

THESIS

AUTOMATED SCORING IN ASSESSMENT CENTERS: EVALUATING THE  
FEASIBILITY OF QUANTIFYING CONSTRUCTED RESPONSES

Submitted by

Diana R. Sanchez

Department of Psychology

In partial fulfillment of the requirements

For the Degree of Master of Science

Colorado State University

Fort Collins, Colorado

Fall 2014

Master's Committee:

Advisor: Alyssa Gibbons

Kurt Kraiger  
Kate Kiefer  
Lucy Troupe

Copyright by Diana Ruth Sanchez 2014

All Rights Reserved

## ABSTRACT

### AUTOMATED SCORING IN ASSESSMENT CENTERS: EVALUATING THE FEASIBILITY OF QUANTIFYING CONSTRUCTED RESPONSES

Automated scoring has promised benefits for personnel assessment, such as faster and cheaper simulations, but there is yet little research evidence regarding these claims. This study explored the feasibility of automated scoring for complex assessments (e.g., assessment centers). Phase 1 examined the practicality of converting complex behavioral exercises into an automated scoring format. Using qualitative content analysis, participant behaviors were coded into sets of distinct categories. Results indicated that variations in behavior could be described by a reasonable number of categories, implying that automated scoring is feasible without drastically limiting the options available to participants. Phase 2 compared original scores (generated by human assessors) with automated scores (generated by an algorithm based on the Phase 1 data). Automated scores had significant convergence with and could significantly predict original scores, although the effect size was modest at best and varied significantly across competencies. Further analyses revealed that strict inclusion criteria are important for filtering out contamination in automated scores. Despite these findings, we cannot confidently recommend implementing automated scoring methods without further research specifically looking at the competencies in which automated scoring is most effective.

*Keywords:* assessment centers, automated scoring, qualitative content analysis, technology

## ACKNOWLEDGEMENTS

The first acknowledgement goes to my family for all their love and support through this process. This paper would not have been possible without assistance from Steve Raymer, Margaret Vicker, Chelsey Green, Kayla Fertman & Amber Anthenien who assisted with developing and applying the coding schema in this project. A special thank you to them for their time and intellectual contributions to this project.

## TABLE OF CONTENTS

Introduction.....	1
Assessment Centers .....	3
Technology Integration in Assessment Centers.....	4
Competitive need to stay technologically current.....	5
Reducing the high cost of assessment centers .....	6
Assessment centers must stay current with job requirements.....	7
Automated Scoring as a Key Technology in Assessment Centers .....	9
Constructed-responses .....	10
Selected-responses .....	10
Automated scoring.....	11
Psychological fidelity in automated scoring.....	14
Automated Scoring in Other Contexts.....	15
Current Study .....	16
Methods.....	20
Qualitative Content Analysis .....	21
Deductive content analysis .....	21
Inductive content analysis.....	21
Analysis Plan .....	23
Automated scoring method.....	24
Qualitative data method .....	24
1. Created coding schema .....	25
2. Behaviors coded using coding schema .....	27
3. SME effectiveness ratings and links to original competencies.....	28
4. Created automated scores .....	30
5. Final scores calculation and data analysis .....	34
Results.....	38
Participants.....	38
Descriptive Information .....	38
Hypothesis Tests: Hypothesis 1 .....	40
Hypothesis 1: Mid-level managers .....	40
Hypothesis 1: High-level managers .....	41
Hypothesis Tests: Hypothesis 2 .....	41
Hypothesis 2: Mid-level managers .....	41
Hypothesis 2: High-level managers .....	42
Hypothesis Tests: Hypothesis 3 .....	42
Hypothesis 3: Mid-level managers .....	43
Hypothesis 3: High-level managers .....	43
Direction of interaction .....	44
Comparing Scoring Methods .....	44
Exploratory Analyses.....	45
Inter-correlations within scoring methods .....	45
Original scoring method .....	45

Conservative automated scoring method .....	46
Liberal automated scoring method.....	46
Convergent and discriminant validity .....	47
Exploratory Factor Analysis .....	48
Discussion .....	50
Implications of Research Findings.....	50
Outcome Measures.....	51
Content Restriction .....	53
Future Research .....	56
Limitations .....	56
Sample size .....	57
Cognitive process.....	58
Resources for development.....	58
Contributions.....	61
Conclusion .....	62
References.....	109
Appendix I .....	117

## INTRODUCTION

The availability of new technologies to the general public has continued to increase over the years. These technological advancements play an increasingly important role in organizations and for the field of selection and training assessments, as evidenced by the growing number of technologically themed submissions and sessions presented at the Society for Industrial and Organizational Psychology (SIOP) annual conference. The 2013 conference hosted over 36 sessions (out of 300) that were focused on ways to utilize various forms of technology (e.g. mobile devices, computer animations, online simulations) in evaluative environments (SIOP, 2013). The 2014 conference hosted over 55 sessions (out of 339) regarding technology within organizations (SIOP, 2014).

Technology integration is not a new concept for organizations, which have been using technology for decades to simulate work situations that may be considered unsafe, unethical, or particularly challenging to recreate. This has included situations where companies trained workers using simulations of dangerous environments (e.g., underground mines) or training physicians and nurses to perform medical surgeries and procedures on simulated patients (Denby & Schofield, 1999; Wood & McPhee, 2011; Zajtchuk & Satava, 1997). Assessment centers (ACs) have been no exception to this trend of using both technology and simulations within organizational settings. ACs use complex behavioral simulations to assess employee competencies for selection and development, which makes them time-consuming and resource-intensive to implement. AC users have long been concerned with ways to reduce these costs while preserving the many benefits of the method (e.g., Thornton & Potemra, 2010; Tziner, Meir, Dahan, & Birati, 1994). Computer and web delivered simulations offer the potential to

reduce expenses associated with AC administration while maintaining the realism of the simulations (Hawkes, 2013; Latorre, 2013). However, there are few empirical evaluations of high-technology ACs, and both academics and practitioners have highlighted the need for more rigorous research in this area (Gibbons, Hughes, Riley, Thornton, & Sanchez, 2013; Hatfield, 2013).

One area in need of particular attention is the use of automated scoring methods for simulation exercises. Automated scoring appears to be of considerable interest to many AC practitioners, due to perceived benefits such as improving consistency and reducing (or even eliminating) the cost of human assessors (Hawkes, 2013). However, research on automated scoring is relatively sparse, particularly when applied to ACs (Handler, 2013), and even in other industries (e.g., medicine, education; Clauser, Margolis, Clyman, & Ross 1997) it appears that the application of automated scoring has outpaced the research (Xi, Higgins, Zecher, & Williamson, 2012).

In this paper, I examine the feasibility and validity of using an automated scoring method to evaluate a qualitative response to an AC simulation, and create a scoring algorithm to evaluate those qualitative responses. In the first phase, I categorize and summarize the behaviors captured in participants' open-ended responses according to common qualitative analysis methodologies (Schilling, 2006). During this categorization phase, I explore whether the behaviors exhibited in response to a traditional constructed-response simulation can be condensed into a manageable number of meaningful categories, which could be used to partially or fully automate the scoring process (e.g., by converting the simulation to a selected-response format). Those categories of behaviors are then be weighted by effectiveness using ratings from subject matter experts (SMEs) and then linked to the competencies from the current AC simulation. The effectiveness



weights are then translated into a scoring algorithm using a judgment-derived statistical weighting system. Specifically, the behaviors identified in the constructed-responses are weighted according to their frequency and effectiveness to create an automated score. The original score from human assessors is then regressed on the automated scores to determine whether there is convergent validity between the scores. This research provides a meaningful contribution to the AC research literature by directly comparing automated and human-based scoring methods using the same AC simulation responses collected from the same participants. This proof-of-concept study helps to fill the gap in research regarding the possibility of using automated scoring methods within an AC context.

### **Assessment Centers**

An AC is a battery of real-to-life simulations given to participants over the course of one or more days (Thornton & Rupp, 2006). Organizations who invest in an AC typically commit substantial time and resources to development and implementation (Lievens & Patterson, 2011), because organizations wish to ensure that the simulations will accurately replicate the work done on the job (Roth, Bobko, & McFarland, 2005). In addition to the time and resources that must be invested in developing realistic simulations, ACs are also expensive to implement and score. The assessors who evaluate a participant's performance in the simulations are carefully trained to observe, document, and rate behavior according to the standards of the AC. The widely accepted *Guidelines and Ethical Considerations for Assessment Center Operations* (International Task Force, 2009) specify detailed requirements for assessor training, and further note that input from multiple assessors is necessary for a procedure to qualify as an AC. Thus, administering an AC requires a number of highly trained and qualified personnel, which contributes significant costs and limits the number of participants who can be assessed at any given time.

The popularity of ACs has been sustained over the last several decades, despite the high level of monetary investment that is required by organizations (Thornton & Potemra, 2010). The combination of benefits that ACs offer illustrates why organizations have continued to use ACs in spite of the cost and resource commitment (Roth et al., 2005). ACs have a wide range of benefits including effective use for selection and screening purposes (Arthur, Day, McNelly, & Edens, 2003). Other positive outcomes associated with ACs and simulations include predicting a broad range of constructs, reducing adverse impact, and making it difficult for participants to fake their responses (Lievens & De Soete, 2012). Further, ACs can test participants on constructs that are difficult to assess using traditional paper and pencil tests or interviews, such as conflict management or public speaking ability (Lievens & Patterson, 2011). Regardless of an organization's commitment to using the AC process, integrating technology into AC procedures may provide some ease to the high monetary investment ACs require.

### **Technology Integration in Assessment Centers**

Researchers have recognized that there are untapped technological capabilities that could influence substantial change for ACs (Rupp, Gibbons, & Snyder, 2008). Efforts over the last few years to integrate computer-based systems into ACs have included computerized simulations and automated scoring (Lievens, Van Keer, & Volckaert, 2010; Rupp et al., 2008). There are many potential benefits to using computerized simulations: they elicit positive responses from participants, they generally provide a realistic job preview, and they make sense to clients and internal stakeholders (O'Connell, 2013). Despite these well-known advantages to using simulations, researchers recognize that there are still other unknown benefits (DeMaria et al., 2010). One major benefit of using a live simulation is that it provides a high degree of realism (i.e., higher physical fidelity). However, contemporary computer-based simulations offer a high

level of physical fidelity as well, while reaching a broader audience without the same demand on personnel and resources for training and administration (Marsch, 2011).

As previously mentioned, technology is attractive to AC users and practitioners because it has the potential to reduce costs in both development and administration of ACs. Depending on the extent to which the target job requires technology use, high-technology simulations can enhance the fidelity of the assessment to the target job and transfer skill confidence onto the job (Gordon & Buckley, 2009). Technologies such as digital video recording, online delivery of simulations, and video communication may reduce the need for participants and assessors to be in the same physical location at the same time, easing travel budgets and scheduling constraints.

Seeing that technology integration has made its way into various aspects of AC development, researchers have predicted that organizations will want technology to be integrated into assessments to a greater extent, specifically pointing to ACs (Aguinis, Henle & Beaty, 2001). In the applied field, practitioners who develop ACs are facing pressures to integrate technology into their assessment processes (Krause & Thornton, 2009). This pressure is at times driven by external factors such as (1) addressing the competitive need to stay current with new technologies as they become available, (2) utilizing technology to reduce the high cost of ACs, and (3) staying true to the fidelity of technological changes that take place on the job. Each of these potential sources of pressure will be discussed in turn.

### **Competitive need to stay technologically current**

Practitioners have reported interest from clients to utilize technology in their assessments stating that they gain client confidence when technological capabilities can be demonstrated (Hatfield, 2013). There are a variety of reasons why companies would be interested in having technologically current assessments in their personnel management systems. First, companies

may wish to appear more technologically savvy and up to date to potential job applicants, as research has shown that technology-based selection methods can lead to positive applicant reactions (Anderson, 2003). Second, companies may also choose to request more technologically advanced assessments from a desire to administer assessments more quickly, such as using internet-based assessments (Aguinis et al., 2001). Third, companies may be aware of the adverse consequences that organizations face when failing to stay technologically current with competitors. Examples include companies such as Borders and Blockbuster; both once ubiquitous in their own industries, yet ultimately overcome by a failure to innovate and stay current with technology and trends in their respective fields (Leopold, 2011). Organizations must strike a balance between investing in technology that will remain salient in the future of their industry, yet not overinvesting in technology which will become obsolete (Ericson, 2006). Abstaining from technology can cause companies to be left behind in industry. However, overinvesting in the wrong technology could have similar adverse consequences (Leopold, 2011).

### **Reducing the high cost of assessment centers**

The high cost of ACs can be a deterrent for some administrators when considering new AC projects. Understandably, the resource intensive process of ACs can be burdensome for administrators who are tasked with considering the development and validation of the simulations, which often includes training assessors, paying salaries, acquiring space for administration, and travel costs. However, advancements in technology have created new avenues for technology integration that could reduce some of the costs associated with ACs. The virtualization of communication allows collaborators to work together across great distances on projects without the high cost investment of travel (Tannenbaum, Mathieu, Salas, & Cohen,

2012). Administrative tasks can be streamlined and automated so that registering, signing-in, providing instructions, and debriefing participants can all be done electronically rather than face-to-face. This may result in a greater number of participants being processed more quickly and simultaneously. In addition to these solutions, using automated scoring methods and excluding a trained assessor can reduce the cost of developing and administering a training course for those assessors as well as the cost of the assessor's salary (Luck, Peabody, & Lewis, 2006; Zhang, Williamson, Breyer, & Trapani, 2012).

### **Assessment centers must stay current with job requirements**

Because ACs utilize work samples, simulations, and other real-to-life simulations, participants often report experiences that closely mimic what exists in a true work environment (Lievens & De Soete, 2012). The realism of an assessment, also called fidelity, can be considered in two parts; physical fidelity (i.e., the assessment environment looks, feels, and behaves similarly to that which is experienced on the job) and psychological fidelity (i.e., the individual is behaving and engaging with the assessment in a way that is similar to that which is done on the job; Ward, Williams, & Hancock, 2006). Researchers have recognized that developing an assessment that closely resembles the work that is done on the job (i.e., physical fidelity) can be very difficult, and at times assessments with high physical fidelity do not predict job performance (Aguinis et al., 2001). When integrating new technologies into an assessment, it is important to ensure that the prediction of the intended outcome (e.g., job performance) is maintained after the technological changes are put into place.

The technological trend of using simulations can be attributed to the need to maintain this similarity between the assessment and the construct from the job that is being evaluated (i.e., fidelity). As the use of technology has continued to flourish, most industries have been saturated

with technology related programs, devices, and skills that soon become job requirements for employees (Papadopoulos, 2013). As these technologies and skills become more ingrained in the work, there is a greater need for ACs to replicate that work environment. Since the tasks employees are completing and the tools employees are using involve an increasing number of specific forms of technology, ACs need to also use that technology to accurately mimic the work and the work environment (Lievens & Patterson, 2011). This trend will likely continue, and as organizations improve their technological capabilities, ACs will also need to update their procedures and technology to maintain face validity and physical fidelity. There is research published on the application and benefits of fidelity, primarily looking at the physical fidelity of an assessment (e.g., Mania & Robinson, 2005). However, the research literature does not explore understanding how changes in technology influence the psychological fidelity of an assessment. Changing the technology of an assessment may in some circumstances alter the process being used to complete an assessment or simulation exercise. This alteration could potentially impact how well that process being used resembles a process that would be used on the job (i.e., psychological fidelity). The psychological concern is regarding the method of how information is presented to an individual and how that presentation may change the psychological process that an individual uses in the assessment.

***The gap between technological advancements and both science and practice***

The continued pressure for organizations to implement technology has created an imbalance between science and practice (Anderson, 2003; Handler, 2013; König, Klehe, Berchtold, & Kleinmann, 2010). While this gap exists, both science and practice are being outpaced by the rapid growth in technology (Anderson, 2003). The difficulty that both researchers and practitioners face when attempting to stay current with technology is exacerbated

by the challenge of keeping up with the rapid changes in technology. These constant changes make it challenging for practitioners and researchers to predict which technologies will remain prevalent and for how long. A second difficulty is the risk that the particular technology studied in any given research effort may be obsolete by the time of publication. Despite the challenge of producing research on technology, the value that this research has can benefit practitioners and organizations. Further research on the benefits and consequences of using technology can provide insight and direction to practitioners and their practices. In addition to this, research can guide organizations when they are deciding what forms of technology to invest in by providing information in areas such as effectiveness and potential applications. For example, making job applications available on mobile devices can reduce adverse impact due to the higher availability of smart phones than computers amongst minorities (Scott et al., 2013). As researchers further understand the utility of these technologies, other researchers can continue their work by asking new questions and practitioners can maximize the benefits by applying these technologies. At times, practices are popularized and implemented before conclusive research is completed on the implications of the associated technology. Perhaps the most controversial application of technology in ACs is the use of automated scoring processes to replace human assessors. Automated scoring methods, as a technological solution in ACs, will be the focus for the remainder of this paper.

### **Automated Scoring as a Key Technology in Assessment Centers**

Automated scoring is appealing to AC users because it permits a reduction in the number of human assessors used to observe and score AC simulations (O'Connell, 2013). A recent survey found that automated scoring was among the top three technologies AC users anticipated adopting in the next two years (Gibbons et al., 2013). In understanding how automated scoring

influences the scoring procedure, I must first look at the different types of responses that can be used within an assessment context (i.e., selected-response versus constructed-responses).

### **Constructed-responses**

Constructed-responses are forms of assessments that require participants to develop their own response to a particular prompt (Thornton & Rupp, 2006). The use of constructed-response assessments in ACs is often attributed to the inability of multiple-choice tests to adequately represent and measure the target construct. Other researchers have recognized the limitations of multiple-choice assessments for providing a comprehensive measure of the intended constructs (Williamson, 2012; Williamson, Xi, & Breyer, 2012). However, constructed-response assessments have challenges as well; common criticisms include high costs to administer and low reliability in scoring (Liu, Lee, & Linn, 2011). Developing standardized, automated scoring options can be difficult for constructed-response assessments because the number of responses is potentially infinite. Thus, constructed-response assessments often require a human assessor to observe and score the response. Although this type of holistic assessor scoring is more resource intensive, usually costing more time and money, the benefit of using constructed-response assessments is that they often allow qualitative data to be gathered. Gathering qualitative data can be particularly beneficial in understanding more with regard to the participant's perspective, thought process, or considerations made during the simulation (Liu et al., 2011).

### **Selected-responses**

In contrast to constructed-response assessments, selected-response assessments are a set of pre-determined, discrete responses from which a participant is asked to select the correct or most appropriate response (Klassen, 2006). Common types of selected-response evaluations include situational judgment tests (SJTs) and multiple-choice tests. In an SJT, participants are



presented with a job-related scenario and then asked to indicate the most effective response from a set of provided options (Ployhart, 2006). These different options are usually pre-scored by SMEs on the effectiveness of the response (Weekley, Ployhart, & Holtz, 2006). One concern regarding selected-response assessments is that participants may be able to use judgment to determine which option amongst a set of options is the best course of action, although they may not have that particular behavior within their own repertoire of capable actions (Lievens & Patterson, 2011). Thus, a participant may not be able to behave in a particular way, but may still be able to choose the best response from a set of provided response options. Administrators often choose to use selected-response forms of simulations because they are typically faster and cheaper to implement (Williamson, 2012). Further, these forms of evaluation have definitive levels of correctness and lend well to the use of automated scoring (Lievens & Patterson, 2011). These benefits have caused SJTs and other forms of selected-response assessments, such as multiple-choice tests, to become increasingly popular, particularly within AC contexts. They are often implemented alongside, or in place of, constructed-response simulations (Lievens & Patterson, 2011).

