## ABSTRACT

WILLIAMS, LAURA JANE. Single-Response Situational Judgment Tests: The Roles of Conscientiousness, Openness, Job Tenure, and Intelligence in SJTs with Differing Response Instructions. (Under the direction of Dr. Mark Wilson).

Historically, situational judgement tests (SJTs) have been useful tools for selection and promotion in organizations (Abilitus, 2012). This paper develops and validates a single-response SJT for use in predicting engineering managers' job performance. The current model and procedure explore relationships among 293 engineering managers' single-response SJT scores, personalities, job tenure, intelligence, and job performance, while simultaneously analyzing the differences between behavioral tendency and knowledge response instructions used in the SJT. Behavioral tendency response instructions ask respondents to indicate what they *would* do in a situation, while knowledge response instructions ask what respondents *should* do in the situation. Some researchers argue that construct validity favors SJTs that use behavioral tendency response instructions simply because of the difference in wording (Ployhart & Ehrhart, 2003), while others have found that neither type of response instruction is more valid than the other (Lievens, Sackett, & Buyse, 2009). Results from this study reveal that scores on SJTs using both types of response instructions are equally reflective of applicants' personality, while Intelligence relates slightly, yet not significantly, more strongly with knowledge response instructions. This study attempts to yield insight on the best methods for predicting engineering managers' job performance.

# Single-Response Situational Judgment Tests: The Roles of Conscientiousness, Openness, Job Tenure, and Intelligence in SJTs with Differing Response Instructions

by
Laura Jane Williams

A thesis submitted to the Graduate Faculty of
North Carolina State University
for fulfillment of the
requirements for the degree of
Master of Science

Psychology

Raleigh, North Carolina

2018

APPROVED BY:

| | |
|---|---|
| _____ | _____ |
| Dr. Mark Wilson<br>Chair of Advisory Committee | Dr. S. Bartholomew Craig |
| _____ | |
| Dr. Adam Meade | |

## BIOGRAPHY

Laura Jane Williams was raised on a farm in Wallace, North Carolina. After completing her schoolwork at Croatan High School in Newport, NC in 2009, Laura Jane entered Meredith College in Raleigh, NC. She received a Bachelor of Arts with majors in Psychology and French in May 2013. In August 2014, she entered the Industrial-Organizational Psychology graduate program at North Carolina State University. After finishing the program, Laura Jane hopes to continue her work in selection and leadership.

## ACKNOWLEDGEMENTS

I would first like to thank my thesis advisor, Dr. Mark Wilson, for his guidance over many Starbucks consultations. He allowed this paper to be my own work, but steered me in the right direction when I needed it.

I must also express my profound gratitude to my parents, Jo Ann and Franklin Williams, and my fiancé, Tyler Ferguson, for their continued advice, compassion, and support through this process. I could not have done this without you. Thank you.

# TABLE OF CONTENTS

# LIST OF TABLES

## LIST OF FIGURES

# Single-Response Situational Judgment Tests: The Role of Conscientiousness, Openness, Job Tenure, and Intelligence on SJTs with Differing Response Instructions

The use of situational judgment tests (SJTs) has become popular since their introduction by Motowidlo, Dunnette, and Carter (1990). Since that time, they have been used frequently as a method for selection and promotion in organizations. The implementation of SJTs for this purpose has increased dramatically in recent years with good reason. Not only are these tests cost effective in that they are typically computer-based and require few resources, but they are also convenient because they are user-friendly and provide quick results for recruiters (Abilitus, 2012).

Typically, situational judgment tests present the employee with challenging situations that are likely to occur in the job to which they are applying. Several possible actions are presented for each situation, and participants must consider the effectiveness of each action in order to solve the problem described. These tests can have one of two types of instructions. Behavioral tendency response instructions require the employee to indicate what they *would* do, while knowledge response instructions require the employee to indicate what one *should* do in each situation. Some researchers argue that construct validity favors SJTs that use behavioral tendency response instructions simply because of the difference in wording (Ployhart & Ehrhart, 2003), while others have found that neither type of response instruction is more valid than the other (Lievens, Sackett, & Buyse, 2009). Regardless, SJTs are a method of measuring potential decision-making skills that will likely be used after being selected into the new position. They are typically useful for "assessing soft-skills, non-academic behaviors, and practical intelligence to select the right candidate for a job" (Abilitus, 2012, p. 3), but perhaps the primary reason for recent growth in support of SJTs is their history of strong incremental validity beyond cognitive

ability, the “Big 5” personality traits, and a composite of the two (Clevenger, Pereira, Wiechmann, Schmitt, & Harvey, 2001; Whetzel, 2009).

The purpose of the current study is to create and validate a single-response situational judgment test as a method of predicting engineers’ performance when promoted to leadership positions. Contrary to typical SJT formats, single-response SJTs provide participants with a situation and action, then ask them to rate the effectiveness of that action using a Likert scale. This format has been recommended and validated by Crook et al. (2011) and Motowidlo, Crook, Kell, and Naemi (2009) and was chosen because of its validity and ease of development.

