor solution, but commented on the difficulty inherent with factor analyses of heterogeneous content. Chan and Schmitt (1997) derived four factors from a situational judgment test designed for job applicant screening. That paper does not address whether the factors were interpretable. However, the first author (Personal communication from Neal Schmitt to Michael A. McDaniel, September 7, 2004) indicated that the factors were not interpretable and that has been true in several of his attempts to factor analyze situational judgment items. Note that none of these factor analyses yielded evidence of a single general factor.

Gottfredson’s (2003) review stated that no evidence is offered that there exists a general factor of practical intelligence that is uncorrelated with  $g$ . Northrop’s (1989) review of the Army situational judgment tests and Guilford’s research argued against the likelihood of a general factor. In addition, our review of the situational judgment literature would also argue against the existence of a general factor. The items tend to have construct heterogeneity and thus the tests show construct heterogeneity. Factor analyses of situational judgment tests often have difficulty obtaining any interpretable factors. No factor analysis of any situational judgment test has provided evidence of a general factor of practical intelligence, whether correlated with  $g$  or not.

### 3.3. *Are SJTs correlated with $g$ and, if so, are SJTs a construct measure or a measurement method?*

Sternberg has asserted that there is a general factor of practical intelligence that is distinct from  $g$ . Sternberg et al. (2000) stated, “Thus, there is a growing body of evidence, obtained in work, school, and

Scenario A		Correlation with		
		g	Conscientiousness	Agreeableness
	You assigned a very high profile project to one of your project managers. The project is very complex and involves the coordination of several other project managers. During each of the project update meetings, your project manager indicates that everything is going as scheduled. Now, one week before the project is due, your project manager informs you that the project is less than 50% complete.			
	Responses:			
1.	Personally take over the project and meet with the customer to determine critical requirements.	.10*	.01	-.13*
2.	Meet with the customer to extend the deadline. Talk with the project manager about how the lack of communication has jeopardized the company's relationship with the customer.	.11*	-.03	-.05
3.	Fire the project manager and take over the project yourself.	.08	.00	-.16*
4.	Coach the project manager on how to handle the project more efficiently.	-.17*	.01	.09
5.	Do not assign any high profile jobs to this project manager in the future.	.13*	.07	-.08

  

Scenario B		Correlation with		
		g	Conscientiousness	Agreeableness
	You lead a project that requires specific, accurate data to make business decisions. The data-capturing methods currently being used do not provide you with the information you need. Another department promised to provide you with the information, but failed to do so at the last minute. This setback delayed your project and you are certain that you still require the information to complete your project accurately.			
	Responses:			
1.	Do the time-consuming work yourself even though it is not technically your responsibility.	.07	.11*	-.08
2.	Temporarily allocate some members of your team to capture the data.	-.01	.11*	.00
3.	Ask the customer for a deadline extension and explain that the other department failed to provide the necessary information.	.12*	.06	-.02
4.	Ask your manager to pressure the other department to deliver the information.	.17*	.02	-.10*

Fig. 1. Two SJT items and their relationship of preferred responses to g and personality.

community settings, that suggests that tacit knowledge measures a construct distinct from academic intelligence” (p. 159). McDaniel et al. (2001) summarized data from 79 samples, including 16,994 individuals, and obtained a population correlation of .46 between g and SJTs (.36 uncorrected for measurement error).



Further, McDaniel et al. (2003) reanalyzed the McDaniel et al. (2001) data examining a response instruction moderator. They identified two categories of response to SJTs: behavioral tendency and knowledge. Behavioral tendency instructions ask examinees what they would do in a given situation, what they would most/least likely do in a given situation, or to rate/rank what they would most likely do. Conversely, knowledge tendency instructions ask examinees to select the best response, select the best/worst response, or to rate the effectiveness of various responses.

As shown in Table 1, SJTs with behavioral tendency instructions correlated .23 (.18 uncorrected for measurement error) with *g*. SJTs with knowledge instructions correlated .55 (.43 uncorrected for measurement error) with *g*. The two types of instructions also moderated the correlations of SJTs with agreeableness, conscientiousness, and emotional stability. From these results, McDaniel et al. concluded that SJTs with knowledge instructions assess primarily *g* with some personality and that SJTs with behavioral tendency instructions assess primarily personality with some *g*.

We note that Wagner and Sternberg (1991) examined correlations between a tacit knowledge measure and various personality scales in a sample of 45 managers. Few statistically significant correlates were found. However, with a sample of 45, the power to locate a population correlation of .25 (a typical observed correlation between situational judgment tests and personality scales) is only 40%. Small-sample, low-power studies do little by themselves to enhance the field's understanding of the personality correlates of practical intelligence measures. If one compares the Wagner and Sternberg sample size of 45 with those reported by McDaniel et al. (agreeableness,  $N=14,331$ ; emotional stability  $N=7718$ , conscientiousness  $N=19,656$ ; *g*,  $N=22,553$ ), one would conclude that the preponderance of the data indicates that SJTs typically measure some personality traits and *g*.