### **Automated scoring**

Automated scoring is a process in which a computer program models the actions of a trained assessor to evaluate an assessment, commonly involving a scoring algorithm (Powers et al., 2002; Williamson, 2012). This process can typically be considered in several different ways. By automating the front end of the AC simulation, the format of the simulation is altered into a selected-response structure that facilitates an automated scoring process. For example, an administrator may choose to automate the scoring process of an existing AC simulation by evaluating the common responses to the simulation and then creating a selected-response

simulation based on those frequent responses. Those selected-responses would then be reviewed and scored by subject matter experts (SMEs) before deployment of the assessment. Future iterations of the simulation can be scored automatically without further review from an SME or human assessor. By automating the back end of the AC simulation, a researcher could theoretically use a constructed-response simulation and automate the scoring system using advanced programs such as text-scanning software. A third potential interpretation of automated scoring is to use a human assessor while removing the judgment component of the scoring process. In this situation, the human assessor might check boxes on a checklist but would not apply subjective interpretation to the score. This checklist would theoretically include clearly dichotomous items that don't require personal judgment from the assessor.

Automated scoring has clear advantages, potentially making the assessment process cheaper, faster, and more consistent. Using scoring algorithms in assessments can add some of the benefits typically associated with selected-response items, such as consistent and expedited scoring which is usually associated with lower costs (Zhang et al., 2012). However, researchers have raised important conceptual concerns about the use of automated scoring (Xi et al., 2012). Automated scoring is difficult to use with constructed-responses, and there is reason for concern that automated scoring may change the meaning of simulations in important ways. Current automated scoring processes can only provide a score based on a predetermined set of scoring rules and cannot adjust for situations when necessary, as a human assessor could, nor can they accommodate novel behaviors (Marsh & Fayek, 2010).

Most applications of automated scoring require using selected-response simulations rather than constructed-response simulations such as SJTs. SJTs can easily be scored with automated methods, as the effectiveness of the various response options can be established in

advance. However, only the responses identified in advance by the test developers are available to the participants, who are unable to vary their behavior from the expected behaviors that fit within the parameters established in the scoring algorithm (Luciana, 2003). This means that automated scoring creates a set of confines within which a participant is expected to behave. Yet most AC simulations are open-ended and require constructed-responses from participants. It is not clear from research whether these assessment responses are appropriate for automated scoring methods (Boyce, 2013). In terms of the testing and assessment research literature, most automated scoring methods essentially transform constructed-response assessments into selected-response assessments (e.g., Liu et al., 2011). Williamson (2012) reported that very few large-scale assessments utilize a constructed-response format due to the efficiency of developing and administering automated scoring processes for selected-response assessments.

It can be argued that selected-response assessments require a different psychological process than constructed-response assessments; that choosing the correct behavior from a provided list does not require the same cognitive process as executing behavior (Hawkes, 2013; Klassen, 2006; Williamson, 2012). That is, “life is not multiple-choice,” and the range of behavior that can be assessed is limited by enumerating the realm of possible participant responses in advance (Ryan & Greguras, 1998). Thus far, only one published study has explored the concept of automated scoring and how to properly apply it in an AC context (Lievens et al., 2010). Lievens et al., (2010) presented the participants in their study with a complex set of branching selected-response options; in total, there were over 2,000 distinct behaviors that participants could potentially select across the course of the assessment. This raises an important question: how much variation in behavior naturally occurs in a constructed-response simulation?

Can the range of behaviors participants spontaneously initiate be captured effectively within a limited set of behavioral categories, and if so, how many categories are needed?

### **Psychological fidelity in automated scoring**

Research on technology in ACs generally emphasizes the benefits of physical fidelity and face validity, but commonly excludes the impact that automation has on psychological fidelity (Clauser et al., 1997; Lievens & Patterson, 2011). Automated scoring can impact psychological fidelity because changing how a simulation within an assessment is presented impacts the cognitive processes a participant uses to respond to that simulation. Selecting an action from a predetermined list of possible actions or behaviors (i.e., selected-response) rarely replicates the behaviors expected on the job (e.g., more often people are expected to develop a plan of action or carry out instructions). Thus, the psychological fidelity of these selected-response assessments in most cases will not mimic the behaviors performed on the job. In situations where the psychological fidelity is lowered, the assessment may not provide a clear indication of how the participant would perform on the job, because the response from the simulation no longer reflects the psychological properties of the job and what a participant would be expected to do on the job. Changing the cognitive process of an assessment may impede the usefulness of those results to predict the intended outcomes (e.g., job performance).

Other researchers have recognized this same concern, referencing response fidelity and the range of behaviors that exist in assessment performance (Ryan & Greguras, 1998). When the responses presented to a participant are limited to a predefined set of selected-response options (e.g., SJTs), the psychological fidelity of the simulation may be reduced (Boyce, 2013). Medical researchers have shown that workplace simulations with low psychological fidelity typically represent idealistic rather than realistic scenarios that occur in the workplace (Yardley, 2011).

This means that low psychological fidelity denies participants the opportunity to practice and perform work related behaviors under realistic conditions. During high psychological fidelity simulations, participants may engage in emotionally enhanced simulations such as distressing interactions with unsupportive colleagues (Yardley, 2011).

In considering a study where behaviors are being simplified to fit into an automated scoring method, it is important to consider how the change in the behaviors being accounted for will impact the psychological fidelity of the assessment. Using an automated scoring method could potentially lead to the loss of information that would otherwise be captured and evaluated by a human assessor. Scoring methods progressively lose more variation in behavior as they move away from a participant constructed-response format towards behavioral checklists and a selected-response format. This study aims to discover how much information is lost in this move; that is, to determine if discrete behaviors can account for a similar level of variation in behavior as a constructed-response simulation.

### **Automated Scoring in Other Contexts**

There are gaps in current AC research regarding the application and utility of automated scoring methods. Thus, theories and findings from other industries can be drawn upon to help build theories and predictions for the purposes of this study, as is done here. These other areas can inform thinking and expectations surrounding AC practices until additional research begins to address these critical issues. The topic of automated scoring has surfaced in other industries such as computer science, information technology, education, medicine, and linguistics (Clouser et al., 1997; Xi et al., 2012; Yardley, 2011). Other practitioners have explored the use and integration of technology in training and assessment venues.

Outside the AC context, other researchers have asked questions similar to the present study with regard to automated scoring. Two studies are outlined here wherein the researchers took constructed-responses from individuals and compared an automated scoring method to a human rated scoring method. In the first study, researchers looked specifically at speech recognition to evaluate the pronunciation of English words and phrases from non-native English speakers (Xi et al., 2012). In the second study, physicians were tested on patient management skills using a computer simulation (Clauser et al., 1997). Both studies used a judgment-based approach and found that the automated score predicted the observer rated scores from expert assessors; indicating that a judgment-based scoring model can be used to approximate expert judgment on computer-simulated exercises (Clauser et al., 1997; Xi et al., 2012). My study takes a similar approach to this question by comparing alternative scoring methods. However, my research looks specifically at simulations within an AC context.

### **Current Study**

The purpose of this study is to determine if automated scoring can be used for constructed-response AC simulations that are traditionally scored using human assessors. This study will determine if an automated score can significantly predict the human assessor score given on the same simulation, which to my knowledge has never been done in an AC context. Although Lievens et al., (2010) evaluated automated scoring within the AC context, the data in their study was solely selected-response, whereas the data in this study will be constructed-responses from participants. A common qualitative content analysis approach will be used to reduce the original constructed-responses from participants into discrete behaviors (similar to a selected-response format).

The goal of the study is to determine whether qualitative (i.e., constructed) responses can be condensed into a list of discrete behaviors and to examine whether those behaviors could produce an automated scoring method that can account for a substantial portion of the actual behaviors displayed by participants. In short, will an automated score predict the original human assessed score? An outline of the analysis plan is provided in Figure 1. This process will involve evaluating the original participant responses and organizing them into categories to create a coding schema, which in this context means the structure of discrete, observable behaviors found in participant responses. This coding schema will be weighted using effectiveness ratings from SMEs, which together will form an automated scoring method that will be applied to the participant's responses to produce an automated score. This is considered an *automated* score because it will be generated using discrete coded behaviors that were categorized according to a set of criteria. Although the scoring method at this stage will not be fully automated, it does provide evidence for the next step in this process by addressing the question of categorizing AC constructed-responses into a manageable number of behaviors. This is a critical first step in determining the feasibility of using an automated scoring method in an AC. For this to be considered a plausible alternative, evidence must first be found that AC responses yield discrete, observable behaviors. Additionally, it must be shown that these behaviors could be used in an automated scoring method with substantial behavioral overlap with the original responses (enough to statistically predict the original scores).

As mentioned, the automated score will be compared to the score assigned by the human assessor during the original AC simulation (referred to hereafter as the *original score*). I expect that a substantial portion of the variance in behavior from the original simulation will be retained in the automated score, so that this score will show convergent validity with the original score.

This is based on findings from previous studies in which automated scores significantly predicted human assessor scores (Xi et al., 2012). Thus, I argue that the automated score will predict the original score.

*H1: The automated score will significantly predict the original score.*

Automated scores exclude a substantial amount of behavioral variance from consideration because the content is coded into behavioral categories that best represent the behavior (i.e., translating qualitative data into quantitative codes). A primary theoretical question in this study is how much this content restriction impacts the ability to predict the human-rated score. Based on this question, I would like to learn how well the automated score uniquely predicts the original score after controlling for variance that occurs from other known effects. One known effect that I will account for in this study is word count of participant responses. Previous research has shown that including more content in a response accounts for more variance in the outcome (Shermis, Shneyderman, & Attali, 2008). Qualitative researchers generally accept this finding that the length of a participant's response is an indicator of quality (Barrios, Villarroya, Borrego, & Olle, 2011). "The more words a respondent uses to answer an open-ended question, the more detailed the response will be, and the more useful the information" (Barrios, Villarroya, Borrego, & Olle, 2011, pg. 210). This relationship has been consistently found across different definitions of "quality" including correct answers on knowledge assessments (Jordan, 2012), more truthful responses on self-report questionnaires (Colwell, Hiscock, & Memon, 2002), and the number of arguments included in a persuasive essay (Spörrle, Gerber-Braun, & Fösterling, 2007).

Practically speaking there is some value to verifying that the findings from this study are consistent with other research findings. However, my overall intention for evaluating whether the



length of participant responses is associated with the quality of those responses (i.e., original and automated scores), is not to replicate the results of previous studies but instead to isolate the effect response length has on quality as a potential confounding variable. Response length is a threat to automated scoring because the amount of variance it explains could potentially inflate the estimated effect size for the automated score. Understanding the relationship between length and quality for this study will allow us to find the incremental validity of score automation without the shared variance explained by word count alone. Based on previous research, I expect that participants who write more in the qualitative constructed-response will provide more detailed information and will have higher scores (Shermis et al., 2008). Additionally, I expect that the automated score will significantly predict the original score with word count as a moderator. I anticipate that the automated score will better predict the original score as word count increases.

*H2. There will be a main effect between word count and the overall scores for both scoring methods, showing that word count significantly predicts both the original and the automated scores.*

*H3. Word count will moderate the relationship between the automated scores and the original scores so that the automated scores will better predict the original scores as word count increases.*

## METHODS

The purpose of this study was to evaluate qualitative constructed-responses from an AC simulation. Specifically, I sought to determine (a) whether the responses can be meaningfully organized into categories of behavior, which are then rated using an automated scoring method, and (b) whether that score predicts the original human rated score. My data was obtained from an archival data set using participants employed by a large Colorado school district, all of who were in managerial positions. The data was collected during a 2012-2013 operational AC, which was used for developmental purposes. Individuals participated in a three hour, online simulation that mimicked the experience of “a day in the life of” an executive director for a medical organization. This simulation presented participants with emails, projects, and telephone-based role-plays. Responses to these simulations were used to provide feedback to participants regarding leadership skills and pinpoint areas for developmental improvement. All identifying information about the participants was removed before data analysis.

Two simulations were used from two separate ACs for this study. One AC was for mid-level managers and included a simulation regarding a customer complaint. The second AC was for high-level managers who were presented with a simulation addressing a concern from a client. Although the simulations were considered parallel across the two ACs, the data was analyzed separately due to slight differences in the prompt and responses from participants. Participants in both ACs received several emails that would prompt the simulation and were asked to take action and engage in “damage control” regarding the situation (i.e., respond to the email to defuse the situation). Participants chose to respond to the simulation prompt in a variety of ways. Some ignored the prompts and did not engage in the simulation, whereas others

contacted the customer or client directly to address the problem. In contrast to these, some participants delegated actions to a subordinate to follow-up with the customer or client, framing the situation as a developmental opportunity, and at times giving specific directions on how to address the situation. Trained assessors from the original AC evaluated and scored the responses on six performance competencies (see Table 1). The scores assigned by the trained assessors during the original AC (i.e., six competency scores and one overall score) will be referred to as the “original scores.”

### **Qualitative Content Analysis**

Two types of qualitative content analysis were considered for this study. Research has shown that the selection of one of these methods should be based on the amount of available research literature and on the purpose of the study, as it will determine the process used to develop the coding schema (Moretti et al., 2011).

#### **Deductive content analysis**

When research exists to support the use of an applicable theoretical model, researchers typically use deductive content analysis, a process in which the framework and theory for the coding schema is derived from research. With this method, researchers impose preexisting categories supported by the research literature (Elo & Kyngas, 2008; Hsieh & Shannon, 2005; Mayring, 2000). Although most qualitative studies follow this method, using a pre-existing conceptual framework for categorization, I use an inductive method in which the framework for my scoring method is built from the raw data (Schilling, 2006).

#### **Inductive content analysis**

Research shows that when there is a dearth of research or theory in the subject area being studied, inductive content analysis is typically used to pull categories out of the raw data rather

than imposing a preconceived set of ideas or expectations for the categories. This will lead researchers to allow the categories to “flow from the data.” Through this technique, a greater understanding of the data can be gained. This approach is similar to grounded theory in that an understanding of the data will emerge. The inductive approach is appropriate when theory and research in the subject matter are limited or non-existent (Elo & Kyngas, 2008; Hsieh & Shannon, 2005; Mayring, 2000). As the purpose of this study is to apply automated scoring techniques to a constructed-response in an AC context and this type of study to my knowledge has never been done before, there is no strong theoretical basis on which to build an expectation of which discrete behaviors will be observable. Thus, I will utilize the inductive content analysis method for the development of the coding schema.

These AC simulations evaluate participants on leadership skills and behavior, a domain which is thoroughly covered in the research literature (Breevaart et al., 2014; Egri, & Herman, 2000; Howell & Avolio 1993; Sosik & Megerian, 1999). Despite the availability of numerous existing models and theories to draw from, I chose to adopt an inductive rather than a deductive approach for this study because of the need to identify discrete behaviors for developing an automated scoring method. The behaviors identified for this study must be discrete because implementing an automated scoring method requires concrete behaviors as opposed to the level of abstraction built into typical AC competencies and scoring procedures with human assessors. For example, one of the indicators for the competency Mission and Values Leadership in the current AC was “Communicating organizational values and facilitating their inclusion in daily tasks” (paraphrased to maintain confidentiality). Understandably, this behavior requires human judgment to review and interpret the execution of this behavior. It would be difficult for an automated scoring method to determine whether a participant performed this action. This places

a unique limitation on the categorization system; meaning intuitive or nuanced behaviors were excluded from the automated scoring methods and only observable behaviors were included.

Further, the inductive approach was appropriate for the study rather than the deductive approach because the primary research question involved looking at base rates of the behaviors that occurred amongst participants. The deductive approach includes all possible behaviors in analysis, including ideal performance. This study aims to evaluate behaviors that actually occurred from participants and does not analyze behaviors that did not occur. Thus, a deductive approach would have provided extraneous information (i.e., ideal performance dimensions) that would not be used as part of this study.

### **Analysis Plan**

This study includes processing qualitative data (i.e., email responses) taken from AC simulations. Although qualitative data provides a rich source of information from participants, other researchers have warned that this type of data needs to be analyzed and interpreted using a scientific method that has been shown to be both valid and reliable (Black, 2006; Moretti et al., 2011). Qualitative research is often criticized as a subjective and nonscientific approach to data analysis (Cook, 2012). Qualitative researchers find it challenging to follow a consistent process that has been used by other researchers, since most structured methodologies for processing qualitative data are often “vague and abstract” when published in the research literature (Schilling, 2006). It is difficult to replicate the methodology of another researcher when many of the details for the procedure were unclear or unreported. An additional challenge for qualitative content analysis is that its very definition at times establishes ambiguity regarding the structure and empirical basis of the methodology; it is sometimes defined as either a less empirical method (“subjective interpretation of the content,” Hsieh & Shannon, 2005) or a more concrete

procedure (“a systematic, rule-based process,” Schilling, 2006). Despite some differences and controversies in the research literature, researchers have overwhelmingly agreed for decades that with qualitative content analysis, it is critical that the methodology follow empirically objective rules (Mostyn, 1985; Neimeyer & Gemignani, 2003). The methodological approach used in the current analysis plan will draw from research in both the automated scoring research literature and the qualitative content analysis research literature. The two methodologies being integrated into the current analysis plan will be discussed in turn.

### **Automated scoring method**

From the automated scoring research literature, Bennett and Bejar (1998) outlined two steps for developing an automated scoring system. Their process included, first, identifying the important features to be scored, and second, combining those features into an overall score that would indicate general performance. Their first step involved breaking out the individual features of performance into smaller behaviors that the participant would engage in. The second step involved evaluating the effectiveness of those different behaviors and determining how they contribute to overall performance. The analysis plan for this study combined these steps for developing an automated scoring system with another methodological approach from the qualitative analysis research literature, detailed below.

### **Qualitative data method**

Drawing from the qualitative content analysis research literature, I used an approach that aims to categorize and interpret qualitative data into a meaningful structure. This process was primarily be used to guide development of the coding schema and analyses. My process replicates Schilling’s (2006) recommended steps for processing and decision-making with qualitative data. Schilling’s steps include (1) transcribing information into raw data, (2)

condensing and structuring the data, (3) building and categorizing the data, (4) create a coding protocol based on the categorization system, and (5) displaying results for interpretation and further analysis (Creswell, 1998; Schilling, 2006).

The current analysis combines these two approaches, as also shown in Figure 1. The procedures of each step in the analysis plan for this study are outlined below.

### ***1. Created coding schema***

Participant responses were reviewed from two different ACs (i.e., a mid-level manager AC and a high-level manager AC). The simulations in both ACs consisted of constructed-responses (i.e., written emails) in reply to a presented scenario. Responses from the two ACs were gathered, reviewed and analyzed separately. A group of three raters independently reviewed the constructed-responses and created a list of observable behaviors using inductive qualitative content analysis. As automated scoring necessitates discrete and observable behaviors, the current AC competency structure was too broad and abstract to draw categories from. The initial review of responses was done independently by the raters (i.e., investigator triangulation) to prevent each rater's initial thoughts and reactions from being influenced by the interpretations of others (Kapoulas & Mitic, 2012).