Several individual differences for technical managers (e.g. engineering managers) are depicted as having unique relationships to response instruction type and job performance. Specifically, technical and project managers possess personality traits that differ from other types of business managers. While general managers are said to perform highest when they are strong in conscientiousness and emotional stability (Barrick, Mount, & Judge, 2001), Thal (2010) suggests that, although emotional stability is important, successful technical managers are most likely to score highest on conscientiousness and openness. McDaniel, Hartman, Whetzel, and Grubb (2003) have also found that general mental ability and the amount of time spent in a job should impact the employee’s performance at work. The current model seeks to predict engineering managers’ job performance in light of these findings.

Conscientiousness, which refers to being disciplined, organized, and achievement-oriented (Komarraju, Karau, Schmeck, & Avdic, 2011), should correlate more strongly with scores on SJTs that utilize behavioral tendency response instructions than with those that use knowledge response instructions, as McDaniel et al.’s (2003) meta-analysis findings suggested. When items are framed in a knowledge format (*should*), respondents are essentially asked to

evaluate the general effectiveness of several possible responses to the incident. Likewise, when items assess behavioral tendency (*would*), the SJT is asking the participant to reveal how they would most likely behave in that situation. Results from McDaniel et al. showed that type of response instructions used in SJTs influence the constructs measured by the tests. Specifically, respondents' personalities were reflected by behavioral tendency instructions more than by knowledge instructions. Therefore, it is expected that conscientiousness will relate significantly more with behavioral tendency instructions than with knowledge instructions.

Openness is also predicted to correlate more strongly with scores on SJTs that use behavioral tendency instructions than with scores on SJTs that use knowledge instructions, coinciding with the results found by McDaniel et al. (2003). Openness is exemplified by a strong intellectual curiosity and a preference for novelty and variety (Komarraju et al., 2011). Thal (2010) reports that effective technical managers differ from general managers in that they tend to exhibit higher levels of openness than general managers do. Therefore, openness is expected to be more strongly related to behavioral tendency instructions than knowledge instructions.

Effective engineering managers should also exhibit high levels of intelligence and knowledge of their job requirements and expectations. McDaniel et al. (2003) discovered through meta-analysis that cognitive ability was more correlated with SJTs that used knowledge instructions. Because general mental ability seems to be more related to knowledge instructions than behavioral tendency instructions, this study hypothesizes that engineering managers' intelligence will be more strongly correlated with single-response SJT scores when the test uses knowledge response instructions rather than behavioral tendency response instructions.

In addition to intelligence, job tenure is also said to have a positive relationship with knowledge response instructions on SJTs. Research by McDaniel et al. (2003) hypothesized that

a longer amount of time in a position should contribute to the job knowledge that the employee acquires. This study explores the idea that job tenure, or number of years in a position, will have a stronger relationship with engineering managers' scores on SJTs using knowledge instructions than on those using behavioral tendency instructions.

As discussed, conscientiousness, openness, intelligence, and job tenure have been used in past research to validate SJTs for general managers. However, there is no evidence that these trends are still observed when using a single-response SJT with technical managers. This study attempts to explore SJT-personality and SJT-knowledge relationships specifically within the realm of engineering managers and using this particular form of assessment.

The final component of the current study includes Job Performance. At this time, very limited research exists which explains the relationships that differently phrased SJT response instructions have on overall job performance. One study by McDaniel, Schmidt, and Hunter (1998) suggests that job experience is positively related to job performance. However, while this is evidence that tenure may be related to performance, it reveals nothing with regard to type of response instructions utilized. Thus, Research Question 1 of this study asks whether scores on single-response SJTs that use knowledge response instructions or behavioral tendency response instructions will have greater correlations with engineering managers' self-rating of overall job performance.

In summary, the purpose of this study is to examine the following hypotheses and research questions (see Figure 1): Hypothesis 1 states that Costa and McCrae's (1992) Revised NEO Personality Inventory (NEO-PI-R) measurement of Conscientiousness will correlate with scores on the newly-developed single-response SJT that uses behavioral tendency response instructions significantly more positively than it correlates with scores on the single-response

SJT that uses knowledge response instructions. Hypothesis 2 states that the NEO-PI-R measurement of Openness will correlate with scores on the newly-developed single-response SJT that uses behavioral tendency response instructions significantly more positively than it correlates with scores on the single-response SJT that uses knowledge response instructions. Hypothesis 3 states that intelligence, measured by high school SAT score, will correlate with scores on the single-response SJT that uses knowledge response instructions significantly more positively than it correlates with scores on the single-response SJT that uses behavioral tendency response instructions. Hypothesis 4 states that number of years in an individual's current position will correlate with scores on the single-response SJT that uses knowledge response instructions significantly more positively than it correlates with scores on the single-response SJT that uses behavioral tendency response instructions. Finally, Research Question 1 asks if technical managers' self-report of job performance is more likely to be significantly correlated with scores on the newly-developed single-response SJT that uses knowledge response instructions or the single-response SJT that uses behavioral tendency response instructions.

