

The Effects of Social Desirability on Situational Judgment Tests in Organizational Selection

Sarah Linkletter

Thesis submitted to the
Faculty of Graduate and Postdoctoral Studies
In partial fulfillment of the requirements
For the Master of Arts in Education degree
Faculty of Education
University of Ottawa

© Sarah Linkletter, Ottawa, Canada, 2017

Acknowledgments

I would like to extend my sincerest gratitude to those who have made the writing of this dissertation possible; I am sincerely humbled by the encouragement and assistance that has been provided to me along this journey.

I would first like to thank Dr. David Trumpower; without his mentorship and instruction, this research would not be possible. I would also like to thank my committee members, Peter Milley and Katherine Moreau, for their thoughtful feedback and insight.

In addition, I would like to thank the team at Evaluation Personnel Selection Inc. (EPSI) for supplying the data included in this research and their collaboration. Specifically, I would like to thank Andre Durivage, Nicolas Roy and Sabrina Poirier for their knowledge and support. This organizations generosity has exceeded my expectations and I could not be more honored.

Lastly, I would like to thank my friends, fiancé, and family who have provided unconditional love throughout this journey.

Abstract

Organizational contexts use Situational judgment tests (SJTs) to assess and select individuals for competitive positions. As with other standardized assessments, threats to validity must be identified, examined, and communicated. As such, this research aims to identify the effects of socially desirable responding on the validity of an SJT used in a competitive selection process, and to identify if response latency provides insight into the identification of socially desirable responding. Participants in a competitive organizational selection process were administered an online assessment and the Balanced Inventory of Desirable Responding (BIDR). No significant correlations were identified between social desirability and the online assessment; however, this research provides evidence to suggest that participants who had shorter response latencies were less likely to participate in social desirability responding.

Table of Contents	Page
Acknowledgements	ii
Abstract	iii
Table of Contents	iv
List of Tables	vi
List of Figures	vii
List of Acronyms and Abbreviations	viii
Introduction	1
Research problem	2
Literature Review	3
Overview of Situational Judgment Tests	3
Strengths of Situational Judgement Tests	8
Weaknesses of Situational Judgment Tests	11
Limitations of Situational Judgment Tests	12
Research question	14
Theoretical framework	14
Method	15
Participants	15
Measures	16
Secondary Data Analysis	18
Analysis and Results	19
Discussion of the Results	27
Summary	32
Contributions	33
Limitations	34
Conclusion	34
References	35
Appendix A: Balanced Inventory of Desirable Responding (BIDR), Seventh Edition.	46
B: Management and Leadership Skills Test (MLST) Test Development	48
C: Competencies and Behavioral Indicators	51
D: MLST Confidentiality, Instructions and Sample Situation	54

E: Gender and Age Results.	59
Statement of Contributions	60

List of Tables	Page
Table 1. Mean and standard deviation for the MLST competencies	19
Table 2. BIDR Categorizations	19
Table 3. Frequencies and percentages of BIDR composite and sub-scale	20
Table 4. MLST Categorizations	20
Table 5. Frequency of participants BIDR and MSLT composite results	20
Table 6. Response Latency Categorizations	20
Table 7. Frequency of participants BIDR composite results and response latencies	21
Table 8. Frequency of participants Self-deception Enhancement (SDE) sub-scale results	21
Table 9. Frequency of participants Self-deception Enhancement (SDE) sub-scale results and response latencies	21
Table 10. Frequency of participants Impression Management (IM) sub-scale results	21
Table 11. Frequency of participants Impression Management (IM) sub-scale results and response latencies	21
Table 12. Frequency of participants MLST composite scores and response latencies	22
Table 13. Frequency of participant's response latencies, BIDR composite scores and MLST composite scores	22

List of Figures	Page
Figure 1. Theoretical Framework	15

List of Acronyms and Abbreviations

Balanced Inventory of Desirable Responding (BIDR): is a social desirability scale that consists of forty (40) propositional items.

Canadian Food and Inspection Agency (CFIA): is a science-based regulator that is dedicated to safeguarding food, animals and plants, in order to enhance the health and well-being of Canada's people, environment and economy.

Change Leadership (CL): A competency used in the Management and Leadership Skills Test (MLST), for an overview of the behaviour indicators please see Appendix C.

Decisiveness (D): A competency used in the Management and Leadership Skills Test (MLST), for an overview of the behaviour indicators please see Appendix C.

Evaluation Personnel Selection International (EPSI): a private organization that provided the data used in this research.

Executive Director (EX): An occupational group in the Government of Canada.

Grade point average (GPA): is a 4.0 standardized scale used by colleagues and high schools to judging a student's performance.

Impression management (IM): the behavior demonstrated when individuals consciously and publicly make exaggerated claims of desirable behavior.

Management and Leadership Skills Test (MLST): is a competency based assessment designed to evaluate candidates in a competitive personnel selection context for the role of director.

People Development and Management (PDM): A competency used in the Management and Leadership Skills Test (MLST), for an overview of the behaviour indicators please see Appendix C.

SAT: is a standardized test widely used for college admissions in the United States.

Self-deception enhancement (SDE): statements that are easily rationalized without overt awareness and associated with positive internal events.

Situational Judgment Tests (SJT): assessments designed to measure a participant's judgment in work-related situations.

Socially desirable responding (SDR): a multifaceted and dimensional construct; self-deception enhancement (SDE) and impression management (IM) are identified as the two most prominent constructs associated with socially desirable responding (SDR).

Stakeholder and Relationship Building (SBR): A competency used in the Management and Leadership Skills Test (MLST), for an overview of the behaviour indicators please see Appendix C.

Strategic Thinking (ST): A competency used in the Management and Leadership Skills Test (MLST), for an overview of the behaviour indicators please see Appendix C.

Introduction

Situational judgment tests (SJTs) are assessments designed to measure a participant's judgment in work-related situations. Situational judgment tests (SJTs) present a participant with situations that may be encountered in an organizational context, and instruct the participant to evaluate a set of response items related to the situation in which they must identify or rank the most appropriate response. Situational judgment is the procedural knowledge of effective actions (Motowidlo & Beier, 2010) and the assessment of options and their appropriateness (Rockstuhl et al. 2015). In organizational contexts, SJTs gather insightful data on an individual's performance and their tendency to behave in a predictable way (McDaniel & Whetzel, 2007). Participants consider a situation provided and select the most effective answer out of four (4) to five (5) response options. The assumption is that an assessor can measure a participant's judgment in a specific context and predict how they will behave in a real-life situation. In organizational contexts, personality assessments are used; however, these assessments are criticized for their susceptibility towards social desirability responding. Although it can be assumed, there is less evidence to suggest that SJTs face the same limitation, as research produces diverse conclusions. For the purpose of this study, social desirability is a tendency to bias responses in a favorable direction (Edwards, 1970) in order to improve assessment outcomes.

The ability to predict and measure a participant's personality, cognitive ability, and judgement enables hiring organizations to select best-fit applicants. SJTs can also be oriented towards professional development and learning, especially in educational settings (e.g., medical training and other high-fidelity positions). Psychometric testing paired with structured interviews and reference checks can create a conclusive portfolio in which the organization can insightfully predict how an applicant will integrate and operate in a work-related context. Judgement is an increasingly utilized predictor of job performance alongside conscientiousness and emotional stability; it is an individual's ability to form an opinion or make a decision after clear thinking (Merriam-Webster); thus making it a valued attribute for employers.

The primary objective of this research was to explore the possibility in which participants distort their responses in a socially desirable way during a competitive selection process with the intent to increase their scores. Due to the increasing use of psychometric assessments as a means to assess and select individuals in organizational contexts, the development and administration of such assessments must be evaluated to ensure that participants are provided a fair opportunity to

demonstrate their competence and ability. Additionally, hiring organizations are selecting top performing participants, who achieved top scores through honest means.

As such, the relationship between social desirability responding and SJTs was examined in order to identify if socially desirable behavior positively influences participants' SJT scores. Social desirability, SJTs and response latency were examined to identify if a relationship exists between these three constructs. A social desirability responding scale, the Balanced Inventory of Desirable Responding (BIDR) was used to assess the extent in which participants distort their responses in a socially desirable direction. The Management and Leadership Skills Test (MLST) was used to assess five essential competencies related to the competitive selection process. The MLST results were examined globally and individually across the five competencies; the participant's composite and sub-scale BIDR social desirability scores were cross-examined with the five competencies in order to analyze the relationships between social desirability and competency scores. Participant MLST and BIDR results were also cross-examined with response latency in order to examine if there was a relationship between response latency, competency scores and social desirability scores.

Research Problem

As psychometric research continues to develop and expand, the questions and theories produced extend scientific reach and prompt further academic inquiries. Assessments are gathering momentum in organizational and educational contexts, and are used to identify, assess, and compare individuals across domains (e.g., public servant, doctor, and manager) and measures (e.g., personality, knowledge and competency). In order to accomplish this intended outcome, assessments and their respective results must be reliable and valid across time, administration, and participants. As such, assessments must be vigorously examined for limitations and deficiencies, challenged and analyzed, and developed and adapted to meet the high standards set by the academic and assessment community. One of these challenges is a participant's ability to respond in a socially desirable way in order to positively influence their assessment results and be selected for a competitive position. Current findings are convoluted and provide no clear direction in regards to the possible effects of social desirability responding in organizational contexts; does socially desirable behavior enable a participant to increase their scores? What are the effects on the outcome of the selection process? How can social desirability be identified and quantified? A second chal-

lenge is the use of incumbents in social desirability research; there is a requirement for real-applicants to be integrated into research in order to ensure the results are applicable to real-life contexts. A third challenge is the use of response latencies as a detection and prediction method for social desirability responding. Further research is required in order to identify if response latencies are a reliable way of predicting social desirability responding as there is convoluted evidence in regards to the length of response latencies and their correlation with cognitive ability and social desirability responding.

Literature Review

Overview of Situational Judgment Tests

Situational judgment tests (SJTs) are self-report measures used to assess a participant's judgment when presented with situations that could be encountered in a work-related setting (Weekly & Ployhart, 2006; Whetzel & McDaniel, 2009). This type of assessment presents participants with work-related situations and a list of possible response options in which they must evaluate and select the most appropriate response. Situational judgement tests have gained momentum in organizational selection contexts since the late 1900's. The earliest example of an SJT depends on how it is defined (Weekly & Ployhart, 2006). The United States Civil Service used one of the earliest documented assessments that featured similarities to a SJT in 1873. The Civil Service used situationally based questions to assess action-based behavior, such as, "A banking company asks protection for a certain device, as a trade-mark which they propose to put upon their notes. What action would you take on the application?" (DuBois, 1970). Similarly, the 1905 Binet scale measures intelligence in children; this scale contained questions such as, "When a person has offended you and comes to offer his apologies, what should you do?" The main differentiation between these types of questions and the standard format seen in SJTs today is the absence of response options. One of the first and more popular SJTs that contained response options was the Judgment in Social Situations, a subtest of the George Washington Social Intelligence Test (McDaniel et al., 2001). SJTs were integrated into organizational contexts to assess supervisor potential in the 1940s (Cardall, 1942; Bruce & Learner, 1958) and managerial potential in the 1960s (Campel et al., 1970). SJTs have also been implemented outside of organizational contexts; military psychologists during World War II sought to assess judgment by presenting

soldiers with scenarios and response options (Northrop, 1989). Research began to accumulate during the 1980s and 1990s on the predictive validity of SJTs and their ability to predict future job performance (Corts, 1980; Motowidlo, Dunnette, & Carter, 1990). Because of this research, SJTs are considered to be an effective assessment tool and are frequently used in organizational selection processes (McDaniel et al., 2001; Salgado, Viswesvaran, & Ones 2001). Further research has found that SJTs have useful levels of validity as predictors of job performance, have lower sub-group differences and adverse impacts than cognitive measures (McDaniel & Nguyen, 2008; Motowidlo & Toppins, 1993; Weekly & Jones, 1997, 1999), and have provided significant evidence of face and content validity (Motowidlo, Hanson, & Crafts, 1997; Salgado et al., 2001).

Socially Desirable Responding. Socially desirable responding (SDR) is a term applied to “positive self-presentation on self-report questionnaires” (Paulhus, 1991). When participants are asked to rate their own personalities, competencies, or abilities, they have a tendency to bias their responses in a favorable direction (Edwards, 1970). In other words, participants are likely to participate in self-enhancement behaviors by exaggerating or inflating their positive qualities (Paulhus & Holden, 2009).

Factor analyses may be used to examine socially desirable responding (SDR) (Edwards, 1970; Wiggins, 1959); the results of these analyses have supported the claim that SDR is a multifaceted and dimensional construct. A two-factor model presented by Holden and Paulhus (2009) identifies self-deception enhancement (SDE) and impression management (IM) as the two most prominent constructs associated with SDR. Impression management (IM) is described as the behavior demonstrated when individuals consciously and publicly make exaggerated claims of desirable behavior. Sackeim and Gur (1978) characterize impression management as the behavior in which the individual consciously and deliberately distorts the truth; for example, “I always pick up my litter”. Self-deception enhancement (SDE), in contrast to impression management (IM), is associated with positive internal events such as, “My first impressions about people always turn out to be right.” These types of statements are easier to rationalize without overt awareness (Paulhus & Holden, 2009).