Saturation (i.e., the point at which no new information is gathered from the data) is a common indicator when a generally acceptable sample size has been reached in a qualitative study (Sinnott, Guinane, Whelton, & Byrne, 2013). For this study I followed the "Francis method" for identifying saturation (Francis et al., 2010). As part of this method an *initial sample* and *stopping criteria* are established. I set the initial sample at five participant responses, meaning the three raters first independently reviewed five participant responses before coming together and discussing and agreeing upon how to define the behaviors that emerged from the

data. I set the stopping criteria at five additional participant responses, meaning the raters reviewed five additional participant responses after the initial meeting. The guideline for saturation is that after the initial sample, no new behaviors should emerge in the stopping criteria if saturation has been reached (Francis et al., 2010). All three raters reported that no new behaviors emerged in that second set of five that were reviewed after the initial meeting. Thus, saturation for this data set was reached after five participant responses in the sample. It is important to note that some participants constructed multiple emails in response to this simulation, meaning several of the five participant responses consisted of multiple emails.

After saturation was reached, the raters met to discuss the list of behaviors and condense them into categories of behaviors. In developing the categories, raters decided to exclude behaviors displayed by only one participant unless all three raters agreed that the behavior seemed salient to the topic and was likely to occur again within a larger sample of participants. Some similar behaviors were not grouped together if the three raters all agreed there was some value in keeping the behaviors distinct (e.g., differences between the behaviors might make one more effective than the other). The list of behavioral categories was revised into a final list of discrete behaviors (i.e., the coding schema) found in Figures 2 and 3. All decisions on constructing and revising the coding schema were discussed until consensus was reached among all three raters (O'Donoghue & Punch, 2003). Audio recordings were taken of all rater meetings to document discussions and decisions. The raters considered creating separate coding schemas that would change based on the particular recipient of the email (e.g., client, customer, supervisor). However, early discussions amongst the raters indicated that there would be no substantial difference in the effectiveness of the behavior based on whom the email was directed



to. Based on this determination the raters chose to construct a single coding schema for all email responses disregarding who the email was written to.

## ***2. Behaviors coded using coding schema***

Prior to coding, the coders received training on how to interpret and apply the coding schema. During this training coders were given an applied practice task in which they independently reviewed and coded an example constructed-response. During this practice the coders discussed and calibrated their codes by justifying the behaviors they coded and coming to agreement. This practice was included as part of the training based on research, which has shown improved consistency between raters when frame-of-reference training and calibration between raters are a part of the methodology (Cash, Hamre, Pianta, & Myers, 2012; Sulsky & Day, 1992)

Two waves of coding occurred for this study. Wave 1 occurred in the fall, coding all of the available participant responses at that time. In the spring a new set of participant responses became available and were analyzed as part of Wave 2. Three coders completed each simulation in Wave 1 and four coders completed each simulation in Wave 2. The groups of coders independently reviewed the participants' constructed-responses and used the coding schema to record how many times each behavior occurred. An example of a coded participant response for the high-level managers' simulation can be found in Table 2.

After independently coding the responses, the coders met to review, discuss and come to consensus. The inter-rater agreements of the coded behaviors (i.e., prior to consensus) were examined for each simulation and are reported in Table 3. Prior to consensus, the coders tended to show the most agreement regarding category 6 (i.e., evaluated the situation). Agreement values were generally acceptable and appeared lowest for the group of coders for the Wave 2 high-level manager simulation. This lower score for agreement may suggest greater individual

differences between coders pre consensus. It is important to note that a fully automated score would eliminate these differences by removing human judgment; however, a partial automated score is still dependent on some level of human judgment.

The final consensus meeting recorded the number of times each participant engaged in a particular behavior. The aggregated frequency of each behavior for both ACs is provided in Tables 4 and 5. Some behaviors were included on the coding schema, but not recorded by the coders. Table 4 shows that for the mid-level managers' simulation two behaviors never occurred (i.e., behaviors 6b "Evaluated the situation and implied or stated false, misleading, or deceitful information," and 8c "Showed positive recognition toward a person or relationship by discussing the investment, commitment, time, or effort that has already been put in"). Table 5 shows that all behaviors included on the coding schema were recorded as having occurred in the participant constructed-responses for the high-level managers' simulation. Some participants sent multiple emails in response to the simulations. Behavior frequencies were aggregated across all emails for a participant, meaning one-time behaviors (i.e., 1a. Provided greeting) could occur multiple times for a single participant.

### ***3. SME effectiveness ratings and links to original competencies***

The coding schemas were presented to eight SMEs in an anonymous survey (i.e., four SMEs for each simulation). SMEs were recruited from the assessor staff of the developmental AC that provided the archival data. SMEs were invited to take this survey and were compensated with a \$5 Starbucks gift card for giving their expert opinion on two tasks. In the first task, the SMEs were asked to review the list of behaviors on the coding schema and link each behaviors to the competencies used in the current AC, see Table 1. SMEs were asked to indicate the strength of the link between the behavior and the competency on a scale of 1 (strong and clear

link), 2 (somewhat link that may be situation dependent), and 3 (weak or vague link but worth noting). The high-level manager AC simulation did not measure the competency Change, so this competency was not included in the survey for the high-level manager simulation.

In the second task, SMEs rated the effectiveness of each behavior in the coding schema on a scale of -5 (ineffective) to 5 (effective). The average effectiveness rating for each behavior for the coding schema can be found in Tables 4 and 5. These effectiveness ratings were averaged to determine the weight of each behavior on the coding schema. The question was also raised on how to verify the legitimacy of this simulation as an indicator of leadership. It was determined that a 1-item scale would be included to ask SMEs whether they believed any response to this simulation was better than no response. The SMEs were asked to rate how necessary an email response to this simulation was for indicating good leadership on a scale of 0 (not necessary at all) to 10 (very necessary). The mid-level managers' simulation had an average necessary rating of 9.0, and the high-level managers' simulation had an average necessary rating of 8.5. These values show that in both simulations providing a response email is necessary for indicating good leadership behavior. This confirmed the importance of taking action on this simulation.

The internal consistency and intra-class correlation (ICC) for the effectiveness ratings between SMEs was evaluated to determine if the SMEs had agreement on which items were generally effective and ineffective. The SMEs demonstrated a high level of agreement for both the mid-level managers' simulation ( $\alpha = .93$ ,  $ICC = .75$ ) and the high-level managers' simulation ( $\alpha = .92$ ,  $ICC = .75$ ). These indicate that the SMEs highly agreed on which behaviors were effective and which were ineffective. Consequently, differences in the effectiveness ratings were associated with differences in the behaviors and not based on differences amongst the SMEs. The complete survey that was presented to SMEs can be found in Appendix A.

#### *4. Created automated scores*

An automated scoring algorithm was created using the occurrences of behaviors on the coding schema and the effectiveness ratings from SMEs. This algorithm weighted the number of behavioral occurrences by the average effectiveness rating from SMEs. Because item 9a (spelling and grammatical errors) occurred far more frequently than other behaviors, this behavior was recoded using standardized z-scores for the number of occurrences. This was to avoid overweighting spelling and grammar issues that occurred very frequently for some participants (canceling out their entire simulation score in some situations). Since grammatical errors have the potential to occur more frequently than the other behaviors, this adjustment allowed this behavior to be more equally weighted in comparison to the other behaviors. Since there is no clear correct way to adjust this behavior, a relative standard logic was applied (i.e., z-score). The behavior count was recoded based on the z-score for grammatical error (i.e., a new count of -2 if the z-score was between -2 and -1, a new count of -1 if the z-score was between -1 and 0, a new count of 0 if the z-score was 0, a new count of 1 if the z-score was between 0 and 1, a new count of 2 if the z-score was between 1 and 2). None of the participants had a z-score greater than +/-2. It was determined that a negative occurrence on this behavior (i.e., -1 or -2) conceptually made sense in this particular situation because participants would receive a positive score if they had fewer errors than the average number of errors. This is because the negative occurrence would be weighted by the negative effectiveness rating from SMEs, which would produce a positive score. Thus, the algorithm not only penalizes participants with more grammatical errors than the average, but also benefits participants with fewer grammatical errors than the average.

When evaluating the survey results, I found less agreement among SMEs on the competency links than I had anticipated. I expected more agreement because the behaviors the raters derived from the constructed-responses were considered discrete and concrete rather than abstract and situational. I believed these concrete behaviors would be interpreted very similarly by the SMEs. The variation in SME responses raised the question of what information would be lost by dropping behaviors that were no longer considered distinct by the SMEs. Rather than discounting this level of variation as contamination, I believe that exploring the value of the variation addressed my original research question. When evaluating a constructed-response with an automated scoring method, can enough behavioral variance be accounted for to allow the automated score to predict the original score? Ultimately I am looking to understand how limiting behavioral variance impacts the predictability of the simulation score. By including these variations in behavior, I add an additional element to the study (i.e., degrees of behavioral variation in the automated score). In the original scoring method potential behavior variation is unlimited. However, when applying a limited list of behavioral options (i.e., the coding schema), I am introducing restrictions in the behavioral variations I am able to evaluate. By layering this into two different levels of restriction I can further investigate how increasing or decreasing restriction influences the scoring outcomes (i.e., how well the automated score predicts the original score).

To utilize the discrepancies between SMEs, I decided to test Hypothesis 1 on two automated scores, with the intention of comparing both automated scoring methods to the original score. These two automated scores include a complete (liberal) and condensed (conservative) scoring method. The liberal scoring method would include as much of the information from SMEs (e.g., all links made between behaviors and competencies) as possible.

The conservative scoring method would apply several exclusion criteria to only include behaviors on which there was high agreement among assessors. The intention of the more inclusive model (i.e., the liberal automated scoring method) is to be comprehensive and to more closely represent the internal process an SME uses in making scoring decisions. It is believed that the liberal scoring method will more closely resemble the original score because it provides more freedom in the links between behaviors and competencies and with the interpretations of the effectiveness ratings. This desire to not limit the automated score is similar to what some researchers have reported, that limiting the understanding of an assessor's cognitive processes oversimplifies the interpretation and results of these processes (Lord, 1985). The liberal scoring method is more inclusive and allows me to more completely capture how SMEs are interpreting performance from the behaviors. These discrepancies between SMEs capture variations in interpretation, which will be eliminated in the conservative scoring method. The conservative scoring method applies exclusion criteria to the coding schema so that only the most discrete and clearly identifiable behaviors are included in the scoring algorithm.

The justification for using a more conservative set of criteria was that the disagreement between SMEs could indicate a limitation of the behavior, meaning a behavior will not clearly distinguish good performance from poor performance if the SMEs cannot agree on how the behavior should be classified (i.e., linking to behaviors to the same competency). Because the conservative scoring method applies exclusion criteria, it is more similar than the liberal method to more concrete assessments (e.g., SJTs and multiple-choice assessments) because the behaviors that were accounted for and scored were limited and had more discrete and clearly observable behaviors. Although these findings will not generalize to selected-response assessments because

they are taken from a constructed-response set of data, they will provide evidence concerning limited sets of behaviors as they compare to detailed sets of behaviors in scoring methods.

All behaviors from the coding schema were used in the scoring algorithm to calculate the liberal automated Overall Score. The liberal automated score for each competency included all behaviors that were linked to that specific competency by SMEs with an average score of 2 or less (“somewhat link that may be situation dependent”). The average SME effectiveness ratings were used to weight the behaviors in each of the algorithms. To calculate the conservative automated score, two exclusion criteria were applied to the behaviors based on decisions made by the SMEs. In the conservative automated score not all behaviors from the coding schema were included in the scoring algorithm used to calculate the Overall Score. Behaviors were excluded from the conservative scoring method if (1) the behavior was not linked to a competency with an average score of 1 (strong and clear link) or (2) the average effectiveness ratings had high variation (i.e.,  $SD > 3$ ). For the first exclusion criterion, behaviors were removed if all of the linked competencies had an average greater than 1 (strong and clear link). This exclusion was based on the belief that if a behavior did not clearly link to a competency then including that behavior would likely introduce contamination into the scoring method, limiting the usefulness of indicating performance. For the second exclusion criterion behaviors were removed if the effectiveness scores did not have high agreement between SMEs (i.e.,  $SD > 3$ ). This was to prevent error caused by the SME's different interpretation of the items effectiveness. The justification for this was that if the SMEs greatly disagreed on how effective a behavior is then the interpretation of that behavior must not be clear, making it a potential source of contamination.

Table 4 shows that for the mid-level managers' simulation, 23 of the original 46 behaviors on the coding schema were eliminated with the conservative scoring method. The non-italicized behaviors depict the final set of behaviors that were used in the scoring the automated conservative scores. Table 5 shows that for the high-level managers' simulation, 20 of the original 38 behaviors on the coding schema were eliminated with the conservative scoring method. Behaviors were not excluded if they had a frequency of 0 amongst all participants (i.e., the behavior did not occur from any participant). It was determined that exclusion would be pointless since the behaviors would naturally fall out of the scoring algorithm (i.e., behaviors occurring 0 times multiplied by any weight is 0 added to the score). Additionally, these items could still potentially occur given another set of participants, thus elimination could exclude potentially valuable behavior variations. Participants who did not provide a response to the simulation were removed from the dataset and no automated score was calculated for them.

##### ***5. Final scores calculation and data analysis***

The composition of the automated competency scores will differ based on the scoring method (i.e., conservative or liberal). The liberal scoring method will include more behaviors from the coding schema than the conservative scoring method. Tables 6 – 16 depict the different compositions of the automated competency scores based on the different scoring methods. The top halves of the tables show the behaviors included using the conservative scoring method (i.e., with the exclusion criteria applied). The bottom halves of the tables show the additional behaviors from the coding schema that are included in that competency when the liberal scoring method is used. On the right side of the table is a summary area showing all of the other competencies that particular behavior also linked to. A “C” in the summary area indicates that the link was made at the conservative level (i.e., an average score of 1 [strong and clear link]).



An “L” in the summary area indicates that the link was made only at the liberal level (i.e., an average score of 2 [somewhat link that may be situation dependent] or less). This was included to show how much overlap a behavior has between the different competencies. A behavior that links to few or no other competencies more clearly indicates that specific competency than behaviors that link with multiple other competencies. The gray highlight in the summary area indicates which automated competency score the table represents.

Tables 6 – 16 included several interesting points, which are outlined below. First, the composition of the competency scores was not consistent between the two ACs. For example, the competency Talent for the mid-level managers’ simulation using the conservative scoring method included three behaviors:

- 2b. Expressed dissatisfaction, regret, disappointment, or apologized;
- 2c. Expressed empathy, considered another’s feelings or related own feelings to others, and;
- 4d. Provided or offered another detailed advice, direction, or information about something.

However, the composition for that same competency for the high-level managers’ simulation using the conservative scoring method included three different behaviors:

- 8a. Showed praise;
- 8b. Discussed value or importance, and;
- 8d. Discussed the potential future success or the trust, faith or confidence in it, dedication to it.

Given that the coding schemas were only slightly different and that the competencies were the same for both ACs, I expected a greater consistency between the behaviors that linked

to the automated competency scores. This indicates that the SMEs interpreted different meanings from the competencies based on the AC simulation that a participant was presented with.

Another point of interest is how many other links are made to different competencies. For both the liberal and conservative scores, many behaviors were linked to more than one competency. This implies that the behaviors from the coding schema are not distinct indicators of the competencies but instead share variance between the competencies. This indicates that the way the behaviors are written they involve some degree of construct overlap, meaning we are measuring many of the same things with the behaviors from the coding schema. A final point to note is that for some of the competencies, the conservative criteria has a large impact on the numbers of behaviors that are accounted for, whereas for other competencies there appears to be less impact. For example in the high-level managers' simulation for the competency Talent, only 10 behaviors are added using the liberal scoring method. However, for the mid-level managers' simulation for the competency Communication, 35 additional behaviors are accounted for in the liberal scoring method that aren't included in the conservative scoring method. The primary effect shown here is that the different scoring methods have a very different impact on the composition of the competencies. Some competencies, such as Change, show less change between the conservative and liberal scoring methods than other competencies, such as Mission and Values Leadership, which showed much greater change.

To calculate the automated Overall Score, all behaviors that were linked to a particular competency were each multiplied by their weights and added together to create the automated competency scores. Thus each participant had a set of original scores and two sets of automated scores. The original scores consisted of either six or seven scores (i.e., one Overall Score with six competency scores for the mid-level managers and one Overall Score with five competency

scores for high-level managers). The two sets (i.e., conservative and liberal) of automated scores also consisted of either six or seven scores (i.e., one Overall Score with six competency scores for the mid-level managers and one Overall Score with five competency scores for high-level managers). The high-level managers had one less competency score because the competency Change was not evaluated for high-level managers in the original AC simulation. The original scores were regressed on the automated scores to determine if the automated scores significantly predict the original human assessed scores. The results of these analyses are as follows.

## RESULTS

A basic power analysis was conducted using an online multiple regression power analysis calculator. The results indicated that to detect a medium effect, a minimum of  $n = 67$  was needed. From the two ACs, the mid-level managers' simulation had an  $n = 84$  and the high-level managers' simulation had an  $n = 45$ . Based on this, we achieved our target sample size with the mid-level managers' simulation but not the high-level managers' simulation. Unfortunately, the high-level managers AC had only 50 participants in total and five of those participants were removed from the sample due to missing scores or responses. Because of this it is difficult to determine why the results of the two ACs differ. I cannot be certain if differences are attributed to managerial level, simulation content, or the inability to depend on the high-level manager simulation results due to the small sample size. Because of this, I will first present the mid-level managers' results as the primary findings for this study, based on the dependable sample size. I will also include a summary of the high-level manager simulation results as they relate to our findings. This summary is primarily a point of interest and consideration for future research. However, those findings should be considered with skepticism based on the sample size.

### **Participants**

A total of 129 participants completed the simulations evaluated in this study. Of those participants, 45 completed the high-level managers' simulation and 84 completed the mid-level managers' simulation, as shown in Table 17.

### **Descriptive Information**

The means and standard deviations were compared for each simulation across the three scoring methods (i.e., original score, liberal automated score, and conservative automated score).

Table 20 provides the descriptive information for word count and each scoring method for the mid-level managers' simulation. For the original AC, mid-level managers obtained the highest average score on the competency Talent ( $M = 2.68, SD = 0.52$ ) and the lowest average score on the competency Mission and Values Leadership ( $M = 1.43, SD = 0.61$ ). For the conservative scoring method, the highest average score was on the competency Relationships ( $M = 1.49, SD = 1.08$ ) and the lowest average score was on the competency Strategy and Results ( $M = 0.43, SD = 0.70$ ). For the liberal scoring method, the highest average score was on the competency Change ( $M = 1.24, SD = 0.70$ ) and the lowest average score was on the competency Relationships ( $M = 0.95, SD = 0.42$ ). Surprisingly, the highest average score using the conservative scoring method (i.e., Relationships) was the lowest average score using the liberal scoring method. Similarly, the competency Talent went from being the second lowest score using the conservative scoring method ( $M = 1.06, SD = 1.89$ ) to the second highest score using the liberal scoring method ( $M = 1.06, SD = 0.53$ ). Interestingly, this average score for Talent did not change, but instead the average scores for all of the other competencies dropped when the additional behaviors were included as part of the liberal scoring method. The general trend appears that using the conservative scoring method produces on average a higher score at the competency level and for the Overall Score ( $M = 1.17, SD = 0.63$ ) than the liberal scoring method ( $M = 0.94, SD = 0.42$ ). A second general trend is that applying the conservative scoring method criteria appears to impact the competencies very differently, having a large impact on the average score of some competencies, such as the competency Relationships, and almost no impact on other competencies, such as the competency Talent. In Table 21 the means and standard deviations for the high-level managers' simulation are provided. This information was consistent with the mid-level managers' simulation in that the different automated scoring

methods had a large impact on some competencies and a small impact on others. As was the case with the mid-level managers' simulation, the competency Relationships showed the greatest change between the conservative scoring method ( $M = 0.58, SD = 0.87$ ) and the liberal scoring method ( $M = 1.60, SD = 0.84$ ). For the high-level managers' simulation the trend of average scores between scoring methods was not consistent with the trend from the mid-level managers' simulation, on average scores were higher using the liberal scoring method.