## Method

In order to test each hypothesis and research question in this study, a single-response situational judgment test was both developed and validated. Therefore, the Participants and SJT sections of this method include both a development component and a validation component.

### Research Design

Because the hypotheses in this study inferred no causal relationships, a correlational design was implemented for this study. For this design, participants were simply asked to complete the questionnaire administered to them. With these data, correlations between several relevant variables were computed and analyzed.

**Participants**

The primary component of situational judgment tests is a collection of critical incidences of good and bad job performance (McDaniel & Nguyen, 2001). For the segments involving the development and rating of critical incidents in this study, a total of 11 Professional Engineers were used as subject matter experts (SMEs). Of these eleven, six participants affiliated with North Carolina State University developed the critical incidents. All six participants held differing engineering degrees in Civil and Environmental Engineering, Industrial Engineering, Materials Science, and Mathematics, and all were over 45 years old. Three indicated their race as White, one as Asian, and two chose not to respond. Additionally, three were male, one was female, and two chose not to respond.

A second group of five subject matter experts from Amazon's Mechanical Turk rated the developed critical incidents on effectiveness. In this sample, Civil and Environmental Engineering, Electrical Engineering, Computer Science, and Mechanical Engineering were represented. Of these five, two indicated they were under 30 years old, two were between 30 and 34, and one was between 35 and 39 years old. Three participants indicated their race as white, while two indicated that they were American Indian. The sample consisted of three females and two males, and four of the five were from the United States, while the fifth's origin was India.

All of the subject matter experts in this study were English-speaking managers in the United States who had earned Bachelor's degrees or higher in Engineering and Professional Engineering (PE) Licenses. All participants had also held engineering management positions for at least three years. To ensure the technical credibility of the MTurk participants, they were asked to answer a preliminary engineering-related qualifying item. Specifically, the qualifying item presented participants with a fulcrum held level by three different forces. On one side of the

fulcrum were two forces at differing distances from the center, their weights indicated in kips. On the other side, one force with an undisclosed weight was presented. The participants were required to correctly calculate the mystery force before continuing with the assessment. Participants in the development and effectiveness rating groups were considered SMEs with regard to engineering management. Both the development and the rating segments took SMEs approximately 15 minutes to complete.

The participants for the model testing phase of this study were collected through Amazon's MTurk and included 293 (N=293) English-speaking engineering managers with B.S. degrees in Engineering. All participants were fairly evenly spread among 10 engineering fields. All age ranges from "Under 30" to "Over 50" were represented, but 41% of the participants were between 30 and 34 years old. Race was also spread among Hispanic, American Indian, Asian, Black, and White, but a majority of 203 people indicated their race as White. Additionally, 201 of the participants were male while 92 were female.

Of the 293 total participants in this phase, 145 people were randomly assigned to the knowledge instruction group, and 148 people were assigned to the behavioral tendency instruction group. These participants must have been employed in management roles for at least one year. They were asked to indicate the number of years they have maintained their current positions. Additionally, to ensure their technical credibility, the participants were asked to answer the preliminary qualifying item before starting the questionnaire. This phase of the study took approximately 20 minutes to complete; each participant was compensated accordingly for his or her responses.

**Measures**

**SJT.** The single-response format of the critical incidence portion of this SJT was recommended and validated by Motowidlo et al. (2009) and Crook et al. (2011). This style of SJT was favored for this study because of its known validity, time-saving nature, and ease of development and administration for SMEs and participants alike. To create this format, 6 Professional Engineers at North Carolina State University first generated critical incidents using a form inspired by the LAS Manager's Critical Incident Form (Appendix A). As suggested by Motowidlo et al. (2009), SMEs first provided examples of occasions when they saw an engineering manager do something that struck them as especially effective or ineffective. Each SME gave one example of effective consideration (person-oriented) behavior, one example of ineffective consideration behavior, one example of effective structure initiation (task-oriented) behavior, and one example of ineffective structure initiation behavior as defined by Fleishman (1953), who described these two types of behavior as the two main factors of successful leadership. To help with this distinction, participants were first given a brief description of ideal manager characteristics:

> *"A good engineering manager is believed to be strong in both task-oriented and person-oriented behaviors. Task-oriented managers create clear, organized, easy to follow work schedules with specific requirements and deadlines. Person-oriented managers cultivate positive working relationships and help employees feel energized about the difference they make in the company."*

Participants were instructed to describe the context leading to the behavior, the behavior itself, and any consequences that occurred as a result of the behavior. These responses were then edited to correct grammar mistakes and omit incidents that were not engineering-related. A total

of 16 critical incidents (7 consideration items and 9 structure items) were retained. Next, a second group of 5 SMEs from MTurk rated the effectiveness of each edited incident on a Likert 7-point scale. Means and standard deviations of effectiveness ratings for each incident were computed across raters. A high degree of interrater reliability was found across the critical incidents. The average measured ICC was .883 with a 95% confidence interval from .761 to .953 ($F(15,60)= 9.258$, $p < .001$). This indicated that the participants were generally in agreement on the effectiveness of the incidents' actions. The 16 edited incidents then served as the items in the SJT, and mean SME effectiveness ratings served as a scoring key for the items.