Organizational Context. Situational judgment tests (SJTs), or job simulations (Motowidlo et al., 1997; Thornton & Cleveland, 1990), have emerged as the tool of choice in organizational selection contexts (Patterson et al., 2012) due to their ability to simulate low-fidelity work-related situations (Motowidlo et al., 1990) and identify behavioral consistency (Deborah, Whetzel & McDaniel, 2009). Behavioral consistency rests on the assumption that past behavior is the best

predictor of future behavior; by sampling current work-related behaviors, an organization can theoretically predict how a participant may behave in the future (Wernimont & Campbell, 1986). Research shows that SJTs have significant incremental and predictive validity of job performance (Nguyen, McDaniel & Biderman, 2005).

Although promising, there is evidence to suggest that SJTs are susceptible to socially desirable responding in organizational selection contexts. This claim however has divided the research community. SJTs were identified as “fakable” in one study using behavioral tendency response format (would/would not do) (Vasilopoulos, Reilly & Leaman, 2000) and not “fakable” in another study using the same response format (Juraska & Drawsgow, 2001). McDaniel and Biderman (2005), state that the importance of these types of studies is dependent on the impact that socially desirable behavior has on the participants’ test scores and the outcome of the organizational selection process. In organizational contexts, participants want to achieve the highest score; if this motivational desire is consistent across all applicants, then faking becomes a constant across all applicants’ scores, and its importance minimal. However, McFarland and Ryan (2000) provided evidence that suggests there are individual differences in faking. As a result, it may be inferred that social desirability affects the validity and reliability of the SJT as well as the outcome of the organizational selection process.

If this is the case, social desirability responding has a significant effect on who is selected in organizational selection processes (Rosse et al., 1998; Weiner & Gibson, 2000). Peeters and Lievens (2005) demonstrated that participants instructed to fake on an SJT had significantly higher scores than participants instructed to respond honestly. They computed the results of the highest quartile and found that seventy-six (76) percent of participants in this quartile were instructed to fake and that twenty-four (24) percent of participants in this quartile were instructed to respond honestly. These percentages demonstrate that social desirability responding has substantial effects on participant performance (Whetzel & McDaniel, 2009). Other studies have presented similar findings (e.g., Hooper & Sackett, 2008; Berry & Sackett, 2009; Viwesvaran & Ones, 1999; Whetzel & McDaniel, 2009). Even so, as research accumulates there is still doubt that SJTs and other self-report measures can be faked (Schmitt & Oswald, 2006).

Measuring social desirability. The Balanced Inventory of Desirable Responding (BIDR) is a social desirability scale used to measure two constructs: self-deceptive enhancement (SDE) and impression management (IM). A descendant of the Self and Other Deception Questionnaires developed by Sackeim and Gur (1978), the BIDR consists of forty (40) propositional items; for

example: “My first impressions of people usually turn out to be right.” Participants rate their agreement with each statement on a seven-point scale, ranging from ‘most likely’ (7) to ‘least likely’ (1). The BIDR proves to have significant internal consistency with values of coefficient alpha ranging from .68 to .80 for SDE and .75 to .86 for IM. BIDR items also have concurrent validity aligned with other social desirability scales (for example, the Marlowe-Crown scale and the Multidimensional Social Desirability Inventory) with a correlation of .71 (Paulhus, 1991).

Research by Hough (1998) and Rosse (1998) identified that higher scores on social desirability scales are present with job applicants in comparison with non-applicants; these findings validate the use of such scales. This difference in scores can be attributed to socially desirable responding and job applicant motives to increase test scores. This concern is not new, as social desirability scales have been embedded in personality measures since the 1940s (such as the Minnesota Multiphasic Personality Inventory, and the 16 Personality Factor) (Paulhus, Bruce, & Trapnell, 1995).

See Appendix A for the full forty (40) items, Balanced Inventory of Desirable Responding (BIDR), Seventh Edition.

Response Latency. Paulhus and Holden (2009) introduce response latency as a component of response distortion techniques used by applicants. Cognitive structures, also known as schemas, direct knowledge and are based on experiences and self-representation. These structures include personality traits that represent components of an individual’s schema. A self-schema will prompt processing for both relevant and irrelevant information; specifically, if information is consistent with the individual’s cognitive structure, or schema, then it will be accepted faster and rejected more slowly. The same line of thought follows for information that is inconsistent with the individual’s cognitive structure; information will be rejected faster and accepted more slowly. This self-schema processing endorses positive and negative personality components that are aligned with the individual’s cognitive structure. Concluding their argument, Paulhus and Holden (2009) state that the behavioral components in social desirability scales, criterion comparison methods, and behavioral approaches (such as response latency) enhance self-enhancement methods. Lastly, they provide high praise for response latency and advocate for the necessity of self-reports.

These findings contribute to the mixed evidence that response latencies provide insight into social desirability responding and cognitive ability. McDaniel and Timm (1990), and Follette (1984) suggest that participants who have longer response latencies on personality measures are

more likely to have higher scores on social desirability scales (Holden et al., 2001). In contrast, Hsu, Santelli and Hsu (1989) and George (1990) suggest that social desirability responding is associated with faster response latencies and requires minimal cognitive processing as lying is a primitive function. Hsu, Santelli and Hsu (1989) highlight research that provides evidence to suggest that self-referenced interpretations of words require longer response latency times than semantic, structural, or phonemic interpretations. They also draw from interrogation research that identifies how response latency is affected by the intent to deceive in interrogation-type situations.

Instruction Set. Researchers hypothesize that different constructs are measured depending on the instruction set used (McDaniel & Nguyen 2001). There are two main instruction sets used in SJTs, knowledge based and behavioral tendency based. The knowledge based instruction set “most” and “least” effective response (sometimes described as the best and worst response) is based on the participant’s knowledge; in contrast the “most” and “least” likely response (sometimes described as would do and would not do response) is correlated with the participant’s behavioral tendencies (McDaniel & Nguyen, 2001).

McDaniel et al. (2007) compared these two types of SJT response instructions and correlated the scores with cognitive ability and personality. The Big Five Personality traits have varying degrees of correlation with SJTs (Digman, 1990) and cognitive ability (McDaniel et al., 2007). The magnitude of the correlation between SJTs and personality (the Big Five) and cognitive ability is moderated by the response instructions; behavioral tendency instructions are significantly correlated with personality traits, while knowledge based instructions are significantly correlated with cognitive ability. Behavioral tendency instruction sets are proven to have limited validity due to their susceptibility to socially desirable responding. The criterion related validity and incremental validity of a SJT can be affected by socially desirable responding due to the instruction set; as a result, behavioral tendency instructions in competitive, organizational selection processes may be problematic (McDaniel & Whetzel, 2009).

In a study by McDaniel and Nguyen (2001), three groups completed a SJT in which each group received different instruction sets. The first group was instructed to identify the most and least likely response; the second group received the same instructions, however they were also instructed to fake the test to improve their scores with the promise of monetary incentive; the third group was instructed to identify the best and worst response to each scenario. As a result, the first group and the third group had nearly identical mean scores, which were 0.5 standard de-

viations above the second group; this suggests that the instruction set with the most and least likely response allows for score inflation through faking. Additionally, the best and worst response instruction scores were not improved when the participants were asked and provided monetary incentive to fake. However, this instruction set is not completely resistant to faking (McDaniel and Nguyen, 2001); it is suspected that if participants responded honestly during an SJT with most and least likely instructions, the assessment would have increased validity levels over the best and worst response instructions. Research shows that behavioral tendencies are better predictors of future behavior than knowledge (McDaniel & Nguyen, 2001); however, they are more susceptible to faking.

These two instruction sets prove to be important elements of a useful and valid assessment, although consequently, they increase the difficulty of item and reliability analysis. In other words, the instruction set makes the item responses partially ipsative (Hicks, 1970). In order to resolve the problem of partial ipsativity, researcher's advocate for the most and least effective response instruction set in which the participant is asked to identify the effectiveness of each response. The effectiveness of one response is not dependent on the other, and as a result there is no ipsativity (McDaniel & Nguyen 2001).

Competencies. Competency-based assessments provide insight into the work approach and behavior associated with effective performance in work-related contexts. A competency is a set of defined behavioral indicators that provide a standardized way of identifying, assessing and developing work-related behaviors. Competency-based assessments can be used to screen applicants for positions, training, professional development and performance management. Competencies can be used in psychometric testing like personality tests and SJTs and provide a standardized description of constructs to be assessed. In SJTs, situations and response options often reflect the competency being assessed.

Strengths of Situational Judgement Tests

Cognitive Ability. Situational judgement tests (SJTs) place cognitive demands on participants (Marentette, Meyers, Hurtz, & Kuang, 2012). Research has established cognitive ability and SJT measures as valid predictors of performance (e.g., Lounsbury, Sundstrom, Loveland & Gibson, 2003; Poropat, 2009; Bobat, Caruth & Buitendach, 2012). Alongside the predictive nature of these measures, studies have identified a relationship between SJT scores and cognitive

ability (e.g., Bergman et al., 2006; McDaniel et al., 2001) with a r of .46 (Patterson et al., 2012). SJTs are also thought to measure the knowledge of job related situations, $r=.23$ (Patterson et al., 2012) more so than behaviour (Motowidlo et al., 2008). As knowledge is a cognitive construct, this may explain why SJTs are often related to cognitive ability (McDaniel et al., 2007; Osstrom et al., 2011). Cognitive load is also associated with SJTs due to their response format; participants must cognitively consider multiple response options prior to selecting a final response (Nowakowska, 1970; Holden, Wood & Tomaszewski, 2001). Studies directed at establishing a relationship between SJT situations and cognitive ability (Weekly & Jones, 1999; McDaniel et al., 2001; Mullins & Schmitt, 1998; Motowidlo, Dunette & Carter, 1990) have shown these to be moderate (Bobat, Caruth, Buitendach, 2012).

Validity. As stated by the Principles for the Validation and Use of Personnel Selection Procedures (SIOP Inc, 2003) validity is the extent in which gathered data supports the interpretation of a test score. This traditional concept of validity is being used to utilize a broad and convenient way of organizing and discussing validity (Brualdi, 1999). Research has summarized the validity evidence relevant to constructs assessed by SJTs (McDaniel et al., 2001; McDaniel et al., 2007; McDaniel & Nguyen, 2001) and this evidence suggests that social desirability responding affects the validity of the test and outcomes of selection processes (Heggstad et al., 2006; Berry & Sackett, 2009).

Face Validity. Face and content validity are strengths of SJTs (Osam, 2014) and is a contributing factor to SJT popularity in organizational contexts (Rynes & Connerly, 1993). Due to the low-fidelity, work-related context of an SJT, participants are documented as having positive reactions when completing the test (Motowidlo, Hanson, & Crafts, 1997; Salgado et al., 2001). This positive reaction is linked to a participant's opportunity to demonstrate their ability to perform on work-related tasks, and this type of assessment is also thought to invoke motivation and engagement during participation (Osam, 2014).

Criterion-Related Validity. Criterion-related validity is the extent in which a measure is related to an outcome. This type of validity provides strong evidence towards the correlation between SJTs and work-related contexts (Whetzel & McDaniel, 2009; Dye, Reck & McDaniel, 1993). Motowidlo, Dunnette and Carter (1990) summarized studies pertinent to establishing the criterion validity of SJTs and performance. A meta-analysis by McDaniel et al. (2001), indicated that SJTs have an uncorrected mean validity of .26 with performance; other studies indicate moderate correlations ranging from $r=.30$ to $.35$ (Chan & Schmidt, 2002; Motowidlo et al., 1990; Mo-

to Motowidlo & Tippins, 1993; Weekly & Jones, 1997; Smith & McDaniel, 1998). Research suggests that measures of social desirability typically do not have a significant impact on the criterion related validity of SJTs (Schmitt & Oswald, 2006). Predictive validity research identifies similar findings; meta-analytical reviews of SJT research have suggested useful levels of validity as predictors of job performance. Due to the nature of today's dynamic work environment, the ability to judge effective actions in work-related contexts is essential (McDaniel & Whetzel, 2009). In a study conducted by Chan (2006), situational judgement effectiveness was significantly correlated with predicted work outcomes of pro-activeness in work-related situations.

Construct Validity. Research suggests that construct validity, the extent in which an instrument actually measures the construct that it was designed to measure (Bagozzi, Yi & Phillips, 1991), is evaluated in SJTs through identifying comparisons between established constructs. As SJTs are related to multiple constructs (Osam, 2014), it can be assumed that this approach is an appropriate way to establish construct validity.

Incremental Validity. Another potential strength of SJTs is incremental validity: the increase in predictive validity when a predictor is used with other established selection measures (Patterson et al., 2012). Meta-analysis and primary studies examined the incremental validity of SJTs over measures of cognitive ability (Chan & Schmitt, 2002; Clevenger & Haaland, 2000; O'Connell et al., 2002; Weekly & Jones, 1997, 1999; McDaniel et al., 2001, 2007). These studies provided significant evidence to suggest that SJTs offer incremental validity over composites of cognitive ability and personality with values ranging from $r = .01$ to $.02$ (McDaniel & Whetzel, 2009). In an experiment by McDaniel and Whetzel (2009), SJTs showed incremental validity above cognitive ability, personality, and experience measures.