### **Hypothesis Tests: Hypothesis 1**

Hypothesis 1 stated that a substantial portion of the variance in behavior would be retained in the automated score so that I could significantly predict the original score. This was true for the majority of the competency scores and overall scores for the mid-level managers' simulation, where the automated scores explained a significant portion of the variance.

#### **Hypothesis 1: Mid-level managers**

Hypothesis 1 was tested using simple linear regression. The original AC scores (i.e., the Overall Score and six competency scores) were regressed on the automated scores using the two different scoring methods (i.e., liberal and conservative) to determine if the automated scores predicted the original AC scores. The results for the mid-level managers' simulation support Hypothesis 1. Table 22 shows that the Overall Scores for both automated scoring methods significantly predicted the original Overall Score. For the liberal scoring method, all but one of the competency scores (i.e., Mission and Values Leadership) significantly predicted the original competency scores. For the conservative scoring method, three of the competency scores (i.e., Strategy and Results, Relationships, and Change) significantly predicted the original scores. For the conservative scores, the greatest amount of variance was explained in the original score for the competency Relationships ( $R^2 = .20, CI = .06, .34$ ). For the liberal scoring method the

greatest amount of variance was explained by the competencies Communication ( $R^2 = .30$ ,  $CI = .14, .44$ ) and Relationships ( $R^2 = .30$ ,  $CI = .14, .43$ ).

### **Hypothesis 1: High-level managers**

The findings for the high-level managers' simulation had some consistency with the findings from the mid-level managers' simulation in support of Hypothesis 1. Table 23 shows that three of the competency scores for the conservative scoring method significantly predicted the original scores (i.e., Talent, Relationships, and Communication). Additionally one of the competency scores for the liberal scoring method significantly predicted the original score (i.e., Mission and Values Leadership). Inconsistent with the findings from the mid-level managers' simulation, neither of the Overall Scores for either scoring method significantly predicted the original Overall Scores. The largest amount of variance from the original competency scores was explained by the competency Relationships ( $R^2 = .12$ ,  $CI = .00, .30$ ) for the conservative scoring method and by the competency Mission and Values Leadership ( $R^2 = .13$ ,  $CI = .00, .31$ ) for the liberal scoring method.

### **Hypothesis Tests: Hypothesis 2**

Hypothesis 2 stated that word count would significantly predict both the original and the automated scores. This was true for the mid-level managers' simulation and partially true for the high-level managers' simulation.

### **Hypothesis 2: Mid-level managers**

Linear regression was used to investigate whether word count alone would significantly predict the original and automated scores. The three scoring methods were regressed on word count to determine if word count alone could significantly predict the Overall Scores, see Table 24. Results indicated that word count could be used to significantly predict all three scores for

the mid-level managers' simulation. The relationship between word count and the original Overall Score and both automated scores was positive, see Table 18. This provided support for Hypothesis 2, as there was a main effect of word count on Overall Score for all three scoring methods in the mid-level managers' simulation.

### **Hypothesis 2: High-level managers**

Results for the high-level managers' simulation were consistent across both automated scores; word count significantly predicted both the conservative and the liberal automated scores. However, word count did not predict the original Overall Score for the high-level managers' simulation. The means and standard deviations for word count are shown in Tables 20 and 21. These values indicate that the high-level managers on average wrote more than the mid-level managers and word count had more variance for the mid-level managers' simulation ( $M = 90.80$ ,  $SD = 128.68$ ) than for the high-level managers' simulation ( $M = 234.84$ ,  $SD = 116.40$ ). The lack of variance in the high-level managers' values for word count can assist in explaining the non-significant result. Having little variation between participants in word count may have made it challenging to use that variable to make meaningful predictions, because there are few differences between participants to distinguish them.

### **Hypothesis Tests: Hypothesis 3**

Hypothesis 3 stated that word count would moderate the relationship between the automated score and original score so that the automated scores for higher word count responses would better predict the original scores. For mid-level managers a moderating effect was found for word count. However, lower word count responses predicted better than higher word count responses. Hierarchical models were performed at the Overall Score level rather than at the competency score level because it was not feasible to calculate word count within competencies .



### **Hypothesis 3: Mid-level managers**

Hypothesis 3 was tested using a hierarchical regression model to determine the moderating effects of word count for both the liberal and conservative automated scores. My previous regression model for word count alone indicated that word count significantly predicted the Overall Scores for the original and both automated scores. The hierarchical regression model was used to further investigate this relationship and determine if word count explained any of the variance in the original Overall Score beyond that which was predicted using the automated score, see Table 25. The results for the mid-level managers' simulation showed that word count did not significantly predict the original Overall Score beyond what was explained by the automated Overall Score. This was true for both the conservative and the liberal automated scoring methods. The interaction term (i.e., automated score\*word count) for both automated scoring methods accounted for significant additional variance explained in the original score beyond the automated scores. The significant negative interaction term provides evidence in support of Hypothesis 3, showing that word count moderates the relationship between the original Overall Score and the automated Overall Score for mid-level managers.

### **Hypothesis 3: High-level managers**

Results for the high-level managers' simulation were not consistent with the mid-level managers' simulation, as shown in Table 26. The automated scores alone did not significantly predict the original scores. For the liberal automated score only, word count did account for a significant portion of the variance beyond the automated score. The interaction term was not significant for either the conservative scoring method or the liberal scoring method.

### **Direction of interaction**

It was also predicted that this interaction term would show high word count responses predicted the original scores better than low word count responses. To view these relationships, word count was converted into a dichotomous measure by splitting responses at the mean for word count. Participants with a word count above the mean were labeled as “High word count” and participants with a word count below the mean were labeled as “Low word count.” All participants were plotted to visually compare the regression trend of low versus high word count, see Figures 4 - 7. The scatterplots of the regression trends indicate a disordinal interaction between the Overall Score and word count for the mid-level managers’ simulation, Figures 4 - 5. This indicates that mid-level managers’ responses with low word count were a better predictor of the original Overall Scores than high word count responses. The scatterplots for the high-level managers’ simulation depict a non-interaction term for word count and the relationship between original and automated scores, as shown in Figures 6 – 7. On average, high word count responses received higher original scores than low word count responses.

### **Comparing Scoring Methods**

Using two different automated scoring methods raised the question of which method better predicted the original Overall Scores (see Figures 8-9). For the mid-level managers, the variance explained by the liberal scores in predicting the original Overall Score ( $R^2 = .32$ ,  $CI = .16, .45$ ), was greater than the variance explained by the conservative scores ( $R^2 = .27$ ,  $CI = .12, .41$ ). Despite this, the regions of the confidence intervals overlap substantially. This indicates that neither scoring method can be identified as better than the other because they both potentially account for the same amount of variance. In addition to testing the hypotheses, I

conducted several exploratory analyses to further investigate these findings and to better understand the relationships of the different scoring methods.

## **Exploratory Analyses**

### **Inter-correlations within scoring methods**

Examining the inter-correlations of each scoring method (i.e., original, liberal automated, and conservative automated) was important to compare how the original competencies related and to determine if this was similar to how the competencies related with the automated scoring methods.

#### ***Original scoring method***

The pattern of the inter-correlations for the original scoring method is consistent across both simulations (i.e., mid-level managers and high-level managers). The original scores indicate a positive manifold (i.e., all competencies are positively and significantly related to one another). The competency least consistent with the pattern is Mission and Values Leadership, which had the lowest and greatest number of non-significant relationships. This highlights a common criticism of AC competencies: that the competencies evaluated are small variations of the same underlying latent construct, where measures link better to AC simulations and context specific dimensions and do not link well to dimensions independent from the AC (e.g., Sackett & Dreher, 1982). This causes a particular problem in my study, because my outcome variable is the original scores from the AC, which here are shown to overlap substantially. However, an automated scoring method by definition necessitates discrete observable behaviors, so that the behaviors can be evaluated without human judgment or interpretation. Thus, I am attempting to use a more discrete set of measures to predict a set of largely overlapping measures.

### ***Conservative automated scoring method***

The inter-correlations between competencies substantially dropped in both the mid-level and high-level managers' simulations using the conservative automated scoring method. Neither simulation presents a pattern of positive manifold; only some of the competencies are significantly inter-correlated and several competencies are negatively correlated. Interestingly, Mission and Values Leadership (which was correlated the least with other competencies for the original scores) correlated the highest with other competencies using the conservative scoring method. Additionally, it had the most overlap with behaviors from the coding schema.

### ***Liberal automated scoring method***

The inter-correlations for the liberal automated scoring method for both the mid-level and high-level managers' simulations indicate a positive manifold similar to that seen in the inter-correlations for the original scores. This pattern of inter-correlations is a stronger indicator of a positive manifold than the pattern seen for the original scores. In general the inter-correlations are stronger for the liberal scoring method than they were using the original scoring method.

The patterns of inter-correlations provide interesting evidence that automating the scoring system using a liberal scoring method (that tends towards inclusion of more behavioral variation) strengthens the positive and significant inter-correlations of the competencies. However, in simplifying the automated scoring method so that only discrete behaviors are included (i.e., conservative scoring method) I no longer represent the original scores as well (i.e., does not indicate a pattern of positive manifold). However, I do begin to reach a level of distinction between competencies, which is often the goal when using automated scoring methods. These findings reiterate the construct validity problem common in ACs (Sackett & Dreher, 1982), where the evaluated competencies share large portions of variance. This indicates that as a more

stringent set of exclusion criteria is applied to a scoring method, the ability to distinguish competencies appears to increase.

### **Convergent and discriminant validity**

The concept of distinct measures was further investigated by looking at the average convergent (i.e., correlations of the same competency across different methods of measurement) and discriminant validity (i.e., correlations for different competencies measured using the same method) between each scoring method and the original AC simulation scores (Campbell & Fiske, 1959). The average convergent validity across correlations between the conservative automated scores and the original scores for the mid-level managers was  $r = .25$ . This was lower than the average convergent validity across correlations between the liberal automated scores and the original scores,  $r = .42$ . This indicates that on average the liberal automated scores converged more with the original scores than the conservative automated scores. This supports the assertion above that having a similar pattern (i.e., positive manifold) more closely relates the original and liberal scoring methods. These values were still lower than desired. When evaluating convergence I am looking to see that the same competency measures similarly across different scoring methods. As would be expected, the convergent value between the conservative and liberal scoring methods,  $r = .58$ , was relatively high, showing there is a large amount of overlap between the two automated scoring methods. This trend was not consistent in the high-level managers' simulation, where the conservative automated scores converged slightly better with the original score, average  $r = .27$  than the liberal automated scores did,  $r = .23$ . However, these convergent scores were still lower than I would anticipate. The trend was consistent in that the convergence between the two automated scoring methods was much higher,  $r = .51$ .

The average discriminant validity in the mid-level managers' simulation was lowest for the conservative automated scores  $r = .20$ . The discriminant validity for the original scores,  $r = .50$  and the liberal automated scores,  $r = .80$  were very high. All values for discriminant validity were substantially higher than I would like to see. These high discriminant validity values indicate that there is a large amount of overlap between the competencies in each scoring method. This indicates that although the conservative automated scores are more distinct than both the original scores and the conservative scores, none of our scoring methods clearly distinguish between different competencies very well. This trend was consistent in the high-level managers' simulation in which the conservative scoring method had the smallest discriminant validity,  $r = .27$ , but all values (original score,  $r = .45$  and liberal scores  $r = .78$ ) were relatively high.

In both simulations, convergence was higher than the discriminant value for the conservative automated scoring method, yet the opposite was true for the other two scoring methods (i.e., the discriminant values were higher than the convergent values). This provides preliminary evidence that shifting from a liberal automated scoring method to a conservative automated scoring method improves convergent validity (e.g., the ability to measure the same competency using different methods) and discriminant validity (e.g., the ability to discriminate between dissimilar competencies).

### **Exploratory Factor Analysis**

An alternate method for grouping the behaviors derived from the constructed-responses would be to conduct an Exploratory Factor Analysis (EFA) to determine the natural clustering of the behaviors that emerge from the data. This could assist in determining how discrete behaviors naturally cluster into separate factors. I decided to run an EFA using Mplus 7 statistical software

on the mid-level managers' simulation. I chose to not use the high-level managers' simulation data because of the small sample size and the problems with power that this would have introduced. It is important to note that two behaviors fell out of the EFA for having no variance (i.e., 6b and 8c). An oblique GEOMIN rotation was used in the EFA to allow for the possibility of correlated factors (Bandalos & Boehm-Kaufman, 2008). A parallel analysis revealed a maximum eigenvalue of 2.71, showing no more than three factors could be extracted from our data (Horn, 1965). Our goodness of fit test revealed that a three-factor structure was still not a good fit for our data  $\chi^2 = 1456.71, p < .01, RMSEA = .10, CFI = .41, TLI = .32$ . However, a three-factor model was an improvement over a two-factor model fit  $\Delta\chi^2 = 123.18, p < .01, \Delta CFI = .08$ . The poor fit of the three-factor model is likely attributed to the fact that many of the behaviors did not load on any of the three factors in the final model. The factor loadings for the three-factor model are shown in Table 27. Typically the next step would be to have SMEs evaluate these new factors by reviewing the behaviors and determining the latent construct between them (Haig, 2005). For example, from this new factor structure, factor three could indicate a particular leadership style that involves considerate and thoughtful motivations and actions. However, this data does not appear to be very amenable to a factor analysis methodology and another method, such as latent cluster analysis, may be considered in future iterations.

## DISCUSSION

The primary purpose of this study was to evaluate how much variation naturally occurs in a constructed-response from an AC simulation. Determining these natural behavioral variations allowed me to review and categorize the behaviors in a meaningful way to produce a set of automated scoring algorithms. Once the automated scoring method was constructed, I was able to determine that the automated scores significantly predicted the original AC scores generated by human assessors. However, this was with a moderate effect size meaning at this stage there is not enough evidence to support an automated scoring method being used interchangeably with a human assessed scoring method.

### **Implications of Research Findings**

The first part of the study set out to create a manageable set of behaviors from the constructed-responses to the AC simulation. This resulted in a coding schema of 46 behaviors for the mid-level managers' simulation and a coding schema of 38 behaviors for the high-level managers' simulation. Both sets of behaviors fell into nine broad categories. The three raters involved in the development of the coding schema were able to quickly reach agreement on decisions regarding the behaviors in the coding schema. Even when the coding schema was used by a new group of coders, strong agreement was generally consistent showing stable interpretation and application of the coding schema. This presents a clear answer to the research question and is consistent with findings from other qualitative studies, which have shown that complex constructed-responses can be categorized into manageable sets of behaviors (Schilling, 2006). These findings show that an exhaustive list of behaviors can be produced from the qualitative responses to a constructed-response AC simulation.



The second part of my study was to evaluate the ability of the automated scoring method to predict the original scores. My results showed that an automated score could significantly predict a human assessed score, meaning that under certain constraints a partially automated scoring method may have value in an AC simulation. Evidence was also found that for the mid-level managers' simulation, word count moderated the relationship between the original and automated scores. In general participants received higher scores if their response had a higher word count. However, the interaction showed that the automated score predicted better for participants whose responses had a lower word count. The direction of this interaction was the opposite of what was hypothesized. It appears that an automated scoring method can better predict the original score produced by a human assessor when fewer words are included in the response. This is consistent with previous research which shows accurate predictions can be made from extremely limited responses from participants (Oostrom, Born, Serlie, & van der Molen, 2012). There was no interaction effect for the high-level managers' simulation.

The final objective of this study was to determine whether one of the automated scoring methods (i.e., conservative or liberal) predicted the original scores better. For both mid-level managers and high-level managers, the regions of the confidence intervals for the effect size of the automated scores predicting the original scores overlapped substantially. This indicated that neither scoring method predicted better because they both potentially account for the same amount of variance. Despite designing very different scoring methods, it appears that the amount of behavioral variance accounted for in each method was statistically the same.

### **Outcome Measures**

This study became increasingly complicated when a group of SMEs were asked to review the coding schema to link the behaviors to the AC competencies and provide effectiveness

ratings. Despite the unexpected amount of variation between SME responses, these variations provided insight into the considerations and interpretations that SMEs typically make when evaluating behaviors in an AC simulation. The purpose of linking the behaviors back to the AC competencies was to enable analysis at the competency score level so that the outcome measures could be evaluated at both the Overall Score level and the competency score level. However, this poses an interesting situation as I am attempting to fit discrete and observable behaviors for an automated scoring system back into an abstract competency structure, which was intentionally avoided during development of the automated scoring structure. At least part of the variation in the results can be attributed to the process of linking the behaviors to the AC competencies due to the contrasting nature between the discrete automated scoring behaviors and the abstract competency definitions. By design these two systems were intended to be different, so using one as a predictor of the other introduces additional considerations.

The results of this study draw on comparisons between the automated scores and the original human assessed scores. Understandably, these two scoring systems are conceptually different, as discussed. However, one additional consideration is the assumption that the original ratings from the human assessors were accurate indicators of performance. It must be considered that there was potentially some error in the original human assessed scores that may have influenced the outcomes of this study. A high level of consistency was expected from the survey because the SMEs completing the surveys were the same individuals working as assessors in the ACs. However, the results of the surveys showed that the assessors viewed the competencies as having different behavioral compositions between the two ACs (i.e., assessors' linked different behaviors to the same competency for the different ACs). Based on this and other variations among SMEs on how to classify the behaviors, we might assume that the individual SMEs are

considering the competencies slightly differently. A consideration for future research would be a comparison between the participants' automated scores and later measures of job performance within the organization or another objective measure.

### **Content Restriction**

One of the primary benefits of this study is initiating research on the impact of content restriction on the scoring process within an assessment. Researchers have shown interest in looking at how much information is lost by restricting the set of possible responses participants have to a reduced number of response options which are often presented in multiple choice assessments and SJTs (Ryan & Greguras, 1998). It is important to note that although I discuss both constructed-response and selected response assessments in this study, results of this study do not generalize to selected-response assessments. This is because the participants from this AC used a constructed-response cognitive process for the simulation. However, this information begins to build evidence to address questions that have been posed by other researchers regarding the comparison of selected-response versus constructed-response options. Specifically, this study shows how placing constraints on the behaviors that are scored in an AC simulation impacts participant scores, as selected-response assessments place constraints on the behaviors available to participants.

Although the conservative automated score does not fully represent a selected-response assessment, it helps build evidence for future research on how content restriction (i.e., limiting options to a smaller set of responses) influences the outcome scores for participants. Previous research shows that limited response options can significantly predict the target outcome (Oostrom et al., 2012). Based on this, it was believed that the conservative automated score could still significantly predict the original score despite accounting for less behavioral variation.

This was found to be true in the mid-level managers' simulation. Additionally, the pattern of correlations showed that restricting content through the application of exclusion criteria influences the measured competencies to be more distinct (i.e., the pattern of correlations did not indicate a positive manifold as was seen in the inter-correlations for both the original scores and liberal automated scores). Interestingly, the liberal automated model better predicted (though not to a statistically significant extent) the competency scores from the original AC and looked more similar to the pattern of inter-correlations. This implies that a liberal scoring method may more closely replicate the original scoring method. These trends appear to indicate that more conservative automated scoring methods could produce more distinct measures between competencies and more liberal automated scoring methods may more closely replicate the outcome measures of the original human assessed scoring methods. However, future research needs to further explore the differences and applicability of different scoring methods to determine how they can be used to more reliably predict human assessor scores with a greater effect size.