Because SJTs can and typically do measure multiple constructs, several authors (Chan & Schmitt, 2002; McDaniel et al., 2001; Weekly & Jones, 1999) have argued that situational judgment tests are measurement methods. Like other measurement methods, such as employment interviews or job knowledge tests, situational judgment tests can be built to measure a variety of constructs. To assess interpersonal constructs, one can build tests or interviews with interpersonal items or questions. Alternatively, one can build tests or interviews in which cognitive ability or conscientiousness are assessed.

In sum, these results show that SJTs are not a unidimensional construct measure, but a measurement method capable of measuring a wide variety of constructs. Empirical evidence indicates that the constructs typically assessed by situational judgment tests are *g*, conscientiousness, agreeableness, and emotional stability. No empirical evidence in this literature exists to support that a practical intelligence construct exists or is measured by situational judgment tests.

Table 1  
Meta-analytic results of correlations among situational judgment tests with *g* and the big 5 measures

	Knowledge instructions			Behavioral tendency instructions		
	Total <i>N</i>	<i>k</i>	$\rho$	Total <i>N</i>	<i>k</i>	$\rho$
<i>g</i>	11,704	40	.55	5263	21	.23
Agreeableness	8303	5	.20	5828	11	.53
Conscientiousness	13,754	8	.33	5902	11	.51
Emotional Stability	1990	4	.11	5728	10	.51

Adapted from McDaniel et al., 2003.

### 3.4. Is the validity of SJTs lower than the validity of *g* for predicting job performance?

Sternberg et al. (2000) stated, “. . .we argue that practical intelligence is a construct that is distinct from general intelligence and that general intelligence is not even general but rather applies largely, although not exclusively, to academic kinds of tasks. Moreover, practical intelligence is at least as good a predictor of future success as is the academic form of intelligence that is commonly assessed by tests of so-called general intelligence” (p. xi–xii). Gottfredson (2003) tabulated the available evidence for the correlation between practical intelligence and a variety of criteria. Most of the criteria that Sternberg uses are not job performance measures as typically defined in personnel psychology. We selected the results from Table 5 and 6 of Gottfredson (2003) that would be generally accepted as measures of job performance in personnel psychology and display them in Table 2. As is typical in the definition of job performance, we excluded criteria such as salary, level in the company, length of time in the company, and performance in training. We also excluded the criteria from studies based on faculty members.

The data for sales volume (from Table 5) are from two different years with the same sample of people. The data for Platoon Leaders, Company Commanders, and Battalion Commanders (from Table 6) report validities for two tacit knowledge tests for each sample. We recognize that others might define job performance more broadly and include more data and that other data might exist. The validity correlations in Table 2 suggest that some practical intelligence tests reported in Sternberg-authored publications predict job performance.

More plentiful data about such tests were summarized by McDaniel et al. (2003). They found that the criterion-related validity of SJTs is .32 for predicting job performance across 84 studies using 11,809 individuals. This correlation of .32 is corrected for measurement error in the criterion but does not contain any corrections for range restriction. The observed mean correlation was .24. In this discussion, we focus on the correlations for SJTs that are corrected for measurement error in the criterion as these are the best estimates of the validity of SJTs for predicting job performance at the construct (i.e., latent) level. Thus, these correlations describe the relationship between SJTs and job performance. The observed correlations describe the relationship between SJTs and job performance measured in a substantially inaccurate way (i.e., with substantial measurement error). We place the observed correlations that are downwardly biased estimates of validity in parentheses after the corrected coefficients. The validity was .27 (.21) for SJTs with behavioral tendency instructions and .33 (.25) for SJTs with knowledge instructions. The criterion-related validity of *g* tests for predicting job performance is .51 (about .25 if not corrected for range restriction and measurement error in the criterion) for jobs of

Table 2

Criterion-related validity coefficients of practical intelligence tests in the prediction of job performance in Sternberg-authored manuscripts

Performance measure	Sample size	Correlation with performance	Reference
Performance ratings	20 Bank managers	.37	Wagner and Sternberg (1985, p. 451)
	368 Platoon Leaders	.14 (TKIML),-.06 (TKIM)	Hedlund et al. (1998, pp. 22–33)
	163 Company commanders	.11 (TKIML),-.13 (TKIM)	
	42 Battalion commanders	.42 (TKIML),-.07 (TKIM)	
Sales volume	31, 39 Life insurance salespeople	.22, .15	Wagner, Sujan, Sujan, Rashotte, and Sternberg (1999, p. 166)

Adapted from Gottfredson (2003).