Next, Amazon's Mechanical Turk was used to collect responses for the 16 critical incidents. To prove their technical credibility, all MTurk participants were first required to answer the qualifying engineering-related item. If the participant responded incorrectly to this item, they were immediately disqualified from continuing with the task. Participants who responded correctly were directed to a demographic segment asking them to indicate their specific engineering degree obtained, age, race, gender, and country of origin.

Participants (N=293) were then randomly assigned to either a behavioral tendency (*would*) response instruction group or a knowledge (*should*) response instruction group. All participants were presented with the same incidents, including context, behavior, and consequences; however, instructions asked the participants to indicate either, "How well does this action describe what an Engineering Manager should do in this situation?" or "How well does this action reflect what YOU would do in this situation?" The only difference between the two tests was the response instruction presented for each critical incident. Each incident asked the participant to rate the effectiveness of the presented behavior on a Likert 7-point scale. As suggested by Motowidlo et al. (2009), scoring for this portion of the SJT measured the similarity

between the participant's judgments and SMEs' judgments in the work situations. This score was formulated by computing the difference between the applicant's effectiveness rating and the mean SME rating for that item, squaring the difference, and finding the mean squared difference across all of the SJT items. Finally, individuals' mean squared scores were multiplied by -1 so that higher scores indicated increasing similarity to SMEs' judgments, thus indicating increasing knowledge about effectiveness of the behaviors presented.

**Openness and Conscientiousness.** Items measuring Openness and Conscientiousness were obtained from Costa and McCrae's (1992) Revised NEO Personality Inventory. These items are scored using a 5-point Likert rating scale (1=Strongly Disagree, 5=Strongly Agree). Only Openness ($\alpha$=.87) and Conscientiousness ($\alpha$=.90) items were obtained from the NEO-PI-R inventory. A total of 48 Openness and 48 Conscientiousness items were included. Three attention check items (e.g. "Choose Strongly Agree for this item.") were also added in this segment of the questionnaire to ensure participants were sufficiently attending to the task. After responses were collected for this segment, several items for both constructs were reverse-scored according to the NEO-PI-R scoring manual.

**Intelligence.** General mental ability of participants was measured by a single item asking participants to indicate their SAT scores to the best of their recollection. Participants were given a 5-point scale with each point representing a range of SAT percentiles (e.g. $30^{th}$ to $49^{th}$ percentile) and asked to state the percentile earned. Of the 293 total participants in this phase, 17 responses were missing, presumably because they could not remember their SAT scores. The Intelligence variable was omitted from these participants' responses.

**Tenure.** To measure job tenure, participants were asked to indicate the number of years they have maintained their current position. To encourage precision and induce variability of

responses, participants were allowed to indicate tenure using decimals (e.g., 4.5 years). This number was treated as a count variable (McDaniel et al., 2003).

**Job Performance.** Previous research in job performance has found that employees' past performance is a valid indicator of their future performance (Guion, 2011). Wanous and Hudy (2001) and Wanous, Reichers, and Hudy (1997) have also found sufficient reliability in single-item measures of job satisfaction and effectiveness at work. Further, Mael (1991) advocates that second-hand interpretation of events is a common method of measuring performance through biodata, and that social desirability distortion is not a serious concern in biodata measures. Therefore, participants' job performance was measured in this study by a single item asking them to honestly, and as accurately as possible, indicate how their current supervisor would rate their performance compared to other employees. This self-rating was reported on a Likert-type 10-point scale and the item read, "According to the distribution given, how would your current supervisor rate your overall performance at work compared to others? Please respond as honestly and accurately as possible."

## Procedure

In order to validate the SJT and test the current hypotheses and research question, two different versions of the SJT with differing response instructions were administered with all participants randomly and equally assigned to one of the two versions. For each incident, participants were asked to rate the effectiveness of the presented behavior using a 7-point Likert scale. In both versions of the SJT, these critical incidents were presented alongside the NEO-PI-R questionnaire items with 3 attention checks, the single SAT score item, the single job tenure item, and the single job performance item. A total of 322 participants completed the questionnaire in its entirety. However, because of the length of the personality segment of the

questionnaire, some participants did not pay attention when responding to those items. Participants who responded incorrectly to two or more attention checks, as well as participants who chose the same response for every personality item, were omitted from analysis. With the results of the 293 remaining participants, a series of correlations and Fisher's Z tests were run among the collection of variables. It was expected that SJT scores yielded from tests with knowledge instructions would correlate more highly with intelligence and tenure, and that scores yielded from tests with behavioral tendency instructions would correlate more highly with personality variables.