Although these results indicate that differences between a liberal and conservative automated scoring method are not statistically significant at the Overall Score level, larger differences between scoring methods were observed for the competency scores. Some of the scores at the competency level were dramatically influenced by the stringent criteria applied in the conservative scoring method. This pattern of change between the two scoring methods seems to indicate that some competencies are highly influenced by stringent criteria, while other competencies are less influenced when this criteria is applied. Table 22 shows how the amount of variance accounted for in the Overall Scores overlap substantially between the conservative and liberal automated scores. However, at the competency score levels, several of the confidence

intervals just barely overlap such as Communication and Mission and Values Leadership. This indicates that at the competency level the two different scoring methods are accounting for different portions of the variance in the original scores.

This concerns the potential difference between including contamination or meaningful variance in the automated scoring methods. The intention of making behaviors more concrete was to reduce the amount of contamination, however in this exclusion process the behaviors were clearly reduced to the point where meaningful portions of behaviors were excluded. This is demonstrated by the competency Talent, which significantly predicted the original score using the liberal automated scoring method but not using the conservative scoring method, as seen in Table 22. In this particular situation it is clear that the eliminated behaviors removed important indicators of performance and not just behavioral contamination.

The finding that the liberal automated scores appeared to more closely represent the human assessed scores suggests that human assessors may include more information when determining their scoring processes, and may consider behaviors as contributing to several different competencies and not as distinct indicators of performance in one area or another. The idea that automated scoring methods can replicate human assessed scoring methods is not new the research literature. For example, Balogh et al., (2012) showed that automated scores using speech recognition software to evaluate oral reading fluency reliably predicted the human assessed scores. However, the underlying cognitive processes that go into these decisions that are made by the assessor during evaluation remain relatively unexplored. Other researchers have stated that there is a need to understand how different perceptions can lead to changes in cognition amongst raters when making evaluations (Kaufman et al., 2013). This preliminary evidence suggests that reducing the behavioral variation accounted for in an AC simulation may

impact the ability to evaluate some competencies and reduce the ability to predict a human assessed score with that information.

### **Future Research**

To capitalize on the variation in responses presented by SMEs, two different automated scoring methods were used to evaluate participant responses. My results consistently showed that these different scoring methods impacted the competency scores differently. Specifically, more information was lost between the conservative and liberal scoring methods for some competencies than others. Suggestions for future research would be to continue to explore the impact of content restriction on automated scoring systems specifically looking for patterns of impact. Future research would benefit from understanding the competencies that are most impacted by content restriction and what latent constructs are most susceptible to change when strict exclusion criteria are applied. Researchers are encouraged to look for systematic differences including constructs that are more resilient to content restriction in an automated scoring context.

### **Limitations**

There are a few important limitations for this study that I should point out. One limitation was that the algorithm use for the automated scoring method was prone to several sources of potential bias. The first potential source of potential bias was using the SMEs to rate behaviors on effectiveness. It appears that the SMEs used little variation in their ratings, tending towards rating behaviors as either “effective” or “ineffective.” Understandably, the SMEs found it difficult to rate the items without contextual information and clues. One SME wrote in their feedback, “Many of these linkages really depend on the situation. I could see it going several different ways.” It appears that some of these behaviors may have been interpreted and evaluated

differently had more context been provided. A second potential source of potential bias was using a simplified method for evaluating performance, in which the rules in the scoring algorithm did not change based on different situational constraints. The drawback to this was evident in a situation where scores were inflated for participants who displayed a single behavior several times. In one specific example a participant displayed behavior 3f, “Explained the details of the actions they should take or of the assistance that is needed” 11 times in their responses. The way the current algorithm is designed, more behaviors result in adding the same weight each time the behavior occurs, which in this situation drastically inflated the score. The scoring inflation caused the participant’s score to become an outlier, falling well outside the distribution of the other scores for that competency. An improved scoring algorithm would account for large anomalies by standardizing scores that were skewed or by continually diminishing the weight of a particular behavior each time it repeated. One consideration is that adjustments such as these would likely require a complex computer program to be involved in the automated scoring process.

### **Sample size**

This study intended to combine data across the two AC simulations. After reviewing the constructed-responses, it became clear that the two simulations were different enough to warrant separate coding schemas. In the current study the high-level managers’ simulation data does not have sufficient power to test many of the relationships of interest. Unfortunately splitting the data set had a large impact, limiting my ability to draw conclusions from the statistical analyses completed for the high-level managers’ simulation. Many of my findings that were inconsistent between the two ACs could potentially be attributed to the small sample size for the high-level managers’ simulation. Despite the small sample size, qualitative researchers have emphasized

the importance of sample quality and applicability to the study over sample size (Sandelowski, 1995). Morse (1989) reported that response redundancy (i.e., saturation) is a key factor in determining an adequate sample. When saturation is reached then the sample size is appropriately large for drawing conclusions and building theoretical inferences. This concept of purposeful sampling focuses on strategic recruitment of study applicable participants rather than emphasizing a large sample. Based on this I see that my study met guidelines within the research literature for the qualitative analysis piece. However, for the statistical analyses, the sample size for the high-level managers was inadequate and prevented me from reaching dependable conclusions.

### **Cognitive process**

Another limitation of this study was that I did not have a separate simulation with an altered cognitive process to truly test the comparison of a constructed-response versus a selected-response. Although many of the intentions that were set out in this study were accomplished, these results cannot fully represent the cognitive process of a selected-response assessment, since the participants constructed all of the responses used in this study. Klassen (2006) has reported that selected-response and constructed-response assessments lead to different outcomes and indicators amongst participants.

### **Resources for development**

A final consideration is given to the resources allocated to this project. A common assumption for automated scoring and the process of technology integration in general is that conversion to a new technology will be more convenient, cost effective and preferred by individuals within an organization. However, these assumptions are not always supported in the literature (Cherniack, 2011; Katz & Shapiro, 1986; Laupacis, Feeny, Detsky, & Tugwell, 1992)



and integrating new technologies can at times come with a great expense in time and resources. Organizations may neglect to consider three important cost considerations with regard to implementing an automated scoring process: start-up investment, sustained use and utility of the project. First, the start-up cost that a company invests in a technology integration project can be underestimated or overlooked. Understanding this, some researchers have focused their studies on how to reduce and more accurately estimate the start-up costs for particular types of projects (Ya-Chen, I-Wen, & Yi-Wen, 2009). Start-up costs can include several components (Tekin-Koru, 2009). These can include personnel (e.g., advising on the content, developing the process, testing the system, participating in the exercise), equipment (e.g., software, hardware), location (e.g., where the development will be done, where the piloting will be done), and communication / advertising. Second, organizations must consider the Total Cost of Ownership (TCO) that is associated with the project (Mouritsen, 2013). Having full ownership of the technology process adds additional costs that are separate from the start-up costs and may not be associated with off-the-shelf solutions offered by support services and contractors. The TCO includes infrastructure development, establishing and maintaining support systems and software / hardware costs. Organizations may overlook the fact that costs for developing a system are separate from costs for sustaining the system. Finally, companies need to consider the relative benefits of the new technological solution in comparison to the relative costs. Financial gains for an organization can be considered in a number of ways including profit, job performance, and retention (Schmidt & Hunter, 1998). This cost / benefit consideration (i.e., utility) can be expressed in more than monetary terms. There are circumstances in which a project will have utility (i.e., the benefits of the solution outweighs the investment) even if there was a greater cost for the project than the financial outcome. Research has documented that utility (i.e., practical economic value) can be

evaluated in a number of ways including practical values,, such as employee knowledge, company morale, organizational justice, and motivation (Schmidt & Hunter, 1998; Sitzmann, Casper, Brown, Ely, & Zimmerman, 2008).

Given this consideration, it has already been well established that the cost investment of ACs are quite steep, yet despite this companies continue to use them because of the high level of utility they return (Hoffman & Thornton, 1997). This research study has provided evidence that it is feasible to develop a computer system to automate the scoring process for this type of an assessment. The reason for doing this and the likely benefits of doing so must be weighed against the investments that will be made. In considering the resources allocated to this project specifically, much time and labor went into the development of the automated scoring method. In considering the utility for the automated scoring method for a single exercise of this AC, it seems unlikely that this process would have provided a cost-saving solution for an organization. At various times throughout the project, the personnel involved included 23 coders, 22 SMEs, and 1 project manager. Coders spent on average 40 hours each on this project, SMEs spent on average 1 hour, each and the project manager spent approximately 300 hours on this project. As a note, all of these resources are considered post training, meaning that time for the participants, developing the training and scoring the original exercise responses are not included in this tabulation. This account of resources exclusively accounts for the time involved to develop the automated scoring method. As a realistic summary, this was a large investment in resources for only one short exercise in a complex AC.

It is important to emphasize that various decisions will impact the costs associated with the project and the likely increase in gains. Factors that could potentially offset the cost of the project include simplicity of the assessment, experience of the developers, knowledge and use of

theoretically based models, exposure to automated scoring methods, having a pre-existing training, automated several exercises using the same method and having a higher throughput of participants. A recent study showed that about 66% of ACs assess fewer than 100 participants in a given year (Gibbons et al. 2013). In understanding utility from a perspective that is not strictly monetary, the value of automated scoring methods become more apparent. When accounting for additional gains that can be received and the various methods that can be used to offset the costs, automated scoring methods are potentially more feasible for organizations. AC administrators are advised to consider the high investment in time, cost and resources associated with automated scoring methods. Although it is anticipated that fewer resources would be needed once in operation, the painstaking process of development is not to be overlooked.

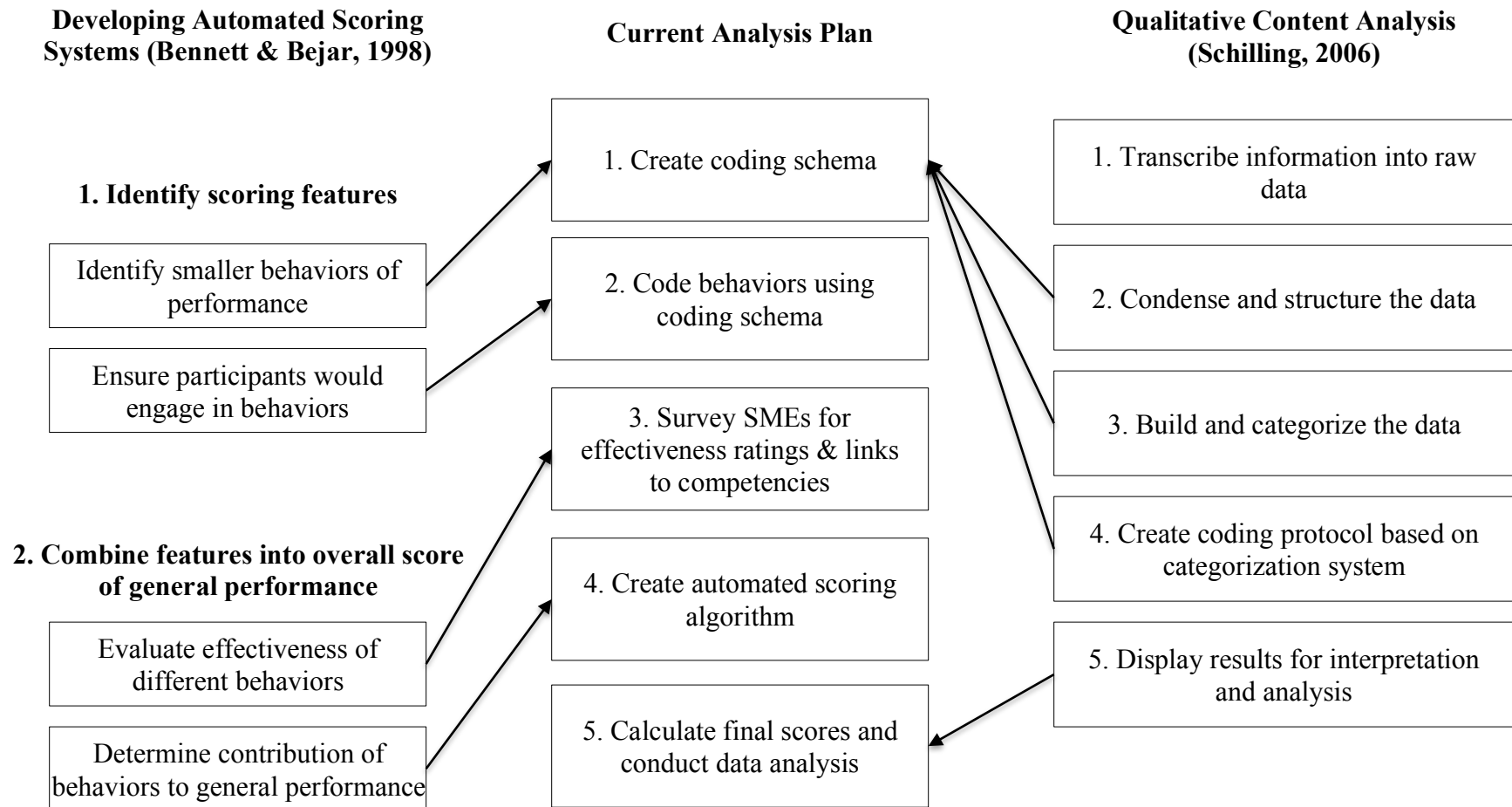
### **Contributions**

In addition to these limitations there were also several strengths from this study. This process included responses from a real AC from professionals in the education industry. In addition to this, our methodology followed a rigorous process outlined in multiple areas of the research literature (i.e., automated scoring and qualitative analysis) to form a process that fits the intended purpose (Bennett & Bejar, 1998; Schilling, 2006). Additionally, the process was well documented and included detailed record keeping of discussions and decisions that were made at each step in the process. Despite having a limited amount of power from one of my samples, I was able to obtain support for all of my hypotheses showing clear patterns of prediction. This suggests that an automated scoring method is a good topic to continue exploring for future use in an assessment center. Despite the smaller sample size in the high-level managers' simulation, there was partial support for several of the hypotheses from this sample as well.

## **Conclusion**

Overall the results of my study provided support for all of the hypotheses. The automated scores could be used to predict the original scores and a majority of the competency scores. Other researchers have similarly found that automated scoring methods can be used to significantly predict human rated scores (Balogh et al., 2012; Xi et al., 2012). Despite these findings, sufficient evidence was not found to support a claim that automated scoring methods can be used interchangeably with human assessed scoring methods. Although overall prediction was rather strong, the results were more inconsistent at the competency level, and some competencies tended to be more convergent than others. Although the general trends seem to imply that liberal methods replicate human assessor scoring methods and conservative methods may produce more distinct competency measures there is not enough evidence to determine at which point the method is distinguishing between contamination and variance accounted for. These findings do provide valuable evidence that automated scoring methods explain a moderate portion of the variance in the original human assessed scores. This implies that automated scoring methods are a valuable topic to continue considering in the assessment center context and that the methodology may one day be applicable after additional evidence is gathered in their support.

## FIGURES AND TABLES



*Note.* Step 1 from the qualitative content analysis method (Schilling, 2006) was not necessary in the current process because the content (i.e., email responses) were received in a typed format.

Figure 1. Combined methodologies to form current analysis plan

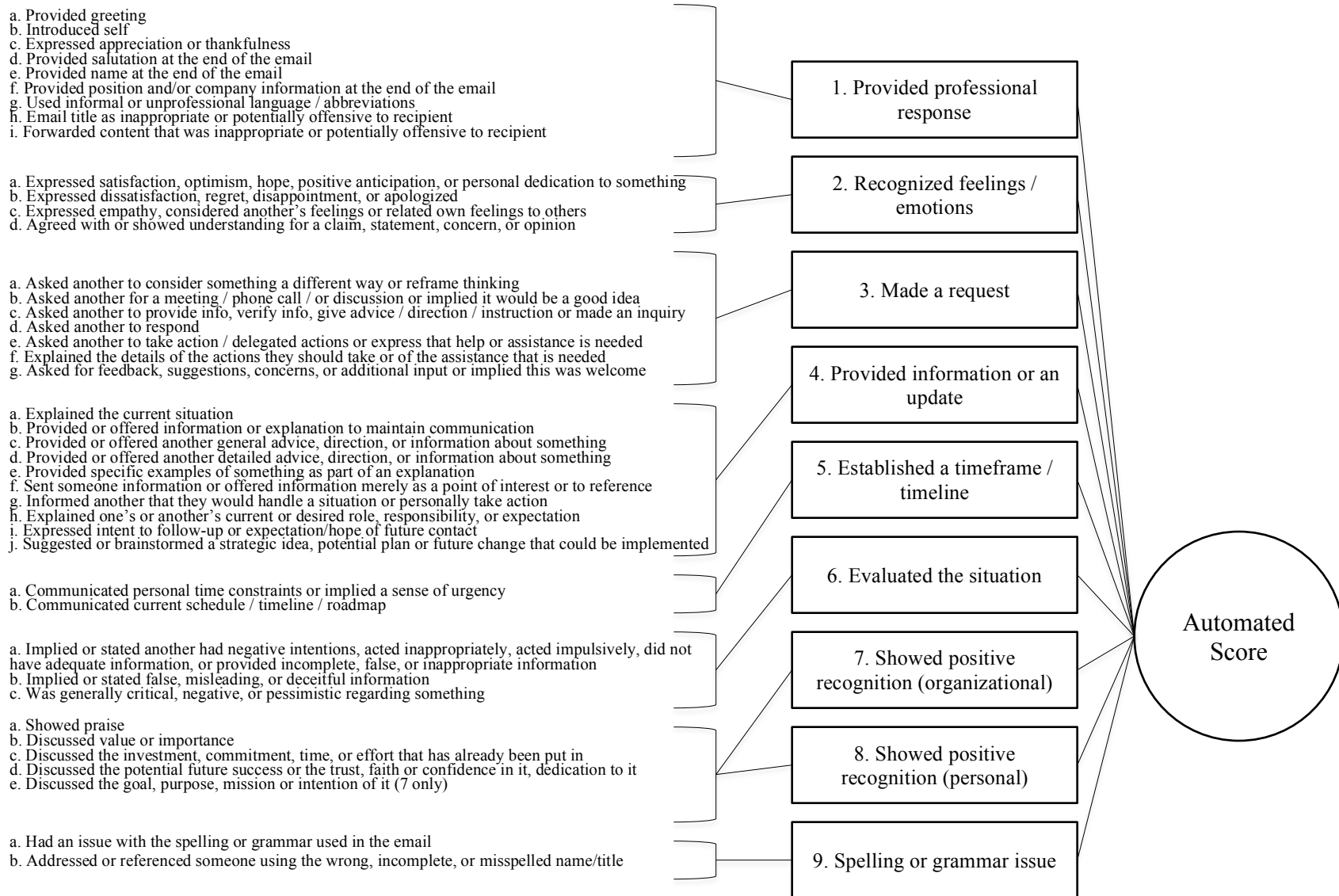


Figure 2. Coding *schema* for mid-level managers' exercise (i.e., customer complaint)