Adverse Impact. Adverse impact measures the effect that assessments and other selection methods have on minority classes, especially those methods that might disadvantage individuals of a particular sex, race or group. SJTs have less adverse impact when compared to cognitive ability assessments (Patterson et al., 2012; Whetzel, McDaniel & Nguyen, 2008; Chan & Schmitt, 1997; Motowidlo & Tippins, 1993; Weekly & Jones, 1999); however, differences across groups still exist. Females tend to score higher than males on SJTs with mean scores ranging from $d = .1$ to $.27$ (O'Connell et al., 2007; Patterson et al., 2012); these differences can be attributed to gender difference in personality traits that may be triggered when responding to SJT situations (Livens et al., 2008). Livens et al., (2008) suggests that race and ethnic differences can be attributed to the cognitive components associated with SJTs and can be mitigated by diminish-

ing situations with high cognitive load. The adverse impacts associated with assessments can be evaluated by consequential validity. Consequential validity refers to how assessment results may be interpreted and utilized in social contexts, and in this instance, competitive selection processes. Consequential validity is used in formal validity investigations (Messick, 1995) and evaluates the validity of an assessment based on the consequences - in other words, the inferences and decisions made based on the assessment results (Mehrens, 1997).

Reliability. The reliability of SJTs has been questioned by the research community, however, most research has acknowledged acceptable levels of reliability (McDaniel et al., 2001) with internal consistency coefficients ranging from 0.43 to 0.94 (Osam, 2014). Due to the multidimensional nature of SJTs, it is difficult to estimate the reliability across multiple constructs. Reliability is most suited for unidimensional tests that measure one construct (Patterson et al., 2012). An example from McDaniel and Whetzel (2005) demonstrates this concept using the response option: “You give your supervisor an update and express your concern about your ability to complete the project within the deadline.” The participant may choose this option because they are conscientious and/or intelligent; as a result, the response option can be attributed to, and correlated with both constructs.

Assessment design. Other limitations include the length of the assessment, as more items prove to have higher levels of reliability, and the format of the instruction set used (Osam, 2014). Ployhart and Ehrhart (2003) indicate that instruction set (whether behavioral tendency or knowledge format) affect the internal consistency of the SJT. The most accurate way to measure reliability is a test re-test or a parallel forms approach (O’Connell et al., 2007). Test re-test reliability coefficients have been documented as high as 0.92 (Ployhart & Ehrhart, 2003), however this high of a correlation is rarely reported (McDaniel & Whetzel, 2009).

Weaknesses of Situational Judgment Tests

There are two main weaknesses of situational judgment tests (SJT), applicant faking and coaching. This section identifies and summarizes research on how applicants may enhance their responses to appear more desirable in competitive selection processes.

Faking. The deliberate distortion of responses in order to score favorably on a selection measure most likely has an effect on the validity of the assessment (Whetzel & McDaniel, 2009). The effect that faking has on validity has created a “rancorous debate” (McDaniel & Nguyen,

2001) between researchers. Two main positions have been debated: first, that participants do not fake, and if they did it would not negatively affect the validity of the assessment (Abrahams, Neumann & Githens, 1971; Ellingson, Smith & Sackett, 2001; Hough, 1998; McCreae & Costa, 1983, Ones & Viswesvaran, 1998), and; faking occurs and attenuates the validity of the assessment (Holden, Wood & Tomashewski, 2001; Heggstad, et al., 2006; Douglas et al., 1996; Dunnette et al., 1962; Kluger, Reilly & Russell, 1991, Schmit & Ryan, 1992; Zickar, 1997). However, there appears to be agreement that faking harms the use of the assessment (Douglas, McDaniel, & Snell 1996; Ones & Viswesvaran 1998) or in other words, the assessment's ability to accurately measure constructs associated with a particular position and the prediction of work performance.

Coaching and Practice Effects. As an assessment gains popularity in organizational contexts, it too will gain popularity in coaching strategies (Lievens, Peeters & Schollaert, 2008). Participants will seek coaching and adopt strategies to increase their chance of being selected. There is limited research on the effects of SJT coaching. Cullen, Sackett and Lievens (2006) examined the coach-ability of SJTs and found that participant performance could be enhanced by coaching, however SJT strategies are more difficult for participants to learn due to the input of subject matter experts in the developmental stages. Other assessments have a clear set of rules that are consistent across test variations, but due to the often customized nature of SJTs, it is much harder to anticipate and learn the strategies beforehand (McDaniel & Whetzel, 2009).

Limitations of Situational Judgment Tests

There are three main limitations of situational judgment tests (SJT), multidimensional constructs, concurrent designs, and the reinforcement of socially desirable behavior. This section identifies and summarizes research on the limitations associated with the design of the SJT.

Multidimensional Constructs. It is desirable for psychometric assessments to consist of uni-dimensional constructs or items (Osam, 2014), however SJT constructs are multidimensional, also defined as construct heterogeneous. As a result, the inability to target specific individual differences across constructs is a major limitation of SJTs (McDaniel & Nguyen, 2001). This limitation proves to make it difficult when identifying what constructs are being measured (Arthur & Villado, 2008; McDaniel & Whetzel, 2009). The assessment of two constructs in one measure limits precision and hinders the ability to explain the predictive validity of

a measure. An additional limitation is the ability to develop new measures that assess the same characteristics (McDaniel & Whetzel, 2009; Mumford & Whetzel, 1997).

Situation judgment tests can be designed to measure a variety of constructs depending on the intended outcome of the assessment (McDaniel & Nguyen, 2001). SJTs may be formatted as a multiple choice assessment and/or video simulation. Video-based simulations are becoming more common (Lievens & Sackett, 2006) as technology advances, especially in medical admission and training.

Concurrent Designs. The theoretical basis for SJT research is lacking, and there has been a recent call for research using real job applicants in organizational contexts, especially with situational judgement tests (McDaniel & Whetzel, 2009). Current research relies heavily on concurrent designs in which participants are incumbents with minimal motivation to distort their responses (McDaniel & Whetzel, 2009). An applicant competing in a selection process is more likely to participate in social desirability responding to improve their chances of being selected. Incumbents are often asked through instruction sets to “fake good”, however as stated by Bikerland, et al., (2006) “research utilizing the induced faking paradigm does not address job applicant faking or response distortion as part of a real-life selection process.” As a result, research using real job applicants is warranted. Response instructions and the constructs assessed in SJT’s have been thoroughly examined; however, the design used in SJT research has proven to have limitations due to the tendency to use concurrent designs. The majority of research in SJTs uses students or incumbents as their population. This limitation may alter the usefulness and representativeness of the study because incumbents have less motivation to distort their responses in order to make a favorable impression, even if instructed to do so. Whetzel and McDaniel (2009) warrant additional research using participants applying in competitive selection processes. Bikerland et al., (2006) make a similar claim suggesting that research utilizing the induced faking paradigm does not address social desirability responding as part of a realistic selection process.

Reinforcing Fakers. Participants are capable of increasing their test score through socially desirable responding. This intentional effort of social desirability responding increases the likelihood that participants will be selected in competitive selection processes (Berry & Sackett, 2006). This capability jeopardizes the fairness of SJT assessments as it alters the rank ordering of participants, both deserving (honest) and undeserving (fakers). As a result, undeserving participants are rewarded in competitive selection processes; this outcome significantly effects which participants are hired (Rosse et al., 1998; Weiner and Gibson, 2000; Whetzel & McDaniel, 2009). To support

this claim, Rosse et al., (1998) and Mueller- Hanson et al., (2003) identified that small groups of extreme fakers affect selection decisions and Bikerland et al. (2006) warrant additional research on how faking can alter the rank order of participant scores and thus hiring decisions.

Research Questions

Further research is required in order to identify and understand how socially desirable responding affects the results and interpretations generated from the use of situational judgement tests (SJTs) in organizational contexts. As such, the research questions are as follows:

What is the relationship between social desirability responding and the scores on the Management and Leadership Skills Test (MLST)? *It is hypothesized that Management and Leadership Skills Test (MLST) composite mean scores and Balanced Inventory of Desirable Responding (BIDR) composite mean scores will be positively correlated. Respondents BIDR composite scores and BIDR subscale scores (Self-deceptive enhancement, and Impression Management) will be compared with the MLST composite scores and MLST scores by competency (Stakeholder and Relationship Building, People Development and Management, Change Leadership, Decisiveness, Strategic Thinking) to identify if a positive relationship exists between social desirability responding and competency.*

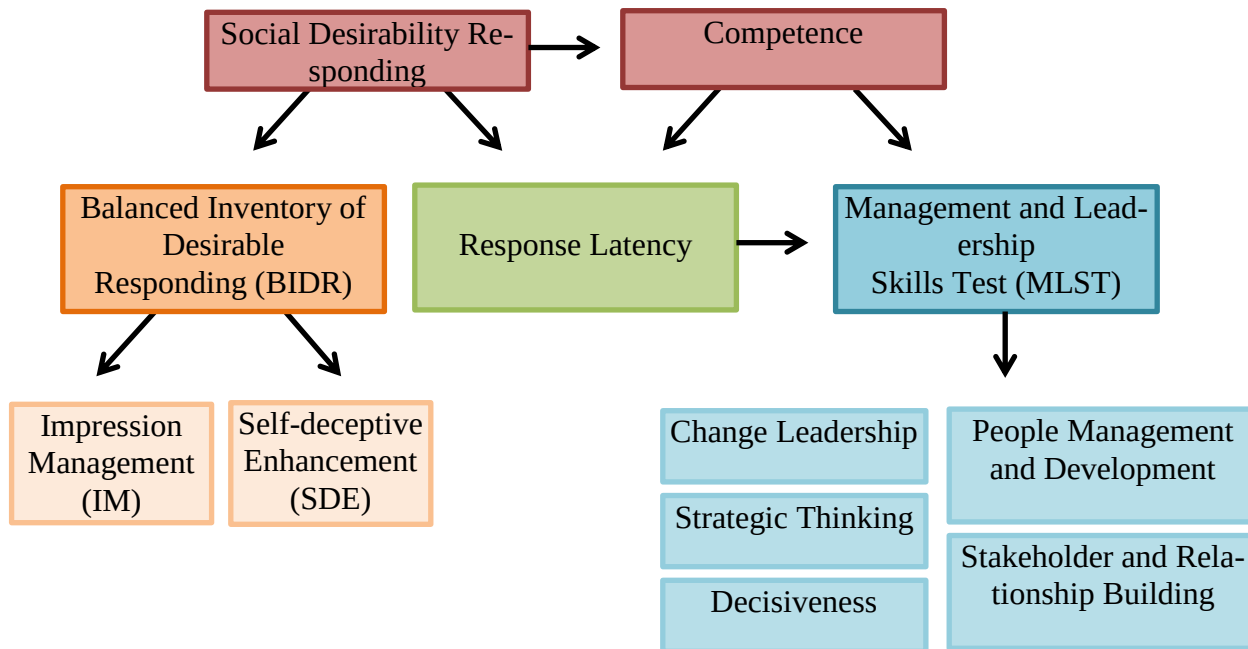
What is the relationship between response times and the Balanced Inventory of Desirable Responding (BIDR) and their effect on Management and Leadership Skills Test (MLST)? *It is hypothesized that longer MLST response latencies will reflect high BIDR composite scores; as longer response latencies are associated with social desirability responding, it is hypothesized that respondents who have high BIDR scores will also have longer response latencies in the MLST thus validating social desirably responding. It is hypothesized that there will be a correlation between the MLST, the BIDR and response latency for each participant.*

Theoretical Framework

The primary objective of this research was to explore the possibility in which participants distort their responses in a socially desirable way during a competitive selection process. The Balanced

Inventory of Desirable Responding (BIDR) was administered to assess social desirability and the Management and Leadership Skills Test (MLST) was used as the competitive selection assessment. Additionally, this research explores the use of response latencies as a predictor of social desirability responding. For an overview, please see Figure 1.

Figure 1. Theoretical Framework



Method

Participants

Twenty-nine (29) participants completed the Management and Leadership Skills Test (MLST) and the Balanced Inventory of Desirable Responding (BIDR) in a competitive selection process with the Canadian Food and Inspection Agency (CFIA) for a first level Director (EX-01) position. Eleven (11) participants were removed from the study because they completed the MLST, but did not complete the BIDR. This data was not considered for the purposes of this research, and only complete data sets were included. In total, eighteen (18) participants were analyzed for the purposes of this research. Six (6) individuals identified themselves as female, and twelve (12) individuals identified as male. Participant's age ranged from 31 – 61+; however, fifty

percent (50%) of participant's ages ranged from 41-50. All participants identified as Canadian citizens who currently reside in the National Capital Region.

Measures

Two assessments, one to assess competence and one to assess social desirability were hosted on an online testing platform.

The Management and Leadership Skills Test (MLST). The MLST was developed by EPSI, a company that specializes in psychometric assessment. EPSI was contracted by the CFIA to develop a situational judgment test to assess five competencies: Stakeholder and Relationship Building, People Development and Management, Change Leadership, Decisiveness, Strategic Thinking. The test consists of twenty-six (26) questions across fifteen (15) situations and participants receive one (1) mark for selecting the most effective response and a zero (0) for not selecting the most effective response. Each situation was developed to reflect the work environment of an executive (EX) position at the organization and is correlated with a competency. The subsequent questions reflect specific behavioral indicators related to that competency. The test is structured as follows; participants are required to electronically sign a consent form confirming confidentiality, and psychological and physical disposition. Once consent is provided, participants may review the instructions for the test and are provided with a sample situation and question. This information can be found in the Appendix D. Next, the participants are introduced to the fictitious organizational context and background information. This introductory information then transitions into the fifteen (15) situations and their respective questions. Participants are able to select from five response options for each question. For a full description of the development of the test, please see Appendix B.