To examine the hypotheses, a series of two-tailed Pearson correlations were run (see Table 1). To test Hypothesis 1, a correlation coefficient was run between SJT score and Conscientiousness for each group. To test Hypothesis 2, a correlation was run between SJT score and Openness for each group. To test Hypothesis 3, a correlation coefficient was found between SJT score and job tenure for each group. Next, Hypothesis 4 was tested using a correlation between SJT score and SAT score in each group. Finally, the coefficients from each of these four pairs of correlations were tested for significant differences between groups using Fisher's Z test (see Table 2). It was expected that SJT scores obtained using knowledge response instructions in the SJT would correlate higher and more significantly ($p < .05$) with general mental ability and tenure, whereas scores using behavioral tendency response instructions would correlate higher and more significantly ($p < .05$) with the Openness and Conscientiousness items. To examine Research Question 1, correlations were run between SJT score and job performance rating for each group. These coefficients were also compared using the Fisher's Z test method.

## Results

A significant correlation between Conscientiousness and SJT Score was found for both behavioral tendency instructions (r = .342, $p < .001$) and knowledge instructions (r = .340, $p < .001$); however, Fisher's Z test indicated that the difference between the correlations was not significant (z = .02, $p = .492$). Therefore, Hypothesis 1 was not supported. Additionally, there was a significant correlation between Openness and SJT Score for both behavioral tendency (r = .192, $p < .05$) and knowledge instructions (r = .287, $p < .001$). Again, Fisher's Z test indicated that there was not a significant difference between the two correlations (z = .85, $p = .198$). Hypothesis 2 was not supported. Third, there was no significant relationship between Intelligence and SJT Score using behavioral tendency instructions (r = .117, $p = .168$), but there was a significant positive correlation for knowledge instructions (r = .206, $p < .05$). This follows the theoretical idea of Hypothesis 3; however, the difference between the two correlations was not significantly different according to Fisher's Z test (z = .77, $p = .221$), so Hypothesis 3 was also not supported. Finally, the relationship between Tenure and SJT Score for both behavioral tendency and knowledge instructions were both not significant (r = .093, $p = .262$; r = -.016, $p = .846$). Fisher's Z test revealed that there was also no significant difference between these correlations (z = .93, $p = .176$); therefore, Hypothesis 4 was not supported.

Research Question 1 explored the relationship between SJT Score and Job Performance, again distinguishing between behavioral tendency (r = -.026, $p = .755$) and knowledge (r = -.082, $p = .328$) response instructions. Results indicated that the relationships were not significant and that there was no significant difference between the two correlations (z = .48, $p = .316$).

Additionally, in order to determine whether the SJT contributed to the overall fit of the conceptual model in this study, a stepwise multiple linear regression was run on the sample. This

stepwise MLR compared model fit before and after the SJT was included (see Table 3). Although Openness was excluded because it did not provide a significant improvement, all other variables were included in the regression. Model 1 included Tenure only ($R^2$ = .038), Model 2 included Tenure and Conscientiousness ($R^2$ = .062), Model 3 included Tenure, Conscientiousness, and Intelligence ($R^2$ = .081), and Model 4 included Tenure, Conscientiousness, Intelligence, and SJT Score ($R^2$ = .099). Results from this multiple linear regression indicated that the SJT provided a significant improvement in fit over models that do not include the SJT.

## Discussion

As a reflection of the findings of McDaniel et al. (2003), the current study inferred that personality variables would link significantly more strongly with behavioral tendency SJT response instructions, while intelligence and job tenure would have a significantly stronger relationship with knowledge response instructions. While none of this study's hypotheses were fully supported, there are several interesting findings worth discussing here.

First, results indicate that, regardless of response instruction type, scores on this SJT were significantly correlated with participants' Conscientiousness and Openness. Because Thal (2010) posits that successful engineering managers are high in these two personality traits, we may infer that a high correlation between personality and SJT score would suggest that the SJT will be a valid predictor of engineering managers' success. However, because there is no significant difference between response instructions, we cannot confirm that one type of response instruction is superior over the other.

Second, results of this study show that intelligence is significantly related to scores on SJTs using knowledge instructions and is not significantly correlated with scores on SJTs using

behavioral tendency instructions. This finding potentially gives knowledge response instructions an advantage when researchers are tasked with deciding which type of instructions to use. However, since the difference between correlations is not significant, we cannot firmly determine that there is a true difference between response instruction types.

Third, participants' tenure was found to have no relationship with SJT score, regardless of instruction type. This means that engineering managers' performance on SJTs is not influenced by the amount of time they have spent in their jobs, contrary to the results found by McDaniel et al. (2003). The current study differs from the McDaniel et al. study in that it applies specifically to engineering managers; the finding that tenure is inconsequential could mean that this construct is less important in engineering managers' success than it is in general business managers' success.