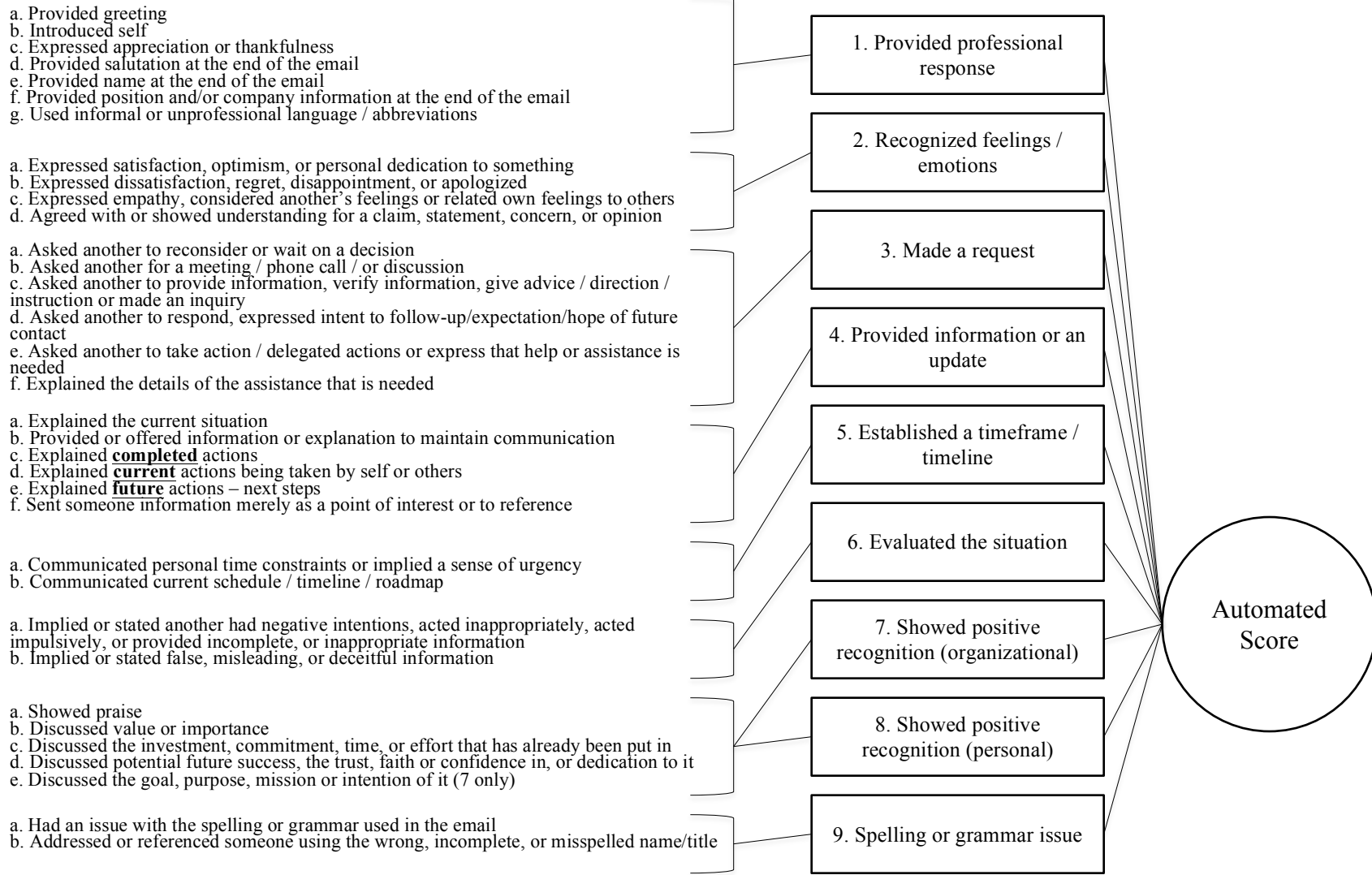


Figure 3. Coding *schema* for high-level managers' exercise (i.e., client concern)

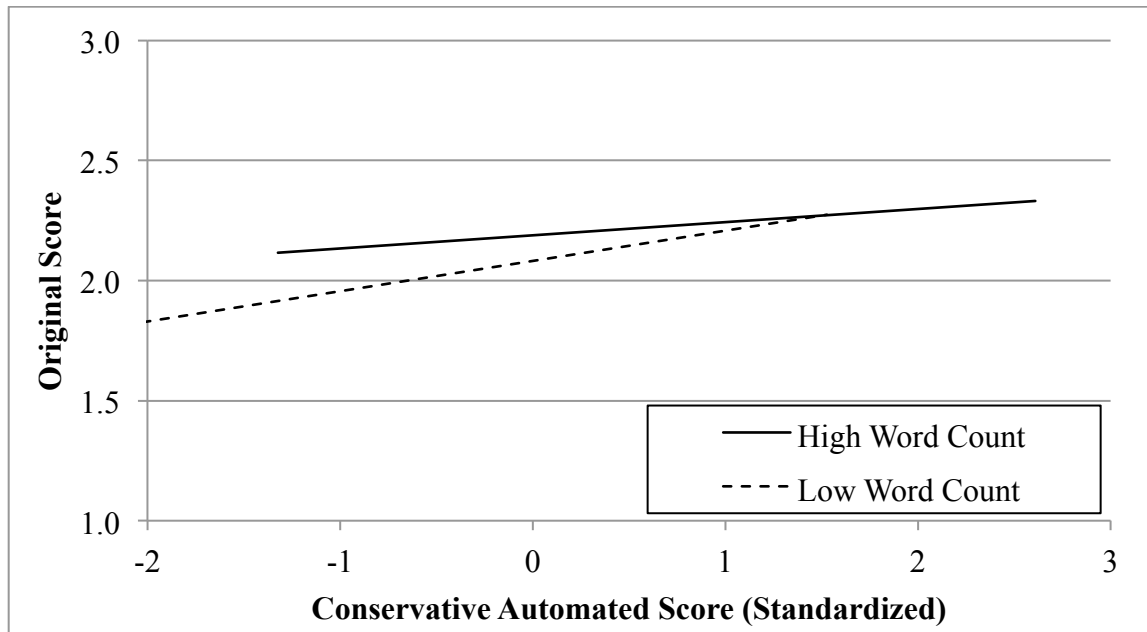


Figure 4. Original and conservative scores for mid-level managers by high and low word count.



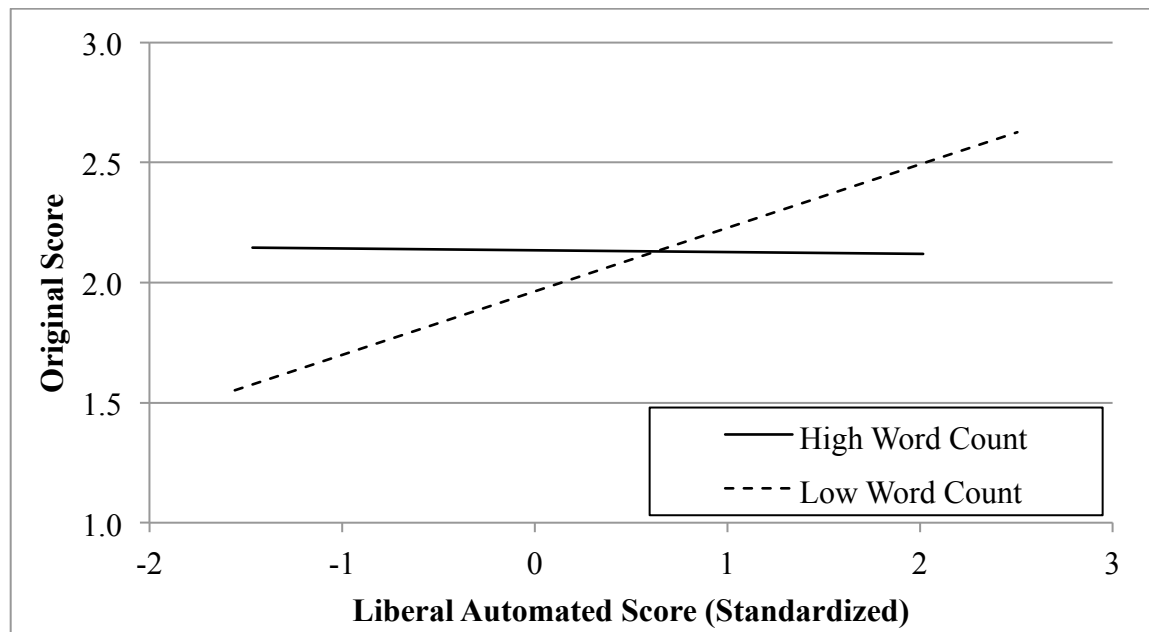


Figure 5. Original and liberal scores for mid-level managers by high and low word count.

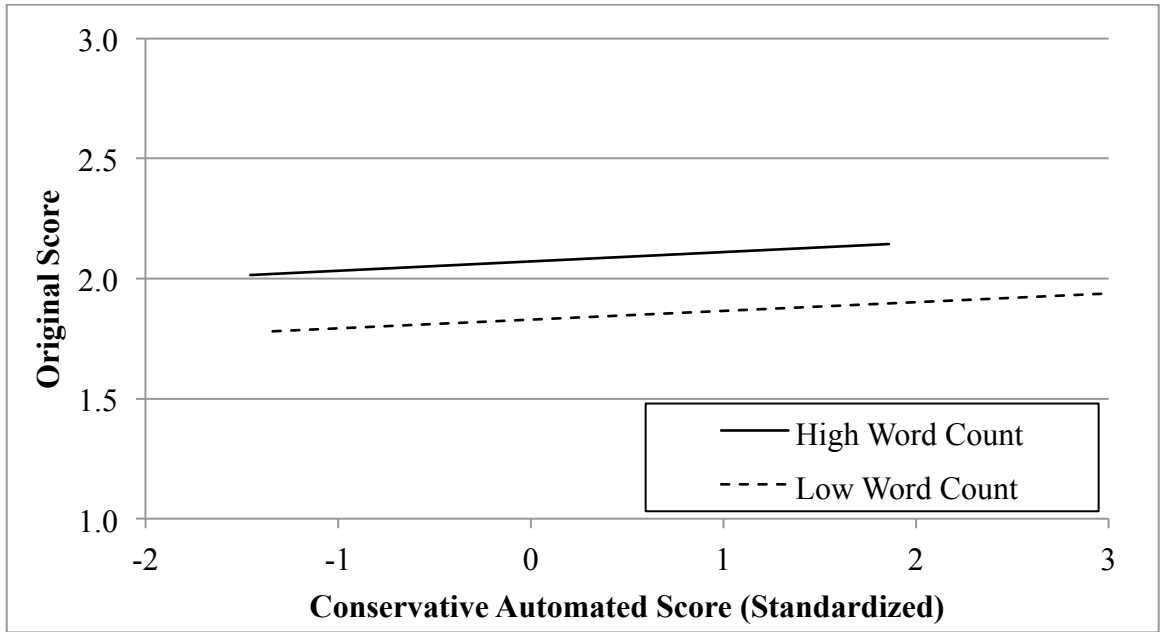


Figure 6. Original and conservative scores for high-level managers by high and low word count.

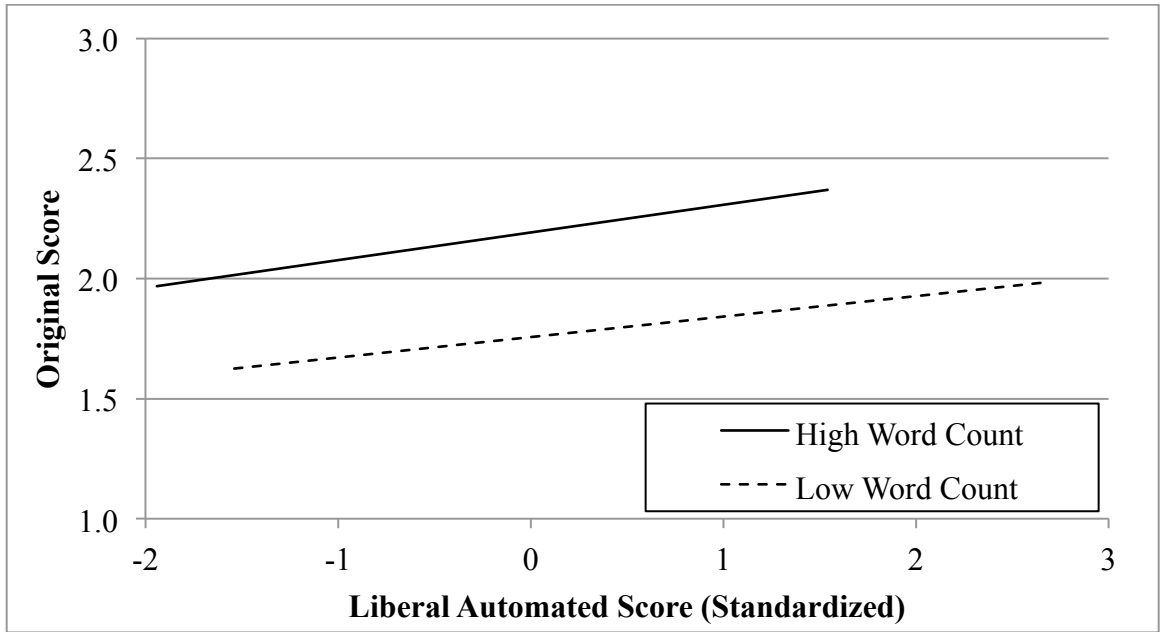


Figure 7. Original and liberal scores for high-level managers by high and low word count.

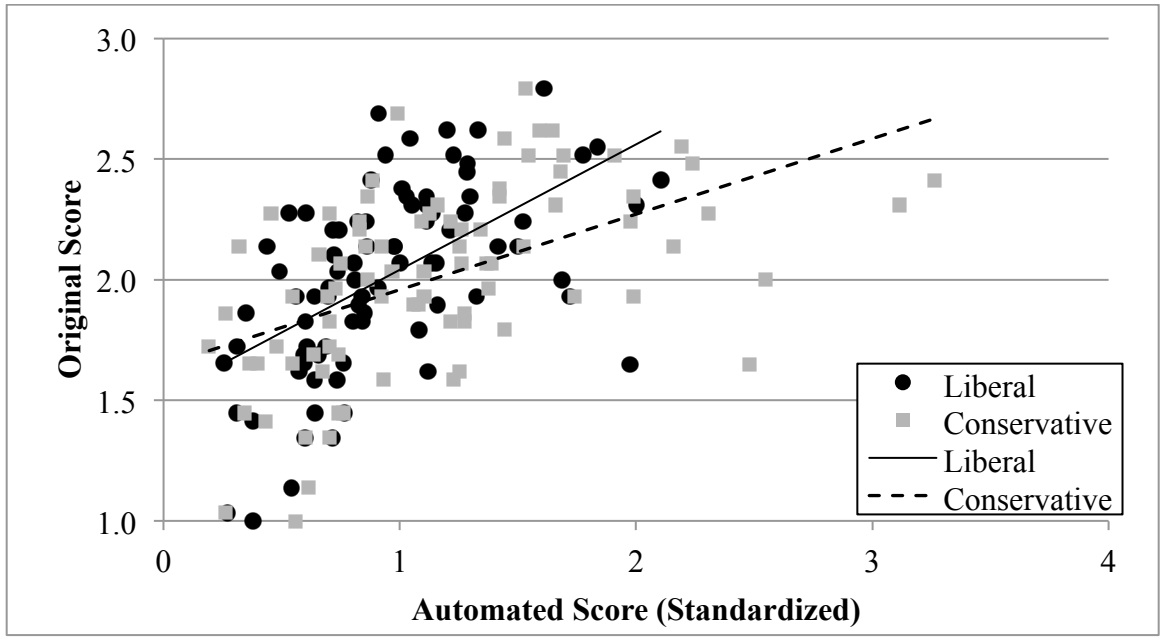


Figure 8. Comparison of scoring methods for the mid-level managers' exercise.

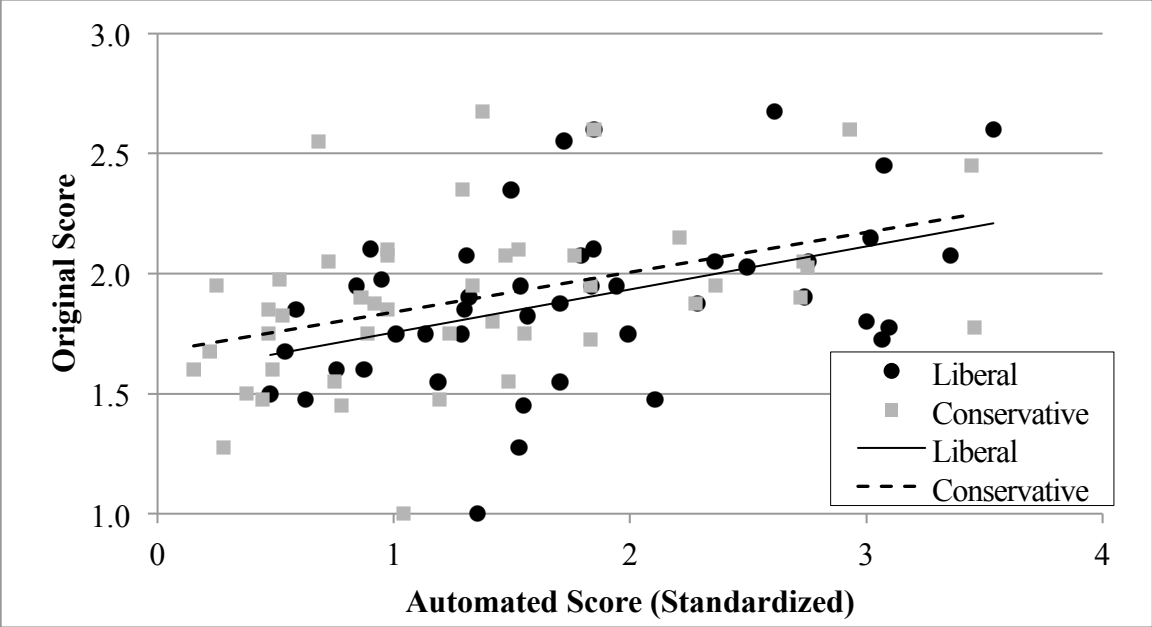


Figure 9. Comparison of scoring methods for the high-level managers' exercise

Table 1. Definitions for the AC competencies

<b>Competency</b>	<b>Definition</b>
Mission and Values Leadership	Leads commitment to district mission and values that motivate and engage the team and others.
Strategy and Results	Demonstrates broad thinking and decision-making focused on district goals and succeeds in accomplishing these goals.
Talent	Seeks opportunities for self, team, and overall district to learn and develop.
Relationships	Builds strong relationships with key internal and external customers and partners.
Communication	Influences others through effective communication about district priorities.
Change	Supports change initiatives within the district by seeking opportunities for continual improvement.

*Note.* The competency Change was not assessed in the high-level managers' exercise.