The response format selected for the MLST is the 'most' effective response. This type of instruction also induces a forced choice design, in which participants must choose between the response options and select the one that they consider to be the most effective.

The Balanced Inventory of Desirable Responding (BIDR). The BIDR is a social desirability scale used to measure two constructs: self-deceptive enhancement (SDE) and impression management (IM). Impression management (IM) is the behavior demonstrated when individuals consciously and publicly make exaggerated claims of desirable behavior; for example, "I always

pick up my litter”. Self-deception enhancement (SDE), is associated with positive internal events such as, “My first impressions about people always turn out to be right.”

A descendant of the Self and Other Deception Questionnaires developed by Sackeim and Gur (1978), the BIDR consists of forty (40) propositional items; for example: “My first impressions of people usually turn out to be right.” Participants rate their agreement with each statement on a seven-point scale, ranging from ‘most likely’ (7) to ‘least likely’ (1). The scoring key of the BIDR is balanced, negatively keyed items are reversed, and one point is allotted for each extreme response (6 or 7). This type of scoring procedure is designed so that participants who respond in a socially desirable manner are more likely to have higher scores. The BIDR proves to have significant internal consistency with values of coefficient alpha ranging from .68 to .80 for SDE and .75 to .86 for IM. The BIDR demonstrates significant levels of concurrent, construct, convergent and discriminate validity (Paulhus, 1984; Paulhus 1988).

Response Latency. Response latency is measured by the elapsed time between the selection of a response in one question and the selection of a response in the next question. Self-schemas prompt information processing; theoretically, if information is consistent with the individual’s schema, then it will be accepted faster and rejected more slowly, and if information is inconsistent with the individual’s schema; information will be rejected faster and accepted more slowly (Paulhus & Holden, 2009). In order to calculate response latency, an approach by Paulhus and Holden (2009) was used. As information processing models were applied to psychometrics, self-schema theories led to the use of adjusted response times rather than raw response times. Raw response latencies do not reflect the distinct stages for stimulus encoding, comprehension, decision making, and selection. Standardized response latencies consider various item properties that influence the speed of processing and item factors such as item length, complexity, number of response items, motor speed, reading ability, etc. The control of main effects for items and individuals is critical in the calculation of response latencies and are considered more reliable by researchers (Holden et al., 1991; Holden & Hibbs, 1995; Holden, Kroner, Fekken, & Popham, 1992; Mayerl, 2013).

The following approach was used to compute and standardize the response latency variable used in this study:

1. Raw response latency times (in minutes and seconds) were identified per participant for each of the twenty-six questions.

2. Raw response latency data was removed to exclude participants who did not provide consent or who did not complete the BIDR.
3. Raw response latency times were reset so that maximum latencies were 3.00 and minimum latencies were -3.00 (values outside of this range were regarded as outliers that will unduly influence the analyses (Paulhus & Holden, 2009; Mayerl, 2013)).
4. The data was then standardized within a respondent to adjust for irrelevant factors such as reading speed, verbal ability and motor speed.
5. Data was standardized within an item to correct for irrelevant factors such as item length, complexity, and order. As such, results are standardized times that represent latencies relative to the respondent and relative to the item.
6. The standardized times were reset so that maximum latencies were 3.00 and minimum latencies were -3.00.
7. The data was aggregated by computing mean latencies within a respondent.

Secondary Data Analysis

For the purposes of this research, a secondary data analysis was utilized; researchers were provided access to the MLST and BIDR results. With the approval of the organization and the University of Ottawa Ethics Approval Board, the data set provided valuable information on social desirability responding through the use of situational judgment tests in competitive organization contexts. The organization and participants have consented to the use of this data, with the condition of security and confidentiality of the information. In order to maintain the confidentiality of the participants, their names have been anonymized and replaced with a random six digit number. The composite scores and sub-scale scores for both assessments and response latencies were analyzed for the purposes of this study.

The data used in this research underwent three separate quality checks at various stages of the process to ensure that the data entry, cleaning and calculations were completed correctly. Data entry was validated after completion, and the cleaning and manipulation of data was reviewed to ensure the appropriate modifications had been made. An expert at the organization who supplied the data participated in these quality checks. The expert's background is in psychometrics and organizational psychology in the public service.

Analysis and Results

Descriptive Statistics

MSExcel was used to calculate the descriptive statistics in this study, including the means and standard deviations. The mean for the MLST composite scores was 35.56 out of a possible 63 (56.4%) with a sd = 9.11. The group mean was below the expected pass mark required to be considered for the position. The means for decisiveness, strategic thinking, stakeholder building relationships, people development and management, and change leadership competencies are as follows (Table 1). The mean composite score for the BIDR was 18.39 out of a possible 40 with a sd = 9.38. The mean for the BIDR sub-scale SDE was 9 with a sd = 4.47, and the sub-scale IM mean was 9.58 with a sd = 3.18.

Table 1. Mean and standard deviation for the MLST competencies

Competency	Mean and Standard Deviation
Decisiveness	13.61 out of 26 (52.3%) with a sd = 3.68
Strategic thinking	6.33 out of 11 (58%) with a sd = 2.11
Stakeholder building relationships	3.22 out of 8 (40%) with a sd = 1.55
People development and management	6.56 out of 11 (59%) with a sd = 1.50
Change leadership	3.83 out of 7 (55%) with a sd = 1.51

Table 2. BIDR Categorizations

Label	Range	Description
High BIDR respondents	24-40	Participants who had high BIDR scores were categorized as individuals who would be likely to respond in a socially desirable manner during an assessment.
Moderate BIDR respondents	16-23	Participants who received moderate BIDR scores are considered neither purposefully faking nor being completely honest.
Low BIDR respondents	0-15	Participants who had low BIDR scores were categorized as low or honest respondents. These participants did not

		respond in a socially desirable manner during the BIDR.
--	--	---

Table 3. Frequencies and percentages of BIDR composite and sub-scale

		BIDR Scales		
		Composite	SDE	IM
BIDR Results	High	7 (38%)	5 (28%)	9 (50%)
	Moderate	5 (28%)	4 (22%)	4 (22%)
	Low	6 (33%)	9 (50%)	5 (28%)

Table 4. MLST Categorizations

Label	Range	Description
Passed MLST	38-60	Participants who scored thirty-eight (38) and higher are identified as receiving high MLST scores. These participants were met the pass mark established and have the potential to be selected for the position.
Failed MLST	0-37	Participants who scored thirty-seven (37) and lower are identified as receiving low MLST scores. These participants would not be selected or considered for the position.

Table 5. Frequency of participants BIDR and MSLT composite results

		MLST Result	
		Fail	Pass
BIDR Result	High	5 (28%)	2 (11%)
	Moderate	3 (16%)	2 (11%)
	Low	5 (28%)	3 (16%)

Table 6. Response Latency Categorizations

Label	Range	Statistics
Short response latencies	-1.61 to -1.79	12 (67%)
Medium response latencies	-1.41 to -1.60	4 (22%)

Long response latencies	-1.21 to -1.40	2 (11%)
-------------------------	----------------	---------

Table 7. Frequency of participants BIDR composite results and response latencies

		BIDR Result		
		Low	Moderate	High
Response Latency	Short	5 (28%)	5 (28%)	3 (16%)
	Moderate	1 (5%)	1 (5%)	2 (11%)
	Long	0 (0%)	0 (0%)	2 (11%)

Table 8. Frequency of participants Self-deception Enhancement (SDE) sub-scale results

Group	Statistics
Low SDE	9 (50%)
Moderate SDE	4 (22%)
High SDE	5 (28%)

Table 9. Frequency of participants Self-deception Enhancement (SDE) sub-scale results and response latencies

		SDE Result		
		Low	Moderate	High
Response Latency	Short	7 (38%)	4 (22%)	2 (11%)
	Moderate	2 (11%)	0 (0%)	1 (5%)
	Long	0 (0%)	1 (5%)	1 (5%)

Table 10. Frequency of participants Impression Management (IM) sub-scale results

Group	Statistics
Low IM	5 (28%)
Moderate IM	4 (22%)
High IM	9 (50%)

Table 11. Frequency of participants Impression Management (IM) sub-scale results and response latencies

		IM Result		
		Low	Moderate	High
Response Latency	Short	5 (28%)	2 (11%)	6 (33%)
	Moderate	0 (0%)	2 (11%)	2 (11%)
	Long	0 (0%)	0 (0%)	2 (11%)

Table 12. Frequency of participants MLST composite scores and response latencies

		MLST Result	
		Fail	Pass
Response Latency	Short	10 (55%)	5 (28%)
	Moderate	2 (11%)	2 (11%)
	Long	2 (11%)	0 (0%)

Table 13. Frequency of participant's response latencies, BIDR composite scores and MLST composite scores

Group	Statistics
Short response latencies, low BIDR, fail MLST	3 (16%)
Short response latencies, low BIDR, pass MLST	2 (11%)
Short response latencies, moderate BIDR, fail MLST	2 (11%)
Short response latencies, moderate BIDR, pass MLST	2 (11%)
Short response latencies, high BIDR, fail MLST	2 (11%)
Short response latencies, high BIDR, pass MLST	1 (5%)
Moderate response latencies, low BIDR, fail MLST	1 (5%)
Moderate response latencies, low BIDR, pass MLST	1 (5%)
Moderate response latencies, moderate BIDR, fail MLST	0 (0%)
Moderate response latencies, moderate BIDR, pass MLST	0 (0%)
Moderate response latencies, high BIDR, fail MLST	1 (5%)
Moderate response latencies, high BIDR, pass MLST	1 (5%)
Long response latencies, low BIDR, fail MLST	0 (0%)
Long response latencies, low BIDR, pass MLST	0 (0%)
Long response latencies, moderate BIDR, fail MLST	0 (0%)

Long response latencies, moderate BIDR, pass MLST	0 (0%)
Long response latencies, high BIDR, fail MLST	2 (11%)
Long response latencies, high BIDR, pass MLST	0 (0%)

Hypothesis 1. It is hypothesized that Management and Leadership Skills Test (MLST) composite mean scores and Balanced Inventory of Desirable Responding (BIDR) composite mean scores will be positively correlated. Respondents BIDR composite scores and BIDR sub-scale scores (Self-deceptive enhancement, and Impression Management) will be compared with the MLST composite scores and MLST scores by competency (Stakeholder and Relationship Building, People Development and Management, Change Leadership, Decisiveness, Strategic Thinking) to identify if a positive relationship exists between social desirability responding and competency.

It was hypothesized that the Management and Leadership Skills Test (MLST) composite mean scores and Balanced Inventory of Desirable Responding (BIDR) composite mean scores would be positively correlated. A bivariate correlation showed no significant relationship between MLST and BIDR composite scores with a two-tailed significant level of $r(18) = .215, p = .391$.

There was no significant correlation between decisiveness and BIDR composite, $r(18) = .190, p = .451$; between decisiveness and the BIDR IM scale, $r(18) = .105, p = .678$; or between decisiveness and the BIDR SDE scale, $r(18) = .247, p = .322$. There was no significant correlation between strategic thinking and BIDR composite, $r(18) = .013, p = .960$; between strategic thinking and the BIDR IM scale, $r(18) = .082, p = .746$; or between strategic thinking and the BIDR SDE scale, $r(18) = .101, p = .689$. There was no significant correlation between SBR and BIDR composite, $r(18) = .188, p = .456$; between SBR and the BIDR IM scale, $r(18) = .076, p = .764$; or between SBR and the BIDR SDE scale, $r(18) = .271, p = .276$. There was no significant correlation between PDM and BIDR composite, $r(18) = .225, p = .370$; between PDM and the BIDR IM scale, $r(18) = .197, p = .443$; or between PDM and the BIDR SDE scale, $r(18) = .225, p = .370$. There was no significant correlation between change leadership and BIDR composite, $r(18) = .403, p = .097$; between change leadership and the BIDR IM scale, $r(18) = .360, p = .143$; or between change leadership and the BIDR SDE scale, $r(18) = .398, p = .102$. These results

demonstrate that there is no significant correlation between social desirability and the five competencies assessed in the MLST.

Despite a non-significant correlation, a categorical analysis was conducted in order to further analyze the social desirability. To do so, the BIDR scores were categorized into three distinct groups. These groups were identified by ranking high, moderate, and low scores on the BIDR composite and sub-scales through equal intervals. Please see Tables 2 and 3. With the expectation that social desirability responding facilitates a participant's ability to increase their assessment score, it was suspected that participants with high BIDR scores will receive higher MLST scores than their low or average counterparts. As such, the BIDR and MLST results were further analyzed to identify if differences exist between the three groups categorized in Table 3 and the MLST scores. The MLST scores were categorized into two distinct groups (pass and fail) by establishing a performance requirement of sixty percent (60%) (a standard pass mark for competitive selection processes for this type of position and assessment). Please see Table 4. Based on these categorizations, thirty-eight percent (38%) of participants passed the MLST, sixty-one percent (61%) of participants had low MLST scores.

The categorical analysis of the BIDR and MLST composite scores is consistent with the correlational analyses reported above and no significant relationship was identified. Twenty-eight percent (28%) of participants had high BIDR scores and failed the MLST; twenty-eight percent (28%) had low BIDR scores and failed the MLST; sixteen percent (16%) had low BIDR results and passed the MLST, and moderate BIDR scores and failed the MLST. Eleven percent (11%) of participants had high BIDR scores and passed the MLST; and moderate BIDR scores and passed the MLST. For an overview please see Table 5.