Finally, neither SJTs with behavioral tendency instructions nor knowledge instructions were significantly correlated with job performance. Although only a research question, not a hypothesis, was associated with job performance, this result is certainly not what we would expect. An explanation of the results is described below.

**Limitations**

The most influential limitation of this study may be the source from which the sample was taken. MTurk was chosen because of the ease, volume, and speed at which data can be collected using the site. However, there was no way to ensure that all participants were verified current, English-speaking engineering managers. The engineering-related qualifier item attempted to distinguish engineers from non-engineers, but it was impossible to be entirely sure of its effectiveness since some participants could have either guessed correctly or calculated the correct response using basic knowledge of torque, irrespective of occupation. Additionally, since

this was a low-stakes situation that was also paid, MTurk workers may have been more apt to lie or respond carelessly in order to get through the questionnaire and get paid more quickly.

Perhaps the issue behind the non-significant relationship between SJT score and job performance was the job performance measure itself. Job performance was measured by a single item asking participants to rate their own performance. With an MTurk sample, it was impossible to collect data from supervisors or colleagues. Self-ratings of performance are generally noted to be more lenient in comparison to supervisory performance appraisals (Heidemeier & Moser, 2008); thus, many participants in this study likely rated their performance higher than their true performance. Further, range restriction may have been an issue; on a scale of 1 to 10, the mean job performance rating was 7.9 with a standard deviation of 1.342. The job performance measure used here is likely not valid enough to accept the non-significant findings of this study at face value.

**Future Research**

Although the hypotheses of this study were largely not supported, there are several details that could be adjusted to improve the procedure. Future replications of this study should be performed on a sample of current engineering management incumbents in an assortment of American engineering firms. The same personality, SJT, and tenure measures should be completed by the incumbents, and the job performance measure should be completed by the incumbents' supervisors. This would ensure that the sample is truly comprised of engineers, and it would validate the job performance measure better than self-reports. This situational judgment test has the potential to affect selection practices in engineering, were it to be administered to a more representative sample in the future.

Additionally, multiple linear regression results indicated that while model fit improved significantly for assessments including the SJT, the beta coefficient for SJT score was -.156. This means that for every unit increase in SJT score, participants' job performance decreased by .156. Although this trend is not what we would expect, a more accurate measure of job performance may alter this result. Future iterations of this study with accurate performance data should involve stepwise multiple linear regressions to determine if the SJT does indeed provide better goodness of fit over assessments that do not include the SJT.

**Conclusion**

Results of this study attempted to yield information helpful in determining which types of response instructions will be most effective in future assessments of this nature. Both behavioral tendency and knowledge response instructions seem to reflect personality well, while knowledge instructions were able to significantly reflect participants' intelligence. However, the SJT developed in this study failed to correlate with participants' self-reported job performance. Therefore, we are still unsure if using this type of assessment in selection processes will predict engineering managers' success. Future replications of this study should be performed with engineering incumbents and their supervisors in order to ensure the most accurate results possible.

## REFERENCES

Abilitus (2012, March 20). Situational judgment tests: What they look like. Retrieved from http://www.abilitus.com.

Barrick, M. R., & Mount, M. K. (1991). The Big Five personality dimensions and job performance: A meta-analysis. *Personnel Psychology, 44*, 1-26.

Clevenger, J., Pereira, G.M., Wiechmann, D., Schmitt, N., & Harvey, V.S. (2001). Incremental Validity of Situational Judgment Tests. *Journal of Applied Psychology, 86*, 410-417.

Costa, P. T. Jr., & McCrae, R. R. (1992). Revised NEO Personality Inventory (NEO-PI-R) and NEO Five-Factor Inventory (NEO-FFI) professional manual. Odessa, FL: Psychological Assessment Resources.

Crook, A. E., Beier, M. E., Cox, C. B., Kell, H. J., Hanks, A. R., & Motowidlo, S.J. (2011). Measuring relationships between personality, knowledge, and performance using single-response situational judgment tests. *International Journal of Selection and Assessment, 19*, 363-373.

Fleishman, E. A. (1953). The description of supervisory behavior. *Journal of Applied Psychology, 37*, 1-6.

Guion, R. M. (2011). *Assessment, Measurement, and Prediction for Personnel Decisions*. New York, NY: Taylor & Francis Group.

Heidemeier, H. & Moser, K. (2008). Self-other agreement in job performance ratings: A meta-analytic test of a process model. *LASER Discussion Papers, 17*, 1-54.

Judge, T. A., Piccolo, R. F., & Ilies, R. (2004). The forgotten ones? The validity of consideration and initiating structure in leadership research. *Journal of Applied Psychology, 89*, 36-51.

Komarraju, M., Karau, S. J., Schmeck, R. R., & Avdic, A. (2011). The Big Five personality traits, learning styles, and academic achievement. *Personality and Individual Differences, 51*, 472-477.