Table 2. Example coded response from the high-level managers' exercise (i.e., client concern)

Response Content	Code	Category	Subcategory
“Please know that we are conducting an investigation into the source of the leaked information.”	4d	Provided information or an update	Explained current actions being taken by self or others
“We highly value our partnership with your organization and would never intentionally put this at risk.”	7b	Showed positive recognition (for a company, project, program, product, or work)	Discussed value or importance
“As we conduct our investigation, I would ask that you not withdraw your presentation from the conference until more information can be obtained.”	3a	Made a request	Asked another to reconsider or wait on a decision
“...from the conference...”	9a	Spelling or grammar issue	Had an issue with the spelling or grammar used in the email
“Thank you,”	1d	Provided professional response	Provided salutation at the end of the email
“John Smith”	1e	Provided professional response	Provided name at the end of the email

Table 3. Inter-rater agreement on coding *schema* for each wave of data coding

Category	Wave 1		Wave 2	
	Mid-Level	High-Level	Mid-Level	High-Level
1. Provided professional response	87%	89%	84%	64%
2. Recognized feelings / emotions	85%	80%	80%	49%
3. Made a request	74%	68%	77%	48%
4. Provided information or an update	72%	68%	67%	46%
5. Established a timeframe / timeline	88%	65%	89%	61%
6. Evaluated the situation	94%	92%	98%	62%
7. Showed positive recognition (for a company, project, program, product, or work)	91%	82%	83%	81%
8. Showed positive recognition (for a person, relationship, or collaboration)	85%	83%	89%	76%
9. Spelling or grammar issue	57%	59%	78%	53%

*Note.* Wave 1 included 3 raters for both managers' exercises. Wave 2 included 4 raters for both managers' exercises.



Table 4. Frequency and effectiveness of behaviors in mid-level managers' exercise

Coding Schema	<i>F</i>	<i>E</i>
1. Provided professional response	--	--
a. Provided greeting	61	2.75
<i>b. Introduced self</i>	12	3.75
c. Expressed appreciation or thankfulness	85	3.50
<i>d. Provided salutation at the end of the email</i>	88	2.25
<i>e. Provided name at the end of email</i>	100	2.00
<i>f. Provided position and/or company information at the end of the email</i>	23	1.75
<i>g. Used information or unprofessional language / abbreviations</i>	15	-2.00
<i>h. Email title was inappropriate or potentially offensive to recipient</i>	9	-3.50
i. Forwarded content that was inappropriate or potentially offensive to recipient	2	-4.75
2. Recognized feelings / emotions	--	--
a. Expressed satisfaction, optimism, hope, positive anticipation, or personal dedication to something	10	4.00
b. Expressed dissatisfaction, regret, disappointment, or apologized	8	0.75
c. Expressed empathy, considered another's feelings or related own feelings to others	6	4.25
d. Agreed with or showed understanding for a claim, statement, concern, or opinion	38	3.75
3. Made a request	--	--
<i>a. Asked another to consider something a different way or reframe thinking</i>	5	4.00
<i>b. Asked another for a meeting / phone call / or discussion or implied it would be a good idea</i>	7	2.50
<i>c. Asked another to provide information, verify information, give advice / direction / instruction or made an inquiry</i>	15	3.75
<i>d. Asked another to respond</i>	7	3.00
<i>e. Asked another to take action / delegated actions or express that help or assistance is needed</i>	65	4.25
f. Explained the details of the actions they should take or of the assistance that is needed	99	4.25
<i>g. Asked for feedback, suggestions, concerns, or additional input or implied this was welcome</i>	14	4.00
4. Provided information or an update	--	--
a. Explained the current situation	29	3.50
<i>b. Provided or offered information or explanations to maintain communication</i>	44	3.00
<i>c. Provided or offered another general advice, direction, or information about something</i>	85	2.75
d. Provided or offered another detailed advice, direction, or information about something	51	3.00
e. Provided specific examples of something as part of an explanation	7	4.25
<i>f. Sent someone information or offered information merely as a point of interest or to reference</i>	17	2.25
g. Informed another that they would handle a situation or personally take action	53	2.25
h. Explained one's or another's current or desired role, responsibility, or expectation	12	4.25
<i>i. Expressed intent to follow-up or expectation/hope of future contact</i>	26	3.50

Table 4. Frequency of behaviors in mid-level managers' exercise (cont.)

Coding Schema	<i>F</i>	<i>E</i>
4. Provided information or an update	--	--
j. Suggested or brainstormed a strategic idea or potential plan or change that could be implemented in the future	13	4.00
5. Established a timeframe / timeline	--	--
a. Communicated personal time constraints or implied a sense of urgency	8	3.25
b. Communicated current schedule / timeline / roadmap	16	3.75
6. Evaluated the situation	--	--
<i>a. Implied or stated another had negative intentions, acted inappropriately, acted impulsively, did not have adequate information, or provided incomplete, false, or inappropriate information</i>	9	-0.50
b. Implied or stated false, misleading, or deceitful information	0	-5.00
c. Was generally critical, negative, or pessimistic regarding something	7	-3.50
7. Showed positive recognition (for a company, project, program, product, or work)	--	--
a. Showed praise	5	3.25
b. <i>Discussed value or importance</i>	11	3.75
c. <i>Discussed the investment, commitment, time, or effort that has already been put in</i>	2	2.75
d. Discussed the potential future success or trust, faith or confidence in it, dedication to it	11	4.25
e. Discussed the goal, purpose, mission or intention of it	66	4.75
8. Showed positive recognition (for a person, relationship, or collaboration)	--	--
a. <i>Showed praise</i>	22	4.25
b. Discussed value or importance	15	4.50
c. <i>Discussed the investment, commitment, time, or effort that has already been put in</i>	0	3.50
d. Discussed the potential future success or trust, faith or confidence in it, dedication to it	14	4.75
9. Spelling or grammar issue	--	--
a. <i>Had an issue with the spelling or grammar used in the email</i>	250	-2.75
b. <i>Addressed or referenced someone using the wrong, incomplete, or misspelled name/title</i>	8	-2.75

*Note.*  $N = 84$ .  $F$  = frequency that behavior occurred in the participant responses,  $E$  = average effectiveness ratings from SMEs. Italicized items were excluded in the conservative scoring method.

Table 5. Frequency and effectiveness of behaviors in high-level managers' exercise

Coding Schema	<i>F</i>	<i>E</i>
1. Provided professional response	--	--
a. Provided greeting	49	2.50
b. Introduced self	6	2.75
c. Expressed appreciation or thankfulness	18	3.25
<i>d. Provided salutation at the end of the email</i>	97	2.75
<i>e. Provided name at the end of the email</i>	127	2.75
<b><i>f. Provided position and/or company information at the end of the email</i></b>	43	2.50
<i>g. Used information or unprofessional language / abbreviations</i>	19	-3.00
2. Recognized feelings / emotions	--	--
<i>a. Expressed satisfaction, optimism, or personal dedication to something</i>	46	4.00
<i>b. Expressed dissatisfaction, regret, disappointment, or apologized</i>	29	4.25
c. Expressed empathy, considered another's feelings or related own feelings to others	18	4.50
d. Agreed with or showed understanding for a claim, statement, concern, or opinion	25	4.00
3. Made a request	--	--
<i>a. Asked another to reconsider or wait on a decision</i>	48	3.50
b. Asked another for a meeting / phone call / or discussion	20	4.00
<i>c. Asked another to provide information, verify information, give advice / direction / instruction or made an inquiry</i>	43	4.00
<i>d. Asked another to respond, expressed intent to follow-up, or expectation/hope of future contact</i>	48	3.00
e. Asked another to take action / delegated actions or express that help or assistance is needed	28	3.75
f. Explained the details of the assistance that is needed	13	4.50
4. Provided information or an update	--	--
<i>a. Explained the current situation</i>	48	4.50
<i>b. Provided or offered information or explanations to maintain communication</i>	26	4.00
<i>c. Explained completed actions</i>	43	4.25
d. Explained current actions being taken by self or others	52	4.25
e. Explained future actions – next steps	86	4.75
<i>f. Sent someone information merely as a point of interest or to reference</i>	15	1.25
5. Established a timeframe / timeline	--	--
<i>a. Communicated personal time constraints or implied a sense of urgency</i>	62	3.00
<i>b. Communicated current schedule / timeline / roadmap</i>	18	4.50
6. Evaluated the situation	--	--
<i>a. Implied or stated another had negative intentions or acted inappropriately or provided incomplete, false, or inappropriate information</i>	21	1.25
b. Implied or stated false, misleading, or deceitful information	3	-4.75
7. Showed positive recognition (for a company, project, program, product, or work)	--	--
a. Showed praise	6	4.00
b. Discussed value or importance	13	4.50
<i>c. Discussed the investment, commitment, time, or effort that has already been put in</i>	10	4.25
<i>d. Discussed the potential future success or the trust, faith or confidence in it, dedication to it</i>	15	4.50

Table 5. Frequency of behaviors in high-level managers' exercise (cont.)

<b>Coding Schema</b>	<i>F</i>	<i>E</i>
7. Showed positive recognition (for a company, project, program, product, or work)	--	--
e. Discussed the goal, purpose, mission or intention of it	9	4.25
8. Showed positive recognition (for a person, relationship, or collaboration)	--	--
a. Showed praise	14	4.75
b. Discussed value or importance	22	4.75
<i>c. Discussed the investment, commitment, time, or effort that has already been put in</i>	5	4.25
d. Discussed the potential future success or the trust, faith or confidence in it, dedication to it	13	4.50
9. Spelling or grammar issue	--	--
<i>a. Had an issue with the spelling or grammar used in the email</i>	140	-2.75
b. Addressed or reference someone using the wrong, incomplete, or misspelled name/title	34	-4.00

*Note.*  $N = 45$ .  $F$  = frequency that behavior occurred in the participant responses,  $E$  = average effectiveness ratings from SMEs. Italicized items were excluded in the conservative scoring method. Item 1f. was also eliminated from the liberal model since the behavior did not link to any of the listed competencies.

Table 6. Behaviors linked to *Mission and Values Leadership* for the mid-level managers' exercise

<b>Behaviors Linked to <i>Mission &amp; Values Leadership</i> Using Conservative Criteria</b>		<b>M</b>	<b>S</b>	<b>T</b>	<b>R</b>	<b>C</b>	<b>Ch</b>
3f	Explained the details of the actions they should take or of the assistance that is needed	C	L			C	L
4a	Explained the current situation	C					L
4d	Provided or offered another detailed advice, direction, or information about something	C	L	C	L	L	L
4g	Informed another that they would handle a situation or personally take action	C	L				
5b	Communicated current schedule / timeline / roadmap	C	L				L
6b	Implied or stated false, misleading, or deceitful information	C	C	L	L		C
6c	Was generally critical, negative, or pessimistic regarding something	C	L			L	C
7a	Showed praise	C		L	L	L	L
7e	Discussed the goal, purpose, mission or intention of it	C	L			L	C
8b	Discussed value or importance	C		L	L	L	
<b>Additional Behaviors Linked to <i>Mission &amp; Values Leadership</i> Using Liberal Criteria</b>		<b>M</b>	<b>S</b>	<b>T</b>	<b>R</b>	<b>C</b>	<b>Ch</b>
1a	Provided greeting	L		L	C	L	
1b	Introduced self	L		L	L	L	
1h	Email title was inappropriate or potentially offensive to recipient	L	L		L	L	
2a	Expressed satisfaction, optimism, hope, positive anticipation, or personal dedication to something	L	L	L	L	L	C
2b	Expressed dissatisfaction, regret, disappointment, or apologized	L	L	C	L	L	L
2c	Expressed empathy, considered another's feelings or related own feelings to others	L		C	C	L	
2d	Agreed with or showed understanding for a claim, statement, concern, or opinion	L	L		C	L	
3a	Asked another to consider something a different way or reframe thinking	L	L	L	L	L	L
3c	Asked another to provide information, verify information, give advice / direction / instruction or made an inquiry	L		L		L	
3e	Asked another to take action / delegated actions or express that help or assistance is needed	L	L	L		L	
3g	Asked for feedback, suggestions, concerns, or additional input or implied this was welcome	L					L
4c	Provided or offered another general advice, direction, or information about something	L	L	L	L	L	L
4e	Provided specific examples of something as part of an explanation	L	L			C	L
4f	Sent someone information or offered information merely as a point of interest or to reference	L				L	L
4h	Explained one's or another's current or desired role, responsibility, or expectation	L	L	L	L	L	C
4i	Expressed intent to follow-up or expectation/hope of future contact	L			L	L	L
4j	Suggested or brainstormed a strategic idea or potential plan or change that could be implemented in the future	L	C				L

Table 6. Behaviors linked to *Mission and Values Leadership* for the mid-level managers' exercise (cont.)

<b>Additional Behaviors Linked to <i>Mission &amp; Values Leadership</i> Using Liberal Criteria</b>		<b>M</b>	<b>S</b>	<b>T</b>	<b>R</b>	<b>C</b>	<b>Ch</b>
6a	Implied or stated another had negative intentions, acted inappropriately, acted impulsively, did not have adequate information, or provided incomplete, false, or inappropriate information	L	L		L	L	
7b	Discussed value or importance	L			L	L	
7c	Discussed the investment, commitment, time, or effort that has already been put in	L	L			L	
7d	Discussed the potential future success or trust, faith or confidence in it, dedication to it	L	L				C
8a	Showed praise	L		L	L	L	
8c	Discussed the investment, commitment, time, or effort that has already been put in	L	L	L		L	
8d	Discussed the potential future success or trust, faith or confidence in it, dedication to it	L	C	L	L	L	L

*Note.* C = behaviors linked to competency based on conservative scoring criteria only, L = behaviors linked to competency based on liberal scoring criteria, M = Mission and Values Leadership, S = Strategy and Results, T = Talent, R = Relationship, C = Communication, Ch = Change. Values outside the highlighted column indicate additional competencies the behavior was linked to by the SMEs.

Table 7. Behaviors linked to *Strategy and Results* for the mid-level managers' exercise

<b>Behaviors Linked to <i>Strategy and Results</i> Using Conservative Criteria</b>		<b>M</b>	<b>S</b>	<b>T</b>	<b>R</b>	<b>C</b>	<b>Ch</b>
4j	Suggested or brainstormed a strategic idea or potential plan or change that could be implemented in the future	L	C				L
5a	Communicated personal time constraints or implied a sense of urgency		C			L	L
6b	Implied or stated false, misleading, or deceitful information	C	C	L	L	C	
8d	Discussed the potential future success or trust, faith or confidence in it, dedication to it	L	C	L	L	L	L
<b>Additional Behaviors Linked to <i>Strategy and Results</i> Using Liberal Criteria</b>		<b>M</b>	<b>S</b>	<b>T</b>	<b>R</b>	<b>C</b>	<b>Ch</b>
1h	Email title was inappropriate or potentially offensive to recipient	L	L	L	L		
1i	Forwarded content that was inappropriate or potentially offensive to recipient		L	L	C	L	
2a	Expressed satisfaction, optimism, hope, positive anticipation, or personal dedication to something	L	L	L	L	L	C
2b	Expressed dissatisfaction, regret, disappointment, or apologized	L	L	C	L	L	L
2d	Agreed with or showed understanding for a claim, statement, concern, or opinion	L	L	C	L		
3a	Asked another to consider something a different way or reframe thinking	L	L	L	L	L	L
3b	Asked another for a meeting / phone call / or discussion or implied it would be a good idea		L	L			L
3d	Asked another to respond		L				L
3e	Asked another to take action / delegated actions or express that help or assistance is needed	L	L	L		L	
3f	Explained the details of the actions they should take or of the assistance that is needed	C	L			C	L
4b	Provided or offered information or explanations to maintain communication		L	L	L	L	
4c	Provided or offered another general advice, direction, or information about something	L	L	L	L	L	L
4d	Provided or offered another detailed advice, direction, or information about something	C	L	C	L	L	L
4e	Provided specific examples of something as part of an explanation	L	L			C	L
4g	Informed another that they would handle a situation or personally take action	C	L				
4h	Explained one's or another's current or desired role, responsibility, or expectation	L	L	L	L	L	C
5b	Communicated current schedule / timeline / roadmap	C	L				L
6a	Implied or stated another had negative intentions, acted inappropriately, acted impulsively, did not have adequate information, or provided incomplete, false, or inappropriate information	L	L	L	L		
6c	Was generally critical, negative, or pessimistic regarding something	C	L	L	C		
7c	Discussed the investment, commitment, time, or effort that has already been put in	L	L			L	
7d	Discussed the potential future success or trust, faith or confidence in it, dedication to it	L	L				C
7e	Discussed the goal, purpose, mission or intention of it	C	L			L	C

Table 7. Behaviors linked to *Strategy and Results* for the mid-level managers' exercise (cont.)

<b>Additional Behaviors Linked to <i>Strategy and Results</i> Using Liberal Criteria</b>		<b>M</b>	<b>S</b>	<b>T</b>	<b>R</b>	<b>C</b>	<b>Ch</b>
8c	Discussed the investment, commitment, time, or effort that has already been put in	L	L	L		L	

*Note.* C = behaviors linked to competency based on conservative scoring criteria only, L = behaviors linked to competency based on liberal scoring criteria, M = Mission and Values Leadership, S = Strategy and Results, T = Talent, R = Relationship, C = Communication, Ch = Change. Values outside the highlighted column indicate additional competencies the behavior was linked to by the SMEs.



Table 8. Behaviors linked to *Talent* for the mid-level managers' exercise

<b>Behaviors Linked to <i>Talent</i> Using Conservative Criteria</b>		<b>M</b>	<b>S</b>	<b>T</b>	<b>R</b>	<b>C</b>	<b>Ch</b>
2b	Expressed dissatisfaction, regret, disappointment, or apologized	L	L	C	L	L	L
2c	Expressed empathy, considered another's feelings or related own feelings to others	L		C	C	L	
4d	Provided or offered another detailed advice, direction, or information about something	C	L	C	L	L	L
<b>Additional Behaviors Linked to <i>Talent</i> Using Liberal Criteria</b>		<b>M</b>	<b>S</b>	<b>T</b>	<b>R</b>	<b>C</b>	<b>Ch</b>
1a	Provided greeting	L		L	C	L	
1b	Introduced self	L		L	L	L	
1c	Expressed appreciation or thankfulness			L	C	L	L
1d	Provided salutation at the end of the email			L	L	L	
1i	Forwarded content that was inappropriate or potentially offensive to recipient		L	L	C	L	
2a	Expressed satisfaction, optimism, hope, positive anticipation, or personal dedication to something	L	L	L	L	L	C
3a	Asked another to consider something a different way or reframe thinking	L	L	L	L	L	L
3b	Asked another for a meeting / phone call / or discussion or implied it would be a good idea		L	L			L
3c	Asked another to provide information, verify information, give advice / direction / instruction or made an inquiry	L		L		L	
3e	Asked another to take action / delegated actions or express that help or assistance is needed	L	L	L		L	
4b	Provided or offered information or explanations to maintain communication		L	L	L	L	
4c	Provided or offered another general advice, direction, or information about something	L	L	L	L	L	L
4h	Explained one's or another's current or desired role, responsibility, or expectation	L	L	L	L	L	C
6b	Implied or stated false, misleading, or deceitful information	C	C	L	L	C	
7a	Showed praise	C		L	L	L	L
8a	Showed praise	L		L	L	L	
8b	Discussed value or importance	C		L	L	L	
8c	Discussed the investment, commitment, time, or effort that has already been put in	L	L	L		L	
8d	Discussed the potential future success or trust, faith or confidence in it, dedication to it	L	C	L	L	L	L

*Note.* C = behaviors linked to competency based on conservative scoring criteria only, L = behaviors linked to competency based on liberal scoring criteria, M = Mission and Values Leadership, S = Strategy and Results, T = Talent, R = Relationship, C = Communication, Ch = Change. Values outside the highlighted column indicate additional competencies the behavior was linked to by the SMEs.