Hypothesis 2. It is suspected that longer MLST response latencies will reflect high BIDR composite scores; as longer response latencies are associated with social desirability responding, it is hypothesized that respondents who have high BIDR scores will also have longer response latencies in the MLST thus validating social desirably responding.

It was suspected that longer response latencies on the MLST would reflect higher BIDR composite scores; as longer response latencies are associated with social desirability responding. A bivariate correlation showed a significant relationship between response latency and social desirability results with a two-tailed significant level of $r(18) = .498, p = .035$. This outcome identifies a significant relationship between participants' response latencies during the administration of the MLST and their BIDR results. A significant relationship was also identified between re-

response latency and the SDE sub-scale, $r(18) = .501, p = .034$. There was no significant relationship between response latency and the IM sub-scale, $r(18) = .433, p = .073$.

The mean baseline speed was calculated (1.67) and used to compare response latencies; differences in response latencies ranged from -1.21 to -1.79. Based on these statistics, three groups were identified in order to categorize short, medium, and long response latencies. The response latencies were rank ordered, and equal intervals were established to identify three groups: short, medium and long response latencies. Sixty-seven percent (67%) of participants were identified as having short response latencies; twenty-two percent (22%) were identified as having medium response latencies; and eleven percent (11%) were identified as having long response latencies. Please see Table 6.

Response latencies were also categorically analyzed with the BIDR results. Twenty-eight percent (28%) of participants had short response latencies and low BIDR scores; twenty-eight percent (28%) had short response latencies and moderate BIDR scores; and sixteen percent (16%) had short response latencies and high BIDR scores. Eleven percent (11%) of participants had long response latencies and high BIDR scores. Eleven percent (11%) of participants had moderate response latencies and high BIDR scores; five percent (5%) had moderate response latencies and low BIDR scores; and five percent (5%) had moderate response latencies and moderate BIDR scores. Please see Table 7.

In order to assess the significance of the BIDR results within this study, the BIDR sub-scales, self-deception Enhancement (SDE) and Impression Management (IM) were also examined. The SDE categorical analysis identified that participants did not have high SDE results within the BIDR. Fifty percent (50%) of participants had low SDE scores, twenty-two percent (22%) had moderate SDE scores, and twenty-eight percent (28%) had high SDE scores. Please see Table 8. SDE and response latency were also analyzed; however, the categorical analysis varied and no distinct groups were identified. Thirty-eight percent (38%) of participants had short response latencies and low SDE scores; twenty-two percent (22%) had short response latencies and moderate SDE scores; eleven percent (11%) had short response latencies and high SDE scores; and eleven percent (11%) had moderate response latencies and high SDE scores. Five percent (5%) of participants had low response latencies and moderate SDE scores, and five percent (5%) had low response latencies and high SDE scores. Please see Table 9.

A similar categorical analysis was conducted for the BIDR sub-scale IM. Twenty-eight percent (28%) of participants had low IM scores, twenty-two percent (22%) had moderate IM

scores, and fifty percent (50%) of participants had high IM scores. Please see Table 10. Overall, participants were more likely to respond with impression management tendencies than self-deceptive tendencies.

IM and response latency was also analyzed; in contrast to the results for the BIDR, participants were more likely to demonstrate impression management tendencies during the administration of the MLST, despite having shorter response latencies. Thirty-three percent (33%) of participants had short response latencies and high IM scores. Twenty-eight percent (28%) of participants had short response latencies and low IM scores; twenty-two percent (22%) had short response latencies and moderate IM scores; twenty-two percent (22%) had moderate response latencies and high IM scores; twenty-two percent (22%) had moderate response latencies and moderate IM scores; eleven percent (1%) had long response latencies and high IM scores. Please see Table 11.

Hypothesis 3. It is hypothesized that there will be a correlation between the MLST, the BIDR and response latency for each participant.

It was hypothesized that there would be a correlation between the MLST competencies, the BIDR composite score and response latencies for each participant. A regression analysis was used to identify if BIDR scores and response latencies were able to predict MLST results. As such, the MLST was set as the dependent variable, and the BIDR results and response latencies were set as the independent variable (predictor). There was no significant correlation identified, $F(18) = .996$, $p = .393$. As such, response latency ($\beta = -.17$) and the BIDR ($\beta = -.01$) were not significant predictors of MLST results. The model was $R^2 = 0.12$ and was not significant. This outcome indicates that the results generated from BIDR results and response latencies were not conducive to predicting MLST scores.

Despite this outcome, the MLST, BIDR and response latencies were categorically analyzed to identify other relationships. MLST scores and response latency were categorically analyzed and fifty-five percent (55%) of respondents had short response latencies and failed the MLST; eleven percent (11%) of participants had moderate response latencies and failed the MLST; eleven percent (11%) had moderate response latencies and passed the MLST; and eleven percent (11%) of participants had long response latencies and failed the MLST. Please see Table 12.

Aligned with the correlational analysis, the data is inconclusive in relation to a multi-dimensional comparison: response latency, social desirability, competency results. Please see Ta-

ble 13. For participants who received medium response latency times, fifty percent (50%) had high BIDR results and passed the MLST.

Discussion of the Results

Situational judgment tests (SJTs) have emerged as the assessment of choice in organizational selection contexts; however, the effects of social desirability responding may threaten their validity. Research indicates that social desirability responding increases the likelihood of participants being selected in a competitive selection process (Berry & Sackett, 2006) and as a result, there are significant effects on participants who are selected for positions (Rosse et al., 1998; Weiner and Gibson, 2000; Whetzel & McDaniel, 2009). With this in mind, response latency models identify and predict social desirability responding; Paulhus and Holden (2009) have introduced the concept of self-schemas and their relationship to calculating response latencies. This concept is the basis for the argument that longer response latencies are associated with social desirability responding (McDaniel and Timm, 1990; Follette, 1984; Holden et al., 2001). However, further research is required to further validate this claim as conflicting findings demonstrate that social desirability responding may also be associated with faster response latencies (Hsu, Santelli and Hsu, 1989; George, 1990).

In response to these academic requirements, the primary objective of this research was to explore the extent in which participants distort their responses in a socially desirable way during a competitive selection process with the intent to increase their scores and be selected for a position. This research also aimed to identify the relationship between response latency and social desirability. The theoretical framework proposed on page 24 is no longer appropriate as a result of this research. Response latency should be connected to the BIDR as it has been identified as a predictor for social desirability responding and an additional connector should be included for response latency and the BIDR subscale self-deceptive enhancement. The arrow connecting social desirability and the MLST should be removed as this prediction was not supported.

Hypothesis 1. There was no significant relationship identified between the Management and Leadership Skills Test (MLST) and the Balanced Inventory of Desirable Responding (BIDR). As a result, no significant inferences can be articulated that identify social desirability as influencing the outcome of the MLST. This can be further supported by the percentage of participants who had high BIDR results yet failed the MLST. A similar outcome was identified for the

MLST competency sub-scale scores and the BIDR sub-scale scores in which there was no significant correlation. The results of this study demonstrate that social desirability responding does occur during competitive selection processes, but that despite socially desirable tendencies, there were no significant effects on an individual's ability to positively direct the outcome of the MLST.

These results could be attributed to several factors. In order to identify a significant correlation, a larger sample size (such as the one used in Paulhus, 1991) may increase the probability of identifying significant effects between social desirability and competency results. Research by Quinn (1989) and Paulhus (1988) used sample sizes ranging from 400-900 participants; the significance of their research can be partially attributed to the size of their population. Even a smaller sample size of forty-eight (48) participants by Mellor, Conroy, and Masteller (1986) produced significant results. The quality of the sample does not appear to be an issue as the process was standardized, and the sample was representative of the population required for this scope of research. Additionally, participants had similar motives; to be selected for an executive level position.

The BIDR composite results indicate that respondents participated in social desirability responding; however, this outcome did not prove to be statistically significant. Thirty-nine percent (39%) of participants were identified as high BIDR respondents based on their composite BIDR scores. Interestingly, sixty-one percent (61%) of this group failed the MLST based on a pass score of 60%. These results suggest that responding in a socially desirable manner did not increase the scores of participants as identified in the literature review. This outcome may be attributed to the robust development and administration of the MLST, as psychometric standards and best practices were integrated into the design based on scientific literature. In other words, the design of the assessment may have hindered the effects of social desirability responding on the MLST.

The BIDR was significantly correlated with the IM and SDE scales which signifies the validity of the BIDR and the two sub-scales. These statistics confirm the significant correlation between the two scales and the inventory as a whole, and indicates that the inventory is functioning as expected. As such, the results derived from the inventory can be identified as valid representations of social desirability responding. Similar results were found with the MLST composite score and the five sub-scales in which all correlations proved to be significant. These significant correlations between the MLST composite score and the five sub-scales demonstrate that the test

was well constructed and that the situational based questions were aligned with the test as a whole. In so, one can conclude that the test was a valid measure.

The mean for the BIDR sub-scale SDE and IM mean were higher than what was found with Paulhus (1989) and Quinn (1989) for both sub-scales. This difference could be attributed to the sample, as both researches enlisted participants outside of a competitive organizational context (students and religious adults). Research shows that social desirability results are higher for participants competitively applying for a position compared to non-applicants or incumbents (Hough, 1998; Rosse et al., 1998; Hooper & Sackett, 2008) such as the results presented by Mellor, Conroy and Masteller (1986) who reported a mean of 11.2 ($sd = 4.9$) for IM. These results demonstrate that in the context of applying for a competitive selection process, participants are motivated to engage in focused and intentional efforts to increase their score. However, as stated above, efforts to respond in a socially desirable way did not appear to significantly affect the results obtained on the MLST.

Participants were categorized depending on their BIDR composite and sub-scale scores. Thirty-eight percent (38%) of participants had high BIDR scores; thirty-three percent (33%) had low BIDR scores; and twenty-seven percent (27%) had moderate BIDR scores. Through an analysis of the BIDR composite scores, the distribution across groups is fairly equal with similar variance between groups. These results also demonstrate that social desirability responding does occur during competitive selection processes; however, as mentioned above there were no significant impacts caused by this behavior.

The BIDR sub-scale IM yielded higher socially desirable tendencies; fifty percent (50%) of participants had high IM scores while SDE scores reflected considerably lower tendencies; twenty-eight percent (28%) had low SDE scores. Twenty-two percent (22%) of participants received moderate IM and SDE scores. The BIDR IM sub-scale group identified the highest tendency to respond in a socially desirable manner; the SDE subscale identified the lowest scores. These results could be attributed towards the nature of the indexes. Impression management (IM), similar to Marlowe and Crowne's need for approval, demonstrates a tendency towards positive self-descriptions to an audience; in other words, the presentation of a socially desirable image. Self-deceptive enhancement (SDE) represents a tendency to give positively biased reports that are viewed as a "non-conscious inclination to perceive oneself favorably," (Hart, et al. 2015). As participants are applying for a competitive selection process, the ability to present a socially desirable image is amplified. When considering the behavior presented during a selection process,

participants want to present and be viewed as the best applicant. In doing so, they must consciously and publically voice appropriate behaviors and work-related abilities; this behavior is aligned with the IM sub-scale results.

Hypothesis 2. It was suspected that longer MLST response latencies would reflect high BIDR composite scores. As longer response latencies are associated with social desirability responding, it was hypothesized that respondents who have high BIDR scores will also have longer response latencies in the MLST; this hypothesis was supported. A significant correlation was identified between response latency and social desirability results. This outcome identifies a significant relationship between participants' response latencies during the administration of the MLST and their BIDR results. Significant relationships were also identified with response latency and the SDE sub-scale.

Participants who received short response latencies were less likely to participate in social desirability responding during the administration of the BIDR. Concurrently, participants who had longer response latencies received higher BIDR scores. These results demonstrate that shorter response latencies are correlated with lower social desirability scores; in particular for the BIDR sub-scale Self-Deceptive Enhancement (SDE). These results contribute to the academic research that advocates for the use of response latency as a predictor of social desirability (McDaniel and Timm, 1990; Follette, 1984; Holden et al., 2001).

Sixty-seven percent (67%) of participants were identified as having short response latencies. As such, the majority of participants had shorter response latencies compared to the computed baseline speed. Twenty-two percent (22%) of participants were identified as moderate respondents, and eleven percent (11%) of participants were identified as having longer response latencies.

Fifty percent (50%) of participants had low SDE scores, twenty-two percent (22%) had moderate SDE scores, and twenty-seven percent (27%) had high SDE scores. Overall, participants did not have high SDE results within the BIDR. Thirty-eight percent (38%) of participants who had short response latencies also received low SDE scores. This statistic is similar to the BIDR analysis above - twenty-eight percent (28%) of participants had short response latencies and low BIDR scores. However, the strength of the relationship appears to be more significant for the sub-scale SDE by 0.10. This result also reiterates the hypothesis that shorter response latencies are correlated with lower social desirability scores, and in this case, the sub-scale SDE.

Thirty-three percent (33%) of participants had short response latencies and high IM scores. In contrast to the results for the BIDR, participants were more likely to demonstrate impression management tendencies, despite having shorter response latencies. Twenty-seven percent (27%) of participants had had short response latencies and low IM scores. Eleven percent (11%) of participants who had long response latencies and high IM also failed the MLST.

Gender and age statistics were analyzed and no significant relationships were identified. Thirty-three percent (33%) of women passed the MLST, and 33% failed. Forty-two percent (42%) of men passed the MLST. These findings are slightly different than found in the literature, as females tend to score higher than males on SJTs (O'Connell et al., 2007; Patterson et al., 2012). For more information, please see Appendix E.