medium complexity (Schmidt & Hunter, 1998). We note that the validity of  $g$  tests varies with the cognitive complexity of the job (Hunter & Hunter, 1984) and situational judgment tests vary with the constructs assessed by the tests. Although the validity for  $g$  is higher than for SJTs (.51 vs. .32), the correlations are not directly comparable because the validity of  $g$  has been corrected for range restriction and the validity of SJTs has not been similarly corrected. However, if one compares the observed validities of SJTs (.21, .25) with the observed correlations for  $g$  (about .25), one concludes that SJTs and  $g$  measures have roughly comparable validities. However, the validity of SJTs for predicting job performance is NOT an endorsement of “practical intelligence” as a predictor of job performance. As shown above, SJTs are known to measure  $g$  and personality. Optimally weighted composites of  $g$  and personality typically exceed the validity of  $g$  (Schmidt & Hunter, 1998).

A key question concerning the utility of situational judgment tests concerns their incremental prediction over that explained by  $g$ . The incremental validity of situational judgment tests over measures of  $g$  has been a topic of several studies (Chan & Schmitt, 2002; Clevenger et al., 2001; O’Connell et al., 2002; Weekly & Jones, 1997, 1999). These studies showed that SJTs provide incremental validity over that of  $g$ . The incremental prediction of SJTs over  $g$  is reasonable in that SJTs typically measure job-related personality traits including conscientiousness, agreeableness, and emotional stability. Given that situational judgment tests are measurement methods and can measure a variety of constructs in varying magnitudes, the incremental validity of situational judgment tests over  $g$  can expect to vary with the  $g$ -saturation of the SJT. SJTs that are highly correlated with  $g$  can not be expected to have much incremental validity over  $g$ . SJTs that measure non-cognitive job-related constructs can be expected to have useful levels of incremental validity over  $g$ .

In summary, situational judgment tests, whether they are called practical intelligence or situational judgment tests, predict job performance. The observed validities of situational judgment tests (.21, .25) are approximately the same as the observed validities of cognitive ability tests (about .25). The validity of situational judgment tests varies with their construct loadings with the more  $g$ -loaded knowledge-instruction SJT tests having higher validity than the less  $g$ -loaded behavioral-tendency-instruction tests. However, both have validity. Thus, we do not dispute Sternberg’s claims that practical intelligence tests can predict job performance. In fact, we have summarized substantial evidence that situational judgment tests do predict job performance and can provide incremental prediction over  $g$  in the prediction of job performance.

#### 4. Conclusion

This paper has used the substantial research base of situational judgment tests in personnel psychology to help inform debate concerning practical intelligence. Based on this research, we call into question a central proposition of the theory of practical intelligence. Specifically, Sternberg et al. (2000) stated, “Thus, there is a growing body of evidence, obtained in work, school, and community settings, that suggests that tacit knowledge measures a construct distinct from academic intelligence” (p. 159). The body of evidence reviewed in this paper shows no support for a general factor of practical intelligence and no support that these measures are unrelated to  $g$ . We show that these measures are best viewed not as measures of a single construct, but as measurement methods that can and typically do assess the established constructs of  $g$ , conscientiousness, emotional stability, and agreeableness. We found no evidence for the existence of the practical intelligence construct. We encourage additional

research on situational judgment tests due to their ability to assess multiple constructs and explain unique variance in important criteria. We also encourage a revision of the theory of practical intelligence that recognizes that there is little evidence to support a construct of practical intelligence.