Lievens, F., Sackett, P.R., & Buyse, T. (2009). The effects of response instructions on situational judgment test performance and validity in a high-stakes context. *Journal of Applied Psychology, 94*, 1095-1101.

Mael, F. A. (1991). A conceptual rationale for the domain and attributes of biodata items. *Personnel Psychology, 44*, 763-792.

McDaniel, M. A., Hartman, N. S., Whetzel, D. L., & Grubb, W. L. (2003). Situational judgment tests, knowledge, behavioral tendency, and validity: A meta-analysis. *Personnel Psychology, 60*, 63-91.

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

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.

Motowidlo, S. J., Crook, A. E., Kell, H. J., & Naemi, B. (2009). Measuring procedural knowledge more simply with a single-response situational judgment test. *Journal of Business Psychology, 24*, 281-288.

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.

Ployhart, R.E., & Ehrhart, M.G. (2003). 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, 11*, 1-16.

Thal, A. E., & Bedingfield, J. D. (2010). Successful project managers: An exploratory study into the impact of personality. *Technology Analysis & Strategic Management, 22*, 243-259.

Wanous, J. P. & Hudy, M. J. (2001). Single-item reliability: A replication and extension. *Organizational Research Methods, 4*, 361-375.

Wanous, J. P., Reichers, A. E., and Hudy, M. J. (1997). Overall job satisfaction: How good are single-item measures? *Journal of Applied Psychology, 82*, 247-252.

Whetzel, D. L. & McDaniel, M. A. (2009). Situational judgment tests: An overview of current research. *Human Resource Management Review, 19*, 188-202.

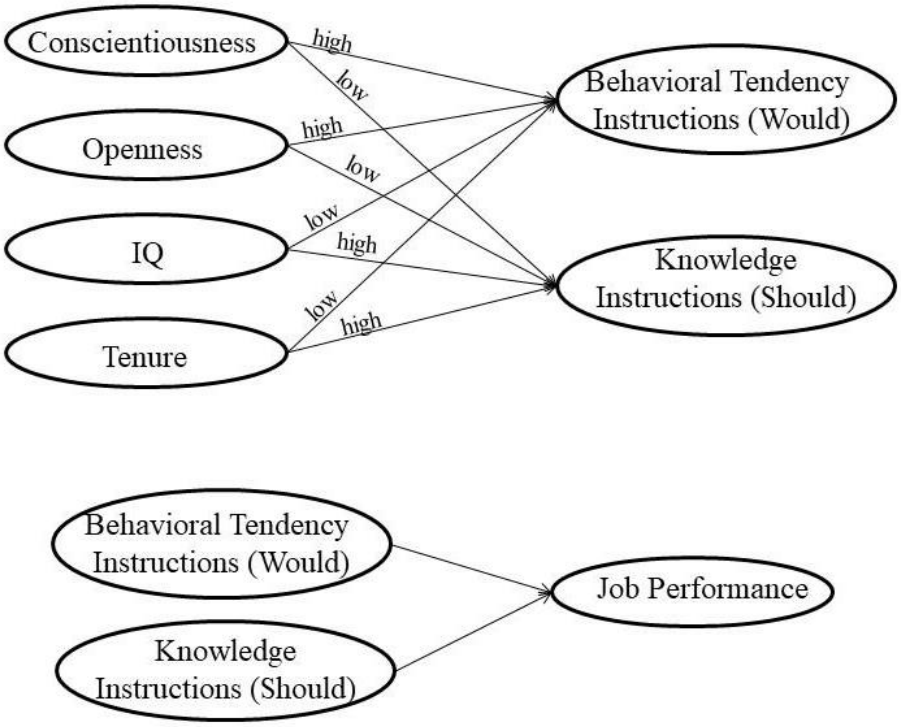


*Figure 1*. Conceptual Model of Knowledge and Behavioral Variables and Their Relationships with Response Instructions and Job Performance.

Table 1

*Pearson Correlations for Knowledge and Behavioral Tendency Response Instructions*

| | SJT Score | Job Performance | Intelligence | Tenure | Openness | Conscientiousness |
|---|---|---|---|---|---|---|
| Knowledge Response Instructions | | | | | | |
| SJT Score | 1 | | | | | |
| Job Performance | -.082 | 1 | | | | |
| Intelligence | .206* | .109 | 1 | | | |
| Tenure | -.016 | .213** | .107 | 1 | | |
| Openness | .287*** | .018 | .259** | -.079 | 1 | |
| Conscientiousness | .340*** | .251** | .039 | .147 | .371* | 1 |
| Behavioral Tendency Response Instructions | | | | | | |
| SJT Score | 1 | | | | | |
| Job Performance | -.026 | 1 | | | | |
| Intelligence | .117 | .213* | 1 | | | |
| Tenure | .093 | .231** | -.008 | 1 | | |
| Openness | .192* | .016 | .018 | -.067 | 1 | |
| Conscientiousness | .342*** | .158 | .029 | .135 | .548*** | 1 |

Note: *** is significant at the .001 level ($p < .001$). ** is significant at the .01 level ($p < .01$). * is significant at the .05 level ($p < .05$).