Hypothesis 3. It was hypothesized that there would be a correlation between the MLST, the BIDR and response latency for each participant; this hypothesis was not supported. There was no significant correlation identified; this outcome suggests that the results generated from BIDR results and response latencies were not conducive to predicting MLST scores.

Eleven percent (11%) of participants who had longer response latencies received high BIDR scores, failed the MLST. This percentage of individuals used more time to respond in the MLST and participated in social desirability responding in the BIDR; this behavior is concurrent with hypothesis 2. It could therefore be implied that these individuals had longer response latencies because the information was not consistent with their cognitive structure, and the information was accepted more slowly (Paulhus and Holden, 2009). This outcome also suggests that the high social desirability scores on the BIDR are associated with the analysis and selection of socially accepted behavior, rather than honest responses. Lastly, these results demonstrate that despite participating in socially desirable responding, the participants were unable to influence their scores in a favorable way on the MLST.

Forty-four percent (44%) of respondents had short response latencies and failed the MLST. It could be argued that participants should have taken more consideration in selecting an appropriate response, and that their fast analysis and selection hindered their ability to receive higher results. Further research should be directed at identifying a range for appropriate latency analysis and the threshold of latency that correlates with social desirability responding. Twenty-eight percent (28%) of respondents had short response latencies and passed the MLST. It appears for a percentage of the individuals, shorter response latencies do not hinder their analysis, selection and performance on the MLST.

For participants who received moderate response latency times, fifty percent (50%) had high BIDR results and passed the MLST. Forty-four percent (44%) of participants received low BIDR results, low competency scores for strategic thinking, decisiveness and stakeholder building, and short response latencies. The results of this research suggest that moderate social desirability responding and short response latencies were conducive to higher results across four of the five competencies. It also demonstrates that short response latencies and low social desirability tendencies were not conducive to high competency results for three of the five competencies. Two of these three competencies are strategic thinking and decisiveness, both of which are correlated to cognitive ability.

The five (5) MLST competencies were analyzed to identify if a relationship existed between competency scores, social desirability and response latency. Fifty-six percent (56%) of participants across all five competencies received moderate BIDR scores, high competency scores, and short response latencies. This outcome suggests that moderate social desirability responding and fast response times were conducive to higher results across four of the five competencies. Similarly, sixteen percent (16%) of participants had moderate BIDR scores, moderate competency scores, and short response latencies.

Summary

There was no significant relationship identified between the Management and Leadership Skills Test (MLST) and the Balanced Inventory of Desirable Responding (BIDR). The results of this study demonstrates that social desirability responding does occur during competitive selection processes, but that despite socially desirable tendencies, there were no significant effects on a participant's ability to positively direct the outcome of the MLST. As such, this research was able to demonstrate that social desirability responding occurs in competitive selection processes (Bikerland, et al., 2006), but could not attest to the threat social desirability has on the validity of the assessment (Holden, Wood & Tomashewski, 2001; Heggstad, et al., 2006; Douglas et al., 1996; Dunnette et al., 1962; Kluger, Reilly & Russell, 1991, Schmit & Ryan, 1992; Zickar, 1997).

It was suspected that longer MLST response latencies would reflect high BIDR composite scores; as longer response latencies are associated with social desirability responding. Participants who had short response latencies were less likely to participate in social desirability re-

sponding during the administration of the BIDR. Concurrently, participants who had longer response latencies received high BIDR scores. These results also identify that shorter response latencies are correlated with lower social desirability scores; in particular for the BIDR sub-scale Self-Deceptive Enhancement (SDE). These results contribute to the academic research that advocates for the use of response latency as a predictor of social desirability (McDaniel and Timm, 1990; Follette, 1984; Holden et al., 2001) and the suggestion that longer response latencies are associated with social desirability responding.

It was hypothesized that there would be a correlation between the MLST, the BIDR and response latency; this hypothesis was not supported. The results of this research suggest that moderate social desirability responding and short response latencies were conducive to higher results across four of the five competencies. It also demonstrates that short response latencies and low social desirability tendencies, were not conducive to high competency results for three of the five competencies.

The two assessments administered proved to function appropriately. The BIDR was significantly correlated with the Impression Management (IM) scale, the Self-Deceptive Enhancement (SDE) scale, and the MLST composite score was significantly correlated with the five subscales. The development of the MLST could have attributed to low BIDR scores as the assessment is designed and administered alongside best practices and standards; such as a *most and least* effective response instruction set (McDaniel & Nguyen 2001), and standardized administration.

Contributions

The theoretical basis for situational judgment research is lacking and there is a scientific requirement to include real-applicants in research. A current limitation of SJT research is the use of concurrent designs in which incumbents, such as students or volunteers, who are enlisted to participate. Although these types of studies produce significant data and outcomes, the academic community recognizes that social desirability responding in organizational contexts is most appropriately assessed with real-applicants (McDaniel & Whetzel, 2009). In response, this research includes an appropriate population and assists in the prediction and application of research outcomes. Recognizing this requirement, this research assessed real applicants who are applying for a competitive position within an organizational context. As such, the results identified in this study contribute to the application and administration of psychometric assessments, and the im-

plications of social desirability and response latency. This research also provides further insight into the use of response latency as a method to predict and identify social desirability responding and the use of technology to collect and store large data sets related to response latencies. This research also contributes to the strengthening and advancement of psychometric assessments, specifically through the use of a forced choice format, introductory warnings in regards to response distortion, and most and least effective behavior prompting; all of which are identified as assessment best practices.

Limitations

There are several limitations of this research. The first limitation was the sample size; in order to identify a significant correlation, a larger sample size (such as the one used in Paulhus, 1991) may increase the probability of identifying significant relationships between social desirability and competency results. A larger sample size would also increase the reliability of the results and be more reflective of the population. The second limitation is the structure of the assessment and the inclusion of multi-dimensional constructs. Psychometric assessments provide more analytical depth when they consist of unidimensional constructs (Osam, 2014). As a result, it is more difficult to target specific differences across the competencies (McDaniel & Nguyen, 2001) and identify direct relationships.

Further research is required to identify the range in which response latencies correlate with social desirability. Additional research that clearly identifies response latency ranges (beyond short, medium and long), and their respective predictions for social desirability scores is warranted in SJT research. The outcome of this research could assist in the generation of key predictor variables that are valid and reliable across different SJT applications.

A third limitation of this study is the MLST, as the assessment has not been validated in concurrent or organizational contexts. Although there are significant correlations between the MLST and the sub-scales, there is no evidence to suggest that the assessment measures what it is intended to measure. For further research, a validity study should be conducted in order to identify that the assessment is a valid measure.

Conclusion

In conclusion, this research demonstrates that there is a significant correlation between social desirability responding and response latencies, and identifies the relationships, and lack thereof, between social desirability responding, response latencies, and competency based assessments. This research demonstrates that participants do respond in a socially desirable manner; however, this behavior does not have a significant effect on the outcome of the assessment. This outcome may provide insight into how individuals present and adapt in different contexts, such as social and professional situations. As current research is convoluted, perhaps the question should be modified from, *do participants fake?* (which appears to be both unproductive and lacking empirical evidence) to *whom and when do participants fake?* This shift in academia could direct new research that advances and strengthens the use of psychometric assessments and the possible effects of social desirability responding. This research provides justification for the use of response latencies as a detection measure for social desirability and suggests that longer response latencies are associated with social desirability responding. The ability to predict and measure social desirability through response latencies would revolutionize and simplify the way the academics and professionals assess, compare and select participants for competitive selection processes. Together, this theme may also identify a range in which social desirability is not a construct that should be viewed as positively or negatively, rather on a spectrum in which socially desirable behavior is an integral and adaptive behavior that is required, and perhaps, even encouraged.

References

- Abrahams, N. M., Neumann, I., & Githens, W. H. (1971). Faking vocational interests: Simulated versus real life motivation. *Personnel Psychology*, 24, 5–12.
- Arthur, W., & Villado, A. J. (2008). The importance of distinguishing between constructs and methods when comparing predictors in personnel selection research and practice. *Journal of Applied Psychology*, 93, 435–442.
- Bagozzi, R. P., Yi, Y., & Phillips, L. W. (1991). Assessing construct validity in organizational research. *Administrative Science Quarterly*, 421-458.

- Berry, C. M., Sackett, P. R. (2009). Faking in Personnel Selection: Tradeoffs in performances versus fairness resulting from two cut-score strategies. *Personnel Psychology*, 62, 4, 833-863.
- Bikerland, S. A., Manson, T. M., Kisamore, J. L., Brannick, M. T., & Smith, M. A. (2006). A Meta-Analytical Investigation of Job Applicant Faking on Personality Measures. *International Journal of Selection and Assessment*, 14, 4, 317-335.
- Bobat, S., Caruth, E. J., Buitendach, J. H. (2012). Relationship Between a Situational Judgement Test and Ability and Personality Measures. *Journal of Psychology in Africa*, 22, 245-250.
- Brualdi, A. (1999). Traditional and modern concepts of validity. Washington, DC: ERIC Clearinghouse on Assessment and Evaluation. 2-6.
- Chan, D. (2006). Interactive effects of situational judgment effectiveness and proactive personality on work perceptions and work outcomes. *Journal of Applied Psychology*, 91, 275–281.
- 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. P. & Haaland, D. E. (2000). Examining the relationship between job knowledge and situational judgment performance. Paper presented at the 15th Annual.
- Cullen, M. J., Sackett, P. R., & Lievens, F. P. (2006). Threats to the operational use of situational judgment tests in the college admission process. *International Journal of Selection and Assessment*, 14, 142–155.
- Dwight, S.A. (1990). An Assessment of the difference effects of warning applicants not to fake. Unpublished doctoral dissertation, State University of New York at Albany.

- Dye, D. A., Reck, M. and McDaniel, M. A. (1993). Moderators of the validity of write job knowledge measures. *International Journal of Selection and Assessment*, 1, 153-157.
- Douglas, E. F., McDaniel, M. A., & Snell, A. F. (1996, August). The validity of non-cognitive measures decays when applicants fake. Paper presented at the annual conference of the Academy of Management, Cincinnati, OH.
- Dubois, P. H. (1970). *A history of psychological testing*. Boston: Allyn & Bacon, 148.
- Dunnette, M. D., McCartney, J., Carlson, H. C., & Kirchner, W. K. (1962). A study of faking behaviour on a forced-choice self-description checklist. *Personnel Psychology*, 15, 13–24.
- Edwards, A. L. (1970). *The measurement of personality traits by scales and inventories*. Oxford, England: Holt, Rinehart & Winston.
- Ellingson, J. E., Smith, D. B., & Sackett, P. R. (2001). Investigating the influence of social desirability on personality factor structure. *Journal of Applied Psychology*, 86, 122–133.
- Follete, W. C. (1984). *A computer administered MMPI and a study of response latency and social desirability*. Unpublished doctoral dissertation, University of Washington.
- George, M. S. (1990). *The use of response latency to study accuracy and consistency in a computerized lifestyle assessment*. Unpublished doctoral dissertation, University of Toronto.
- Hart, C. M., Ritchie, T. D., Hepper, E.G., Gebauer, J.E. (2015). The Balanced Inventory of Desirable Responding Short Form (BIDR-16). *Sage Open*, 5, 1-9.
- Heggestad, E. D., Morrison, M., Reeve, C. L. and McCloy, R. A. (2006). Forced-Choice Assessments of Personality for Selection: Evaluating Issues of Normative Assessment and Faking Resistance. *Journal of Applied Psychology*, 91, 9-24.

- Holden, R. R., Fekken, G. C., & Cotton, D. H. G. (1991). Assessing psychopathology using structured test-item response latencies. *Psychological Assessment*, 3, 111–118.
- Holden, R. R., & Hibbs, N. (1995). Incremental validity of response latencies for detecting fakers on a personality test. *Journal of Research in Personality*, 29, 362–372.
- Holden, R. R., Kroner, D. G., Fekken, G. C., & Popham, S. M. (1992). A model of personality test item response dissimulation. *Journal of Personality and Social Psychology*, 63, 272–279.
- Holden, R. R., Wood, L. L., Tomashewski, L. (2001). Do Response Time Limitations Counteract the Effect of Faking on Personality Inventory Validity? *Journal of Personality and Social Psychology*, 81, 160-169.
- Hooper AC, Sackett PR. (2008, April). Self-presentation on personality measures: A meta-analysis. In Peterson MH (Chair), Examining faking using within-subjects designs and applicant data. Symposium presented at 23rd Annual Conference of the Society for Industrial and Organizational Psychology, San Francisco, CA.
- Hsu, L. M., Santelli, J., & Hsu, J. R. (1989). Faking detection validity and incremental validity of response latencies to MMPI subtle and obvious items. *Journal of Personality Assessment*, 53, 278-295.
- Juraska and Drasgow, F. (2001). Faking in situational judgment test. Paper presented at the 16th Annual Conference of the Society for Industrial and Organizational Psychology, San Diego, CA.
- Kluger, A. N., Reilly, R. R., & Russell, C. J. (1991). Faking biodata tests: Are option-keyed instruments more resistant? *Journal of Applied Psychology*, 76, 889–896.