## References

- Bruce, M.M. (1965). *Examiner's manual: Business judgment test*. Larchmont, NY: Author.
- Bruce, M.M. (1974). *Examiner's manual: Supervisory practices test* (revised). Larchmont, NY: Author.
- Campbell, J. P., Dunnette, M. D., Lawler, E. E., & Weick, K. E. (1970). *Managerial Behavior, Performance and Effectiveness*. New York: McGraw-Hill.
- Cardall, A. J. (1942). *Preliminary Manual for the Test Of Practical Judgment*. Chicago: Science Research Associates.
- 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*, *86*, 143–159.
- Chan, D., & Schmitt, N. (2002). Situational judgment and job performance. *Human Performance*, *15*, 233–254.
- Clause, C. S., Mullins, M. E., Nee, M. T., Pulakos, E. D., & Schmitt, N. (1998). Parallel test form development: A procedure for alternative predictors and an example. *Personnel Psychology*, *51*, 193–208.
- Clevenger, J., Pereira, G. M., Wiechmann, D., Schmitt, N., & Schmidt Harvey, V. (2001). Incremental validity of situational judgment tests. *Journal of Applied Psychology*, *86*, 410–417.
- Corts, D. B. (1980). *Development and validation of a test for the ranking of applicants for promotion to first-line federal trades and labor supervisory positions*. Washington, DC: U. S. Office of Personnel Management, Personnel Research and Development Center, PRR-80-30. NTIS PB81137747.
- French, J. W., Ekstrom, R. B., & Price, L. A. (1963). *A Manual for Kit of Reference Tests for Cognitive Factors*. Princeton, NJ: Educational Testing Service.
- Gottfredson, L. S. (2003). Dissecting practical intelligence theory: Its claims and evidence. *Intelligence*, *31*, 343–397.
- Greenberg, S. H. (1963). *Supervisory Judgment Test Manual*. Technical Series No. 35. Personnel measurement research and development center. Bureau of Programs and Standards, Standards Division. Washington, DC: U.S. Civil Service Commission.
- Guilford, J. P., Hertzka, A. F., & Christensen, P. R. (1953). *A Factor-Analytic Study of Evaluative Abilities: II. Administration of Tests and Analysis of Results*. University of Southern California Psychological Laboratory Report. No. 9.
- Guilford, J. P. & Lacey, J. I. (1947). *Printed Classification Tests*. AAF Aviation Psychology Research Program, Report No. 5. Washington, D.C.: U.S. Government Printing Office.
- Hedlund, J., Horvath, J. A., Forsythe, G. B., Snook, S., Williams, W. M., Bullis, R. C., Dennis, M., & Sternberg, R. J. (1998, April). *Tacit Knowledge in Military Leadership: Evidence of Construct Validity*. Technical Report 1080. Alexandria, VA: US Army Research Institute for the Behavioral and Social Sciences.
- Hunter, J. E., & Hunter, R. F. (1984). Validity and utility of alternative predictors of job performance. *Psychological Bulletin*, *96*, 72–98.
- Lee, S., & Kim, A. (2002). A three-step approach to factor analysis on data of multiple testlets. In S. Nishitani, Y. Baba, H. Bozdogan, & K. Kanefuji (Eds.), *Measurement and Multivariate Analysis* (pp. 315–324). Springer-Verlag.
- Legree, P. J., Heffner, T. S., Psotka, J., Martin, D. E., & Medsker, G. J. (2003). Traffic crash involvement: Experiential driving knowledge and stressful contextual antecedents. *Journal of Applied Psychology*, *88*, 15–26.
- McDaniel, M. A., Hartman, N. S. & Grubb III, W. L. (April, 2003). *Situational judgment tests, knowledge, behavioral tendency, and validity: A meta-analysis*. Paper presented at the 18th Annual Conference of the Society for Industrial and Organizational Psychology. Orlando.
- McDaniel, M. A., Morgeson, F. P., Finnegan, E. B., Campion, M. A., & Braverman, E. P. (2001). Use of situational judgment tests to predict job performance: A clarification of the literature. *Journal of Applied Psychology*, *86*, 730–740.
- McDaniel, M. A., & Nguyen, N. T. (2001). Situational judgment tests: A review of practice and constructs assessed. *International Journal of Selection and Assessment*, *9*, 103–113.
- Moss, F. A. (1926). Do you know how to get along with people? Why some people get ahead in the world while others do not. *Scientific American*, *135*, 26–27.

- Motowidlo, S. J., Dunnette, M. D., & Carter, G. W. (1990). An alternative selection procedure: The low-fidelity simulation. *Journal of Applied Psychology, 75*, 640–647.
- Nihira, K., Guilford, J. P., Hoepfner, R. & Merrifield, P. R. (1964). *A factor analysis of the semantic evaluation abilities*. University of Southern California Psychological Laboratory Report #32.
- Northrop, L. C. (1989). *The Psychometric History of Selected Ability Constructs*. Washington, DC: United States Office of Personnel Management.
- O’Connell, M. S., McDaniel, M. A., Grubb III, W. L., Hartman, N. S., Lawrence, A. (April, 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.
- Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin, 124*, 262–274.
- Sternberg, R. J., Forsythe, G. B., Hedlund, J., Horvath, J. A., Wagner, R. K., Williams, W. M., et al. (2000). *Practical Intelligence in Everyday Life*. New York: Cambridge University Press.
- Stevens, M. J., & Campion, M. A. (1999). Staffing work teams: Development and validation of a selection test for teamwork settings. *Journal of Management, 25*, 207–208.
- Wagner, R. K., & Sternberg, R. J. (1991). *Tacit Knowledge Inventory for Managers: User Manual*. San Antonio, TX: The Psychological Corporation.
- Wagner, R. K., Sujana, H., Sujana, M., Rashotte, C. A., & Sternberg, R. J. (1999). Tacit knowledge in sales. In R. J. Sternberg, & J. A. Horvath (Eds.), *Tacit Knowledge in Professional Practice: Researcher and Practitioner Perspectives* (pp. 155–182). Mahwah, NJ: Earlbaum.
- 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.