Table 2

*Fisher's Z Test for Correlation Comparisons*

| | Z Test | *p* |
|---|---|---|
| Intelligence -> SJT Score | 0.77 | 0.221 |
| Tenure -> SJT Score | 0.93 | 0.176 |
| Conscientiousness -> SJT Score | 0.02 | 0.492 |
| Openness -> SJT Score | 0.85 | 0.198 |
| SJT Score -> Job Performance | 0.48 | 0.316 |

Note: *** is significant at the .001 level ($p < .001$). ** is significant at the .01 level ($p < .01$). * is significant at the .05 level ($p < .05$).

Table 3

*Stepwise Multiple Linear Regression Coefficients Predicting Job Performance, Includes both Knowledge and Behavioral Tendency Response Instructions*

| | Model 1 | | | Model 2 | | | Model 3 | | | Model 4 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Variable | *B* | *SE* | *β* | *B* | *SE* | *β* | *B* | *SE* | *β* | *B* | *SE* | *β* |
| Tenure | .076 | .022 | .185 | .069 | .022 | .185 | .067 | .022 | .178 | .065 | .022 | .174 |
| Conscientiousness | | | | .371 | .131 | .166 | .362 | .130 | .162 | .474 | .136 | .212 |
| Intelligence | | | | | | | .193 | .075 | .149 | .223 | .075 | .173 |
| SJT Score | | | | | | | | | | -.112 | .044 | -.156 |
| $R^2$ | | .038 | | | .062 | | | .081 | | | .099 | |
| *F* for change in $R^2$ | | 11.762** | | | 7.998** | | | 6.610* | | | 6.456* | |

Note: *** is significant at the .001 level. ($p$<.001). ** is significant at the .01 level ($p$<.01). * is significant at the .05 level ($p$<.05).

# APPENDIX

# Appendix A


## Critical Incident Form - Manager

*Return form to XXXXX no later than 17:00 on Friday, May 2, 2015.*

**Name:** **Date:**

**Instructions:** Read the incident listed below. Think back to a time where you witnessed a fellow manager perform at an exceptionally high level. We are interested in the first incident that comes to mind. Please describe the context which led up to the event, a description of the behavior, as well as the consequences following the behavior.

Please note that no classified material should be described and that all responses will go through the CAO process.

**Sample Positive Critical Incident.**

**Context:** Approximately six years ago the U. S. Department of Defense issued an immediate ban of all USB drives and other removable media devices following a worm infiltrating Army networks. Multiple networks were going down and being abused at an alarming rate. At the time, Carter was in charge of managing a large team of researchers and was tasked with resolving the network issue immediately. (*Notice how the larger environmental context was described*)

**Behavior:** A mathematician by training, Carter spent much of his time not only with his own team of researchers but with operations personnel as well. Having credibility on both the developer and mission side gave him a valuable perspective on the resources of his colleagues in multiple offices across agencies. Knowing what had to happen in order to ameliorate the situation, Carter began to identify agency members who were best suited to help resolve the issue. He effectively understood the requirement at hand and translated that requirement appropriately to the right people. Carter assembled a team with maximum likelihood of solving the network crisis. (*Notice the level of specificity used when describing the actual behavior)*

**Consequences:** Carter's management skills and the work of his team resulted in the identification of the worm in considerably less time than was expected. Soon thereafter, Army networks were secured and new protocols were put in place to protect against any future attacks. (*Notice that both proximal and distal outcomes were described*)

**1. Positive Critical Incident:**

**Context:**

**Behavior:**

**Consequences:**

Laboratory for Analytic Sciences
Reflect. Observe. Imagine.

# Critical Incident Form - Manager

**Instructions:** Second, think back to a time where you witnessed a fellow manager perform a particularly *negative* behavior. Again, please describe the context which led up to the event, a description of the behavior, as well as the consequences following the behavior. Refer to the example for guidance.

**Sample Negative Critical Incident.**

**Context:** Sidney is an agency employee who supervises several researchers across multiple projects. He isn't known for listening to others or networking amongst his peers. Sidney is especially interested in one project he began prior to being promoted to a management role that aggregates text data and presents it in a visually interesting way. (*Notice how reputation is included in the contextual description*).

**Behavior:** Fellow technical directors have expressed their skepticism both in private and to Sidney himself regarding the viability of Sidney's pet project. Nevertheless, Sidney has continually committed a significant portion of his budget towards the data visualizer. (*Notice how both observable behavior as well as reactions to said behaviors are included*)

**Consequences:** Recently, after years of sunk resources, both researcher time and project money, Sidney's pet project was deemed unfit to be continued and was defunded.

**2. Negative Critical Incident:**

**Context:**

**Behavior:**

**Consequences:**