- Lievens, F., Peeters, H., & Schollaert, E. (2008). Situational judgment tests: A review of recent research. *Personnel Review*, 37, 426–441.
- Lievens, F., Sackett, P. R. (2006). Video-Based Versus Written Situational Judgment Tests: A Comparison in Terms of Predictive Validity. *Journal of Applied Psychology*, 91, 5, 1181-1188.
- Zieky, M.J. & Perie, M. (2006). *A Primer on Setting Cut Scores on Tests of Educational Achievement*. ETS Publication. Retrieved November 20, 2016.
- Marentette, B. J., Meyers, L. S., Hurtz, G. M., & Kuang, D. C. (2012). Order effects on situational judgment test items: A case of construct- irrelevant difficulty. *International Journal of Selection and Assessment*, 20, 319–332.
- Mayerl, J. (2013). Response Latency Measurement in Surveys. Detecting Strong Attitudes and Response Effects. *Survey Methods: Insights from the Field*. Retrieved from <http://surveyinsights.org/?p=1063>
- McCrae, R. R., & Costa, P. T. (1983). Social desirability scales: More substance than style. *Journal of Consulting and Clinical Psychology*, 51, 882–888.
- McDaniel, M. A., Hartman, N. S., Whetzel, D. L., & Grubb, W. L., III. (2007). Situational judgment tests, response instructions, and validity: A meta-analysis. *Personnel Psychology*, 60, 63–91.
- 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., & Timm, H. W. (1990). Lying takes time: Predicting deception in biodata using response latency. Paper presented at the 98th Annual Convention of the American Psychological Association, Boston.
- McFarland, L.A. & Ryan, A. M. (2000). Variance in faking across non-cognitive measures. *Journal of Applied Psychology*, 85, 812-821.
- Mehrens, W.A. (1997). The Consequences of Consequential Validity. *Educational Measurement: Issues and Practice*, 16, (2), 16-18.
- Mellor, S., Controy, L., & Masteller, M. K. (1986). Comparative trait analysis of long term recovering alcoholics. *Psychological reports*, 58, 411-418.
- Messick, S., (1995), Standards of Validity and the Validity of Standards in Performance Assessment. *Educational Measurement: Issues and Practice*, 14, (4), 5-8.
- Motowidlo, S. J., & Beier, M. E. (2010). Differentiating specific job knowledge from implicit trait policies in procedural knowledge measured by a situational judgment test. *Journal of Applied Psychology*, 95, 321-333.
- 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.
- Motowidlo, S. J., Hanson, M. A., & Crafts, J. L. (1997). Low-fidelity simulations. In D. L. Whetzel and G. R. Wheaton (Eds.), *Applied Measurement Methods in Industrial Psychology*. Pali Alto, CA: Davies-Black.
- Motowidlo, S. J., Brownlee, A. L., & Schmit, M. J. (2008). Effects of personality characteristics on knowledge, skill, and performance in servicing retail customers. *International Journal of Selection and Assessment*, 16, 272-281.

- Motowidlo, S. J., & Tippins, N. (1993). Further studies of the low-fidelity simulation in the form of a situational inventory. *Journal of Occupational and Organizational Psychology*, 66, 337–344.
- Mullins, M. E. and Schmitt, N. (1998). Situational judgement testing: Will the real constructs please present themselves? Paper presented at the 13th annual conference of the Society for Industrial and Organizational Psychology, Dallas, Texas.
- Mumford, M. D. and Whetzel, D. L. (1997). Background data. In D. L. Whetzel and G. R. Wheaton (Eds.), *Applied Measurement Methods in Industrial Psychology*. Palo Alto, CA: Davies-Black.
- Nguyen, N. T., McDaniel, M. A., Biderman, M. D. (2005). Effects of Response Instructions on Faking a Situational Judgement Test. *International Journal of Selection and Assessment*, 13, 250-260.
- Nowakowska, M. (1970). A model of answering to a question item. *Acta Psychologica*, 24, 420-439.
- O’Connell, M.S., Hartman, N.S., McDaniel, M.A., Grubb, W.L., & Lawrence, A. (2007). Incremental validity of situational judgment tests for task and contextual job performance. *International Journal of Selection and Assessment*, 15, 19-29.
- O'Connell, M. S., McDaniel, M. A., Grubb, W. L. III, Hartman, N. S., & Lawrence, A. (2002, April). Incremental validity of situational judgment tests for task and contextual performance. Paper presented at the 17th annual conference of the Society of Industrial and Organizational Psychology, Toronto, Canada.
- Oostrom, J. K., Born, M., Serlie, A. W., & van der Molen, H. T. (2011). A Multi-media Situational Judgement Test With a Constructed-Response Format: Its Relationship With Personality, Cognitive Ability, Job Experience, and Academic Performance. *Journal of Personnel Psychology*, 10, 78-88.

- Osam, E. K. (2014). The Adaption of a Situational Judgement Test to Measure Leadership Knowledge in the Workplace. Masters These and Specialist Projects. Western Kentucky University.
- Patterson, F., Ashworth, V., Zibarras, L., Coan, P., Kerrin, M., & O'Neil, P. (2012). Evaluations of situational judgement tests to assess non-academic attributes in selection. *Medical Education*, 46, 850-86.
- Paulhus, D. L., Bruce, M. N., & Trapnell, P. D. (1995). Effects of Self-Presentation Strategies on Personality Profiles and Their Structure. *Personality and Social Psychology Bulletin*, 21, 2, 100-108.
- Paulhus, D. L. (1991). Measurement control of response bias. In J.P. Robinson, P. R. Shaver, & L. S. Wiggins (Eds.), *Measures of Personality and social psychological attitudes* (pp. 17-60). San Diego: Academic Press.
- Paulhus, D. L., & Holden, R. R. (2009). *Measuring Self-Enhancement: From Self-Report to Concrete Behaviour. Then a Miracle Occurs*. Oxford University Press. 221-240.
- Peeters, H., & Lievens, F. (2005). Situational judgment tests and their predictiveness of college students' success: The influence of faking. *Educational and Psychological Measurement*, 65, 70–89.
- 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.
- Rockstuhl, T., Ang, S., Ng, K. Y., Lievens, F., & Van Dyne, L. (2015). Putting judging situations into situational judgment tests: Evidence from intercultural multimedia SJTs. *Journal of Applied Psychology*, 100, 464–480.

- Rosse, J. G., Stechner, M.D., Levin, R.A., & Miller, J. L. (1998). The impact of response distortion of reemployment personality testing and hiring decisions. *Journal of Management*, 26, 565-606.
- Rynes, S. L., & Connerley, M. L. (1993). Applicant reactions to alternative selection procedures. *Journal of Business and Psychology*, 7, 261-277.
- Salgado, J. F., Viswesvaran, C., & Ones, D. S. (2001). Predictors used for personnel selection: An overview of constructs, methods, and techniques. In N. R. Anderson, D. S. Ones, H. K. Sinangil & C. Viswesvaran (Eds.), *Handbook of industrial, work, & organizational psychology*, Vol. 1. (pp. 165–199) London: Sage.
- Schmitt, N., Oswald, F.L. (2006). The Impact of Corrections for Faking on the Validity of Non-cognitive Measures in Selection Settings. *Journal of Applied Psychology*, 91, 3, 613-621.
- Schmit, M. J., & Ryan, A. M. (1992). Test-taking dispositions: A missing link? *Journal of Applied Psychology*, 77, 629–637.
- Sackeim, H. A., & Gur, R. C. (1978). Self-deception, self-confrontation, and consciousness. In G. E. Schwartz & D. Shapiro (Eds.) *Consciousness and self-regulation: Advances in research* (Vol. 2, pp.139-197). New York: Plenum Press.
- Smith, K.C., & McDaniel, M.A. (1998, April). Criterion and construct validity evidence for a situational judgment measure. Paper presented at the 13th annual conference of the Society for Industrial and Organizational Psychology, Inc., Dallas, TX.
- Society for Industrial and Organizational Psychology, Inc. (2003). *Principles for the validation and use of personnel selection procedures*, 4th Edition Bowling Green, OH: Author.
- Thornton, G. C. and Cleveland, J. N. (1990). Developing managerial talent through simulation. *American Psychologist*, 45, 190-199.

- Vasilopoulos, N.L., Reilly, R.R. and Leamna, 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.
- Viswesvaran, C. and Ones, D.S. (1999). Meta-analyses of faking estimates: Implications for personality measurement. *Educational and Psychological Measurement*, 59, 197-210.
- Weekley, J. A., & Jones, C. (1997). Video-based situational testing. *Personnel Psychology*, 50, 25-49.
- Weekley, J. A., & Jones, C. (1999). Further studies of situational tests. *Personnel Psychology*, 52, 679-700.
- Weekley, J.A., & Ployhart, R.E. (2006). An introduction to situational judgement testing. In J.A. Weekley & R. E. Ployhart (Eds.), *Situational Judgement Tests*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Weiner, J. A., & Gibson, W. M. (2000). Practical effects of faking on job applicant attitude test scores. Paper presented at 15th annual conference of the Society for Industrial and Organizational Psychology, New Orleans, LA.
- Wernimont, P. F., & Campbell, J. P. (1986). Signs, samples and criteria. *Journal of Applied Psychology*, 52, 372-376.
- Whetzel, D. L., & McDaniel, M. A. (2009). Situational Judgement Tests: An overview of current research. *Human Resource Management Review*, 19, 188-202.
- Whetzel, D. L., McDaniel, M. A., & Nguyen, N. T. (2008). Subgroup differences in situational judgment test performance: A meta-analysis. *Human Performance*, 21, 291-309.

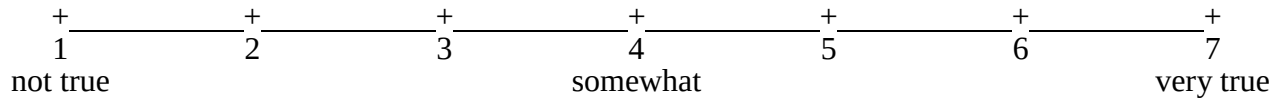
Wiggins, J. S. (1959). Interrelationships among MMPI measures of dissimulation under standard and social desirability instructions. *Journal of Consulting Psychology*, 23, 419-427.

Zickar, M.J. (1997, April). Computer simulation of faking on a personality test. Paper presented at the 12th annual conference of the Society for Industrial and Organizational Psychology, St. Louis, MO.

Appendix A

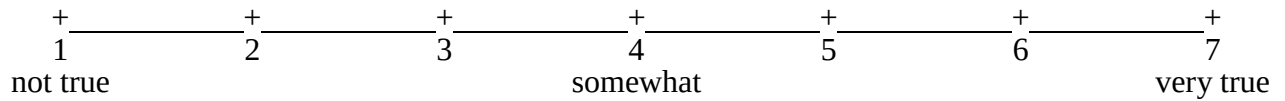
Balanced Inventory of Desirable Responding (BIDR), Seventh Edition.

Using the scale below as a guide, write a number beside each statement to indicate how true it is.



- ___ 1. My first impressions of people usually turn out to be right.
- ___ 2. It would be hard for me to break any of my bad habits.
- ___ 3. I don't care to know what other people really think of me.
- ___ 4. I have not always been honest with myself.
- ___ 5. I always know why I like things.
- ___ 6. When my emotions are aroused, it biases my thinking.
- ___ 7. Once I've made up my mind, other people can seldom change my opinion.
- ___ 8. I am not a safe driver when I exceed the speed limit.
- ___ 9. I am fully in control of my own fate.
- ___ 10. It's hard for me to shut off a disturbing thought.
- ___ 11. I never regret my decisions.
- ___ 12. I sometimes lose out on things because I can't make up my mind soon enough
- ___ 13. The reason I vote is because my vote can make a difference.
- ___ 14. My parents were not always fair when they punished me.
- ___ 15. I am a completely rational person.
- ___ 16. I rarely appreciate criticism.
- ___ 17. I am very confident of my judgments
- ___ 18. I have sometimes doubted my ability as a lover.
- ___ 19. It's all right with me if some people happen to dislike me.
- ___ 20. I don't always know the reasons why I do the things I do.
- ___ 21. I sometimes tell lies if I have to.
- ___ 22. I never cover up my mistakes.
- ___ 23. There have been occasions when I have taken advantage of someone.
- ___ 24. I never swear.

Using the scale below as a guide, write a number beside each statement to indicate how true it is.



- ___ 25. I sometimes try to get even rather than forgive and forget.
- ___ 26. I always obey laws, even if I'm unlikely to get caught.
- ___ 27. I have said something bad about a friend behind his/her back.
- ___ 28. When I hear people talking privately, I avoid listening.
- ___ 29. I have received too much change from a salesperson without telling him or her
- ___ 30. I always declare everything at customs.
- ___ 31. When I was young I sometimes stole things.
- ___ 32. I have never dropped litter on the street.
- ___ 33. I sometimes drive faster than the speed limit.
- ___ 34. I never read sexy books or magazines.
- ___ 35. I have done things that I don't tell other people about.
- ___ 36. I never take things that don't belong to me.
- ___ 37. I have taken sick-leave from work or school even though I wasn't really sick.
- ___ 38. I have never damaged a library book or store merchandise without reporting it.
- ___ 39. I have some pretty awful habits.
- ___ 40. I don't gossip about other people's business.

Appendix B

Management and Leadership Skills Test (MLST) Test Development

In order to develop the Management and Leadership Skills Test (MLST), fourteen work descriptions were analyzed. The work descriptions focused on director level positions within the organization and included descriptions of financial advisory services, operation and assessment, science and business, corporate planning, and food inspection positions. Each work description was analyzed and pertinent, work related elements were identified per description. Through the analysis of the organizational context, the scope of work, and the responsibilities, twelve common trends were identified across the fourteen work descriptions.

The EPSI development team consulted with the organization's hiring management team to review the twelve common trends and the overall analysis of the work descriptions. The discussion confirmed the common trends identified and as a result, six essential competencies were identified. From these trends, the essential competencies were derived from the organization's Leadership Competency Dictionary which include the standard agency-wide competencies required for EX level positions. Of the six competencies deemed essential, five were selected as pertinent to the position – Strategic Thinking, Decisiveness, Change Leadership, People Development and Management, and Stakeholder and Relationship Building. For a description of each competency and their behavioural indicators, please view Appendix C.

The EPSI development team analyzed the behavioural indicators provided for the five competencies and created a Critical Incident Report. A Critical Incident Report provides a standardized way of interviewing subject matter experts and top-performers on essential competencies. Subject matter experts and top-performers were identified by the organization's management team as having in-depth experience and expertise at the EX-01 level.

The Critical Incident Report asked participants to identify the key responsibilities, obstacles, and challenges that a Director may face on a daily basis based on their own experiences. It also asked participants to give an example of past experiences in which they demonstrated the competencies being assessed. Participants were asked to describe the context of the situation, its relation to the competency being assessed, and the actions they took. Each competency was assessed separately, for example:

Can you provide us with an example when you had to face a difficult decision and you had to make an efficient and effective decision?

- a) What was the context?
- b) What was the difficult decision that you faced?
- c) What actions did you take? What were the results?
- d) What other actions could have been taken, but would have not been as effective?

As a result, six subject matter experts and top-performers completed the Critical Incident Report. Based on these Critical Incident Reports, the development team was able to identify relevant, on-the-job situations that are related to the competencies being assessed. The development team edited these responses to reflect a consistent and precise format while maintaining the essence of the situational context. Information provided about the results and consequences of the behaviour were edited out. This approach was taken because response options do not traditionally include information about results of the competency described; the objective is to collect judgments of incidents and their respective behaviours and not the results of those judgments and behaviours. The test development team also conducted an online-based literature review. The basis for the literature review was fuel by situations documented during the Critical Incident Report and provided the test development team with a wider organizational context and scope for the scenarios described.

The EPSI development team began developing test scenarios and items based on the analysis provided above. The fictitious organizational context was developed alongside twenty scenarios, each with one to three questions and five response items. Each question is highly associated with the competency it is assessing, the trends and analysis, and the Critical Incident Reports. The first draft of the test was reviewed by psychometric experts to ensure consistency, content validity, and current testing standards (Standards for Educational and Psychological Testing).

Eight high performing Directors at the EX-01 level at the Canadian Food and Inspection Agency participated in the pre-testing session. Each participant completed the Management and Leadership Skills Test and provided written and oral feedback to test developers. This feedback was taken into consideration by the EPSI development team and participant responses were analyzed for consistency and agreement. Situations and responses were modified or removed to improve the consistency and quality of the test.

Five psychometric experts independently evaluated each situation and its respective responses in order to identify which competencies were being assessed. Each situation was correlated with the competency decisiveness and one or two additional competencies depending on the complexity of the situation and the most effective response. The competencies chosen were associated with the most effective answer. The psychometric expert's competency identifications were cross examined and this identification ensured that the most effective response was associated with the most suitable competencies.

Appendix C

MLST Competencies and Behavioral Indicators

1. Strategic Thinking :

- a) Develops operational and policy objectives that are effective and valid for the short to medium-term.
- b) Adapts plans and strategies to reflect changes in unit's operations, policies, business context, etc. as required.
- c) Effectively evaluates the risks and benefits of different courses of action on operations over the medium term and acts on them.
- d) Translates branch strategy into simple, meaningful explanations that others can grasp and implement in their quarterly and yearly plans.
- e) Integrates branch needs into a short to medium term business plan that articulates and executes the strategies.
- f) Identifies interdependencies in cross-functional projects and sets priorities that ensure coherent implementation of plans.

2. Decisiveness:

- a) Makes difficult, time-sensitive decisions even when only limited information is available.
- b) Makes the right call: stays true and is not influenced by undue interferences or threats.
- c) Acknowledges personal responsibility for outcomes of decisions made during crisis.
- d) Acts promptly and with confidence when a situation requires a quick decision.
- e) Keeps composure during difficult times, in an emergency or under pressure; and acts decisively to resolve business issues.
- f) Promptly decides what activities and projects need to be done even though they may not be easy or have popular support.
- g) Selects an appropriate response by seeing beyond the obvious and superficial through steadied observation rather than emotion.
- h) Anticipates delays and potential problems, plans for contingencies when scheduling work and adjusts time frames accordingly.
- i) Shows a significant level of effort to achieve objectives and maintains commitment even in the face of adversity.

- j) Sets priorities and makes timely decisions that address risks to the organization.

3. Change Leadership:

- a) Publicly describes the need for change in the organization as a positive outcome.
- b) Personally exemplifies or embodies the desired change through strong, symbolic actions that are consistent with the change.
- c) Repeats message for Agency transformation wherever possible to foster engagement and excitement.
- d) Defines an explicit vision for change, defining the burning platform for change.
- e) Is a true agent of change: wins the heart and mind of employee for the change initiative.
- f) Addresses change and obstacles with a positive outlook, composure and resilience.
- g) Celebrates the effort whether things go right or wrong and follows through with implementation of lessons learned.
- h) Takes efforts to deliver the message or vision for change to everyone affected.
- i) Involves others in planning for and implementing changes and in so doing gains buy-in for change.

4. People Development and Management:

- a) Communicates a collective purpose and creates a clear line of sight to CFIA's mandate and agenda.
- b) Fosters a climate of solidarity amongst employees by treating each person as a valued team member.
- c) Involves others in planning for and implementing work plans, and in so doing gains their buy-in.
- d) Rewards the contribution of team members; profiles individual and group accomplishments and talents across different functions, as appropriate.
- e) Creates an engaged work climate where staff understand the goals and objectives of the group or of the change project.
- f) Shares information and the broad context to enrich the teams' understanding of the mandate and how their work makes a difference.
- g) Values diversity and recognizes the unique skills and contributions of all team members.
- h) Evaluates individual performance fairly, taking account of diversity.

- i) Identifies opportunities that challenge and encourage the development of people.
- j) Provides regular feedback; acknowledges success and the need for improvement.
- k) Develops and supports individual career plans and learning opportunities.
- l) Ensures an appropriate cascade of corporate objectives into the performance objectives of employees.

5. Stakeholder and Relationship Building:

- a) Seeks new relationships outside own unit and identifies new partnerships that better position CFIA's programs and activities.
- b) Identifies and creates opportunities to meet new individuals through networking, participating in cross-functional, multi-stakeholder groups, etc.
- c) Identifies opportunities to make new relationships with CFIA and or government organizations.
- d) Encourages the development of trust with stakeholders by promoting open, constructive discussions.

Appendix D

MLST Confidentiality, Instructions and Sample Situation

Consent Form

This is to certify that I have no prior knowledge nor have I received any information concerning the content of the questions relating to this competition.

Also, I will not discuss the content of this examination with any one other than the responsible Human Resources Advisor until the results of the competition have been officially made public as this examination may also take place at another date.

Candidate's Name (Please Print): _____

Candidate's Signature: _____

Date: _____

If before or during the testing session, you experience physical or psychological indisposition of sufficient severity to interfere with your test performance, it is your responsibility to inform the test administrator that you cannot undertake or continue the test.

Please be aware that if you choose to undertake or continue the test despite your indisposition, you must accept its results. If you choose to leave once the administration of the exam has begun, you will be considered as withdrawn from the process.

If you have a disability that could impede your performance on this test, please advise the test administrator now so that alternative arrangements can be made.

Instructions

As part of the evaluation process for the position of EX-01, an assessment will be used to evaluate the competencies that are deemed necessary to occupy this position.

This assessment was developed in order to measure the following essential competencies:

- Decisiveness;
- Strategic thinking;
- Stakeholder Relationship Building;
- People Development and Management, and;
- Change Leadership.

Part A: Management and Leadership Skills Test

You will be introduced to a number of challenging situations that a person might encounter at work while performing the duties of a generic Director position. You will be provided with five (5) response options. All responses may be effective to some degree; however it is your responsibility to select which response is the most effective within the described context.

For this exercise, you must assume the role of the EX-01 Director position of a fictitious agency specifically created for this exercise. Each situation provides you with all of the information you need to respond to the questions. Please be attentive to the content of the exercise and the information provided. You are allowed to make reasonable assumptions, however, your answers should reflect only the information provided per issue. Please note that for the purposes of this exercise, the date is April 2nd of 2025.

The test consists of 26 questions and you will have one (1) hour to complete it. You will obtain one point for each correct response, but no points will be deducted for an incorrect response. Therefore, it is to your advantage to respond to all of the questions, including the ones for which you are unsure.

Part B: Inventory

You will be asked to respond to forty (40) questions from the Balanced Inventory of Desirable Responding. Part B is voluntary and will be for research purposes only. The results will not affect your application or test results. It will take you approximately fifteen (15) minutes to complete.

Sample Scenario and Question

To: Director, Atlantic Region (YOU)

From: J. Jones, Executive Director

Subject: Assignment opportunity

The Safe Water Agency (SWA) has announced an opportunity for a one-year assignment (at level EX-01 or equivalent) to the SWA Atlantic region. The incumbent will be invited to serve on the National Advisory Committee on Works and Distribution. The Committee will be mandated to provide overall guidance and leadership on policies regarding water purification and distribution as well as wastewater treatment and collection. The Committee will also develop and implement new measures in order to improve and maintain inter-agency cooperation and collaboration.

While this assignment represents a tremendous opportunity to recruit, retain and re-engage older employees, the appointment to the position does not require proficiency in both official languages. As you are aware, the SWA lacks bilingual capacities for continuous service in both official languages. This is especially troublesome for the Atlantic region given the higher rate of bilingualism in the Maritimes. As reported in our most recent internal audit, the majority of complaints received focused on a lack of bilingual officers among the Agency.

Given the difficulties associated with recruiting bilingual staff and our current shortage of bilingual officers, I strongly recommend that you consider assigning an employee who meets the minimal language proficiency requirements of the position to which they will be assigned. I am counting on you to act in accordance with the SWA's best interests.

Regards,

J. Jones

Executive Director

Question 1: What would be the most effective answer?

- A. You thank J. Jones for his input and assure him that you will carefully consider all your options before making a decision.
- B. You request a list of the most highly qualified applicants to determine if a unilingual employee would make a suitable incumbent.
- C. You set up a meeting with your branch managers to discuss the impact of possibly reducing bilingual staff on workload.
- D. You conduct an assessment to better understand the consequences of reducing bilingual staff and allocate the remaining resources accordingly.
- E. You post a public request for individuals interested in the position.

Question 2: Given the considerable impacts of reducing bilingual staff for a year, which of the following actions is the most appropriate?

- A. You appoint the most suited employee to the position regardless of their language proficiency and the impact on workload.
- B. You explore the possibility of hiring external bilingual resources should you appoint a bilingual employee to the position.
- C. You appoint a unilingual employee that meets the minimal language proficiency requirements of the position to which they will be assigned.
- D. You appoint the most suited employee to the position regardless of their language proficiency provided that the remaining staff can manage an increased workload.
- E. You do not appoint anyone to the position; there will be opportunities in the future when language proficiency is less of a necessity in your unit.

Appendix E

Gender Results. Of the eighteen (18) participants, six (6) individuals identified themselves as female, and twelve (12) individuals identified as male. Sixty-six percent (66%) of women had low BIDR results, and twenty-five percent (25%) of men had low BIDR results. Fifty percent (50%) of men had high BIDR results, and sixty-six percent (66%) of that group failed the MLST. One hundred percent (100%) of women had short response latencies. Fifty percent (50%) of men had above short response latencies, thirty-three percent (33%) had medium response latencies, and sixteen percent (16%) had long response latencies. One hundred percent (100%) of men who had long response latencies had high BIDR scores and failed the MLST.

Thirty-three percent (33%) of women passed the MLST, and 33% failed. Forty-two percent (42%) of men passed the MLST.

Age. Twelve (12) participants identified their age, ranging from 31-61+. One hundred percent (100%) of men from ages 31-40 had short response latencies, with varying inconclusive results for the BIDR and MLST.

One hundred percent (100%) of women from ages 41-50 had short response latencies; sixty-six percent (66%) of which had low BIDR results. The MLST results were varied and inconclusive. Sixty-six percent (66%) of men from ages 41-50 had short response latencies and moderate BIDR results. The men in age ranges 51-60 all failed the MLST and had high to moderate BIDR results. The response latency times were varied and inconclusive. The men in age ranges 61 - + had high BIDR results, passed the MLST and had short response latencies.

The statement of contributions

Social desirability may threaten the validity and use of psychometric assessments in organization contexts. This research is oriented to analyze the relationship of socially desirability responding and the use of psychometrics assessments to appoint individuals for competitive positions. In doing so, this research included real-applicants to accommodate a research need, and provided insight into response latency and its relationship with social desirability. The present findings demonstrate that participants did participate in socially desirable responding; however, this behavior did not appear to improve their results on the assessment as hypothesized. The research also identified a significant correlation between social desirability and response latency, and suggests that shorter response latencies are associated with lower social desirability scores. As such, these results contribute to the academic research on social desirability and response latency in organizational contexts.