Table 9. Behaviors linked to *Relationships* for the mid-level managers' exercise

<b>Behaviors Linked to <i>Relationships</i> Using Conservative Criteria</b>		<b>M</b>	<b>S</b>	<b>T</b>	<b>R</b>	<b>C</b>	<b>Ch</b>
1a	Provided greeting	L		L	C	L	
1c	Expressed appreciation or thankfulness			L	C	L	L
1i	Forwarded content that was inappropriate or potentially offensive to recipient			L	L	C	L
2c	Expressed empathy, considered another's feelings or related own feelings to others	L		C	C	L	
2d	Agreed with or showed understanding for a claim, statement, concern, or opinion	L	L		C	L	
<b>Additional Behaviors Linked to <i>Relationships</i> Using Liberal Criteria</b>		<b>M</b>	<b>S</b>	<b>T</b>	<b>R</b>	<b>C</b>	<b>Ch</b>
1b	Introduced self	L		L	L	L	
1d	Provided salutation at the end of the email			L	L	L	
1e	Provided name at the end of email				L		
1f	Provided position and/or company information at the end of the email				L		
1g	Used information or unprofessional language / abbreviations				L	L	
1h	Email title was inappropriate or potentially offensive to recipient	L	L		L	L	
2a	Expressed satisfaction, optimism, hope, positive anticipation, or personal dedication to something	L	L	L	L	L	C
2b	Expressed dissatisfaction, regret, disappointment, or apologized	L	L	C	L	L	L
3a	Asked another to consider something a different way or reframe thinking	L	L	L	L	L	L
4b	Provided or offered information or explanations to maintain communication			L	L	L	L
4c	Provided or offered another general advice, direction, or information about something	L	L	L	L	L	L
4d	Provided or offered another detailed advice, direction, or information about something	C	L	C	L	L	L
4h	Explained one's or another's current or desired role, responsibility, or expectation	L	L	L	L	L	C
4i	Expressed intent to follow-up or expectation/hope of future contact	L			L	L	L
6a	Implied or stated another had negative intentions, acted inappropriately, acted impulsively, did not have adequate information, or provided incomplete, false, or inappropriate information	L	L		L	L	
6b	Implied or stated false, misleading, or deceitful information	C	C	L	L	C	
6c	Was generally critical, negative, or pessimistic regarding something	C	L		L	C	
7a	Showed praise	C		L	L	L	L
7b	Discussed value or importance	L			L	L	
8a	Showed praise	L		L	L	L	
8b	Discussed value or importance	C		L	L	L	

Table 9. Behaviors linked to *Relationships* for the mid-level managers' exercise (cont.)

<b>Additional Behaviors Linked to <i>Relationships</i> Using Liberal Criteria</b>		<b>M</b>	<b>S</b>	<b>T</b>	<b>R</b>	<b>C</b>	<b>Ch</b>
8d	Discussed the potential future success or trust, faith or confidence in it, dedication to it	L	C	L	L	L	L
9b	Addressed or referenced someone using the wrong, incomplete, or misspelled name/title				L	L	

*Note.* C = behaviors linked to competency based on conservative scoring criteria only, L = behaviors linked to competency based on liberal scoring criteria, M = Mission and Values Leadership, S = Strategy and Results, T = Talent, R = Relationship, C = Communication, Ch = Change. Values outside the highlighted column indicate additional competencies the behavior was linked to by the SMEs.

Table 10. Behaviors linked to *Communication* for the mid-level managers' exercise

<b>Behaviors Linked to <i>Communication</i> Using Conservative Criteria</b>		<b>M</b>	<b>S</b>	<b>T</b>	<b>R</b>	<b>C</b>	<b>Ch</b>
3f	Explained the details of the actions they should take or of the assistance that is needed	C	L			C	L
4e	Provided specific examples of something as part of an explanation	L	L			C	L
6b	Implied or stated false, misleading, or deceitful information	C	C	L	L	C	
6c	Was generally critical, negative, or pessimistic regarding something	C	L		L	C	
<b>Additional Behaviors Linked to <i>Communication</i> Using Liberal Criteria</b>		<b>M</b>	<b>S</b>	<b>T</b>	<b>R</b>	<b>C</b>	<b>Ch</b>
1a	Provided greeting	L		L	C	L	
1b	Introduced self	L		L	L	L	
1c	Expressed appreciation or thankfulness			L	C	L	L
1d	Provided salutation at the end of the email			L	L	L	
1g	Used information or unprofessional language / abbreviations				L	L	
1h	Email title was inappropriate or potentially offensive to recipient	L	L		L	L	
1i	Forwarded content that was inappropriate or potentially offensive to recipient		L	L	C	L	
2a	Expressed satisfaction, optimism, hope, positive anticipation, or personal dedication to something	L	L	L	L	L	C
2b	Expressed dissatisfaction, regret, disappointment, or apologized	L	L	C	L	L	L
2c	Expressed empathy, considered another's feelings or related own feelings to others	L		C	C	L	
2d	Agreed with or showed understanding for a claim, statement, concern, or opinion	L	L		C	L	
3a	Asked another to consider something a different way or reframe thinking	L	L	L	L	L	L
3c	Asked another to provide information, verify information, give advice / direction / instruction or made an inquiry	L		L		L	
3d	Asked another to respond			L		L	
3e	Asked another to take action / delegated actions or express that help or assistance is needed	L	L	L		L	
4a	Explained the current situation	C				L	
4b	Provided or offered information or explanations to maintain communication		L	L	L	L	
4c	Provided or offered another general advice, direction, or information about something	L	L	L	L	L	L
4d	Provided or offered another detailed advice, direction, or information about something	C	L	C	L	L	L
4f	Sent someone information or offered information merely as a point of interest or to reference	L				L	L
4h	Explained one's or another's current or desired role, responsibility, or expectation	L	L	L	L	L	C
4i	Expressed intent to follow-up or expectation/hope of future contact	L			L	L	L
5a	Communicated personal time constraints or implied a sense of urgency		C			L	L

Table 10. Behaviors linked to *Communication* for the mid-level managers' exercise (cont.)

<b>Additional Behaviors Linked to <i>Communication</i> Using Liberal Criteria</b>		<b>M</b>	<b>S</b>	<b>T</b>	<b>R</b>	<b>C</b>	<b>Ch</b>
5b	Communicated current schedule / timeline / roadmap	C	L			L	
6a	Implied or stated another had negative intentions, acted inappropriately, acted impulsively, did not have adequate information, or provided incomplete, false, or inappropriate information	L	L	L	L		
7a	Showed praise	C	L	L	L	L	L
7b	Discussed value or importance	L		L	L		
7c	Discussed the investment, commitment, time, or effort that has already been put in	L	L			L	
7e	Discussed the goal, purpose, mission or intention of it	C	L			L	C
8a	Showed praise	L	L	L	L		
8b	Discussed value or importance	C	L	L	L		
8c	Discussed the investment, commitment, time, or effort that has already been put in	L	L	L		L	
8d	Discussed the potential future success or trust, faith or confidence in it, dedication to it	L	C	L	L	L	L
9a	Had an issue with the spelling or grammar used in the email						L
9b	Addressed or referenced someone using the wrong, incomplete, or misspelled name/title				L	L	

*Note.* C = behaviors linked to competency based on conservative scoring criteria only, L = behaviors linked to competency based on liberal scoring criteria, M = Mission and Values Leadership, S = Strategy and Results, T = Talent, R = Relationship, C = Communication, Ch = Change. Values outside the highlighted column indicate additional competencies the behavior was linked to by the SMEs.

Table 11. Behaviors linked to *Change* for the mid-level managers' exercise

<b>Behaviors Linked to <i>Change</i> Using Conservative Criteria</b>		<b>M</b>	<b>S</b>	<b>T</b>	<b>R</b>	<b>C</b>	<b>Ch</b>
2a	Expressed satisfaction, optimism, hope, positive anticipation, or personal dedication to something	L	L	L	L	L	C
4h	Explained one's or another's current or desired role, responsibility, or expectation	L	L	L	L	L	C
7d	Discussed the potential future success or trust, faith or confidence in it, dedication to it	L	L				C
7e	Discussed the goal, purpose, mission or intention of it	C	L			L	C
<b>Additional Behaviors Linked to <i>Change</i> Using Liberal Criteria</b>		<b>M</b>	<b>S</b>	<b>T</b>	<b>R</b>	<b>C</b>	<b>Ch</b>
1c	Expressed appreciation or thankfulness			L	C	L	L
2b	Expressed dissatisfaction, regret, disappointment, or apologized	L	L	C	L	L	L
3a	Asked another to consider something a different way or reframe thinking	L	L	L	L	L	L
3b	Asked another for a meeting / phone call / or discussion or implied it would be a good idea		L	L			L
3f	Explained the details of the actions they should take or of the assistance that is needed	C	L			C	L
3g	Asked for feedback, suggestions, concerns, or additional input or implied this was welcome	L					L
4c	Provided or offered another general advice, direction, or information about something	L	L	L	L	L	L
4d	Provided or offered another detailed advice, direction, or information about something	C	L	C	L	L	L
4e	Provided specific examples of something as part of an explanation	L	L			C	L
4f	Sent someone information or offered information merely as a point of interest or to reference	L				L	L
4i	Expressed intent to follow-up or expectation/hope of future contact	L		L	L	L	L
4j	Suggested or brainstormed a strategic idea or potential plan or change that could be implemented in the future	L	C				L
5a	Communicated personal time constraints or implied a sense of urgency		C			L	L
7a	Showed praise	C	L	L	L	L	L
8d	Discussed the potential future success or trust, faith or confidence in it, dedication to it	L	C	L	L	L	L

*Note.* C = behaviors linked to competency based on conservative scoring criteria only, L = behaviors linked to competency based on liberal scoring criteria, M = Mission and Values Leadership, S = Strategy and Results, T = Talent, R = Relationship, C = Communication, Ch = Change. Values outside the highlighted column indicate additional competencies the behavior was linked to by the SMEs.

Table 12. Behaviors linked to *Mission and Values Leadership* for the high-level managers' exercise

<b>Behaviors Linked to <i>Mission and Values Leadership</i> Using Conservative Criteria</b>		<b>M</b>	<b>S</b>	<b>T</b>	<b>R</b>	<b>C</b>
7b	Discussed value or importance	C	L		L	C
<b>Additional Behaviors Linked to <i>Mission and Values Leadership</i> Using Liberal Criteria</b>		<b>M</b>	<b>S</b>	<b>T</b>	<b>R</b>	<b>C</b>
2a	Expressed satisfaction, optimism, or personal dedication to something	L			L	L
3b	Asked another for a meeting / phone call / or discussion	L	L		C	L
3d	Asked another to respond, expressed intent to follow-up, or expectation/hope of future contact	L	L	L	L	L
3e	Asked another to take action / delegated actions or express that help or assistance is needed	L	C	L		L
4e	Explained future actions – next steps	L	L	L		C
5b	Communicated current schedule / timeline / roadmap	L	L			L
7a	Showed praise	L	L	L	C	L
7c	Discussed the investment, commitment, time, or effort that has already been put in	L	L	L		L
7d	Discussed the potential future success or the trust, faith or confidence in it, dedication to it	L	L	L		L
7e	Discussed the goal, purpose, mission or intention of it	L	L		L	C
8a	Showed praise	L	L	C	C	C
8b	Discussed value or importance	L	L	C	L	C
8c	Discussed the investment, commitment, time, or effort that has already been put in	L	L		L	
8d	Discussed the potential future success or the trust, faith or confidence in it, dedication to it	L	L	C	L	L

*Note.* C = behaviors linked to competency based on conservative scoring criteria only, L = behaviors linked to competency based on liberal scoring criteria, M = Mission and Values Leadership, S = Strategy and Results, T = Talent, R = Relationship, C = Communication, Ch = Change. Values outside the highlighted column indicate additional competencies the behavior was linked to by the SMEs.

Table 13. Behaviors linked to *Strategy and Results* for the high-level managers' exercise

<b>Behaviors Linked to <i>Strategy and Results</i> Using Conservative Criteria</b>		<b>M</b>	<b>S</b>	<b>T</b>	<b>R</b>	<b>C</b>
3e	Asked another to take action / delegated actions or express that help or assistance is needed	L	C	L		L
6b	Implied or stated false, misleading, or deceitful information		C		C	L
<b>Additional Behaviors Linked to <i>Strategy and Results</i> Using Liberal Criteria</b>		<b>M</b>	<b>S</b>	<b>T</b>	<b>R</b>	<b>C</b>
2b	Expressed dissatisfaction, regret, disappointment, or apologized	L				L
2c	Expressed empathy, considered another's feelings or related own feelings to others	L			C	C
2d	Agreed with or showed understanding for a claim, statement, concern, or opinion	L			L	C
3b	Asked another for a meeting / phone call / or discussion	L	L		C	L
3c	Asked another to provide information, verify information, give advice / direction / instruction or made an inquiry	L	L	L	L	L
3d	Asked another to respond, expressed intent to follow-up, or expectation/hope of future contact	L	L	L	L	L
3f	Explained the details of the assistance that is needed	L	L	L		C
4a	Explained the current situation	L				L
4b	Provided or offered information or explanations to maintain communication	L	L	L	L	L
4c	Explained completed actions	L			L	L
4d	Explained current actions being taken by self or others	L				C
4e	Explained future actions – next steps	L	L	L		C
4f	Sent someone information merely as a point of interest or to reference	L			L	L
5a	Communicated personal time constraints or implied a sense of urgency	L				L
5b	Communicated current schedule / timeline / roadmap	L	L			L
6a	Implied or stated another had negative intentions or acted inappropriately or provided incomplete, false, or inappropriate information	L	L	L	L	L
7a	Showed praise	L	L	L	C	L
7b	Discussed value or importance	C	L		L	C
7c	Discussed the investment, commitment, time, or effort that has already been put in	L	L	L		L
7d	Discussed the potential future success or the trust, faith or confidence in it, dedication to it	L	L	L		L
7e	Discussed the goal, purpose, mission or intention of it	L	L		L	C
8a	Showed praise	L	L	C	C	C
8b	Discussed value or importance	L	L	C	L	C
8c	Discussed the investment, commitment, time, or effort that has already been put in	L	L			L



Table 13. Behaviors linked to *Strategy and Results* for the high-level managers' exercise (cont.)

<b>Additional Behaviors Linked to <i>Strategy and Results</i> Using Liberal Criteria</b>		<b>M</b>	<b>S</b>	<b>T</b>	<b>R</b>	<b>C</b>
8d	Discussed the potential future success or the trust, faith or confidence in it, dedication to it	L	L	C	L	L
9b	Addressed or reference someone using the wrong, incomplete, or misspelled name/title		L		C	L

*Note.* C = behaviors linked to competency based on conservative scoring criteria only, L = behaviors linked to competency based on liberal scoring criteria, M = Mission and Values Leadership, S = Strategy and Results, T = Talent, R = Relationship, C = Communication, Ch = Change. Values outside the highlighted column indicate additional competencies the behavior was linked to by the SMEs.

Table 14. Behaviors linked to *Talent* for the high-level managers' exercise

<b>Behaviors Linked to <i>Talent</i> Using Conservative Criteria</b>		<b>M</b>	<b>S</b>	<b>T</b>	<b>R</b>	<b>C</b>
8a	Showed praise	L	L	C	C	C
8b	Discussed value or importance	L	L	C	L	C
8d	Discussed the potential future success or the trust, faith or confidence in it, dedication to it	L	L	C	L	L
<b>Additional Behaviors Linked to <i>Talent</i> Using Liberal Criteria</b>		<b>M</b>	<b>S</b>	<b>T</b>	<b>R</b>	<b>C</b>
3c	Asked another to provide information, verify information, give advice / direction / instruction or made an inquiry	L	L	L	L	L
3d	Asked another to respond, expressed intent to follow-up, or expectation/hope of future contact	L	L	L	L	L
3e	Asked another to take action / delegated actions or express that help or assistance is needed	L	C	L		L
3f	Explained the details of the assistance that is needed		L	L	L	C
4b	Provided or offered information or explanations to maintain communication		L	L	L	L
4e	Explained future actions – next steps	L	L	L		C
6a	Implied or stated another had negative intentions or acted inappropriately or provided incomplete, false, or inappropriate information		L	L	L	L
7a	Showed praise	L	L	L	C	L
7c	Discussed the investment, commitment, time, or effort that has already been put in	L	L	L		L
7d	Discussed the potential future success or the trust, faith or confidence in it, dedication to it	L	L	L		L

*Note.* C = behaviors linked to competency based on conservative scoring criteria only, L = behaviors linked to competency based on liberal scoring criteria, M = Mission and Values Leadership, S = Strategy and Results, T = Talent, R = Relationship, C = Communication, Ch = Change. Values outside the highlighted column indicate additional competencies the behavior was linked to by the SMEs.

Table 15. Behaviors linked to *Relationships* for the high-level managers' exercise

<b>Behaviors Linked to <i>Relationships</i> Using Conservative Criteria</b>		<b>M</b>	<b>S</b>	<b>T</b>	<b>R</b>	<b>C</b>
1a	Provided greeting					C L
1c	Expressed appreciation or thankfulness					C L
2c	Expressed empathy, considered another's feelings or related own feelings to others		L			C C
3b	Asked another for a meeting / phone call / or discussion	L	L			C L
6b	Implied or stated false, misleading, or deceitful information		C			C L
7a	Showed praise	L	L	L		C L
8a	Showed praise	L	L	C		C C
9b	Addressed or reference someone using the wrong, incomplete, or misspelled name/title		L			C L
<b>Additional Behaviors Linked to <i>Relationships</i> Using Liberal Criteria</b>		<b>M</b>	<b>S</b>	<b>T</b>	<b>R</b>	<b>C</b>
1b	Introduced self					L C
1d	Provided salutation at the end of the email					L L
1e	Provided name at the end of the email					L L
2a	Expressed satisfaction, optimism, or personal dedication to something	L				L L
2d	Agreed with or showed understanding for a claim, statement, concern, or opinion		L			L C
3a	Asked another to reconsider or wait on a decision					L L
3c	Asked another to provide information, verify information, give advice / direction / instruction or made an inquiry		L	L		L L
3d	Asked another to respond, expressed intent to follow-up, or expectation/hope of future contact	L	L	L		L L
3f	Explained the details of the assistance that is needed		L	L		L C
4b	Provided or offered information or explanations to maintain communication		L	L		L L
4c	Explained completed actions		L			L L
4f	Sent someone information merely as a point of interest or to reference		L			L L
6a	Implied or stated another had negative intentions or acted inappropriately or provided incomplete, false, or inappropriate information		L	L		L L
7b	Discussed value or importance	C	L			L C
7e	Discussed the goal, purpose, mission or intention of it	L	L			L C
8b	Discussed value or importance	L	L	C		L C
8c	Discussed the investment, commitment, time, or effort that has already been put in	L	L			L

Table 15. Behaviors linked to *Relationships* for the high-level managers' exercise (cont.)

<b>Additional Behaviors Linked to <i>Relationships</i> Using Liberal Criteria</b>		<b>M</b>	<b>S</b>	<b>T</b>	<b>R</b>	<b>C</b>
8d	Discussed the potential future success or the trust, faith or confidence in it, dedication to it	L	L	C	L	L

*Note.* C = behaviors linked to competency based on conservative scoring criteria only, L = behaviors linked to competency based on liberal scoring criteria, M = Mission and Values Leadership, S = Strategy and Results, T = Talent, R = Relationship, C = Communication, Ch = Change. Values outside the highlighted column indicate additional competencies the behavior was linked to by the SMEs.

Table 16. Behaviors linked to *Communication* for the high-level managers' exercise

<b>Behaviors Linked to <i>Communication</i> Using Conservative Criteria</b>		<b>M</b>	<b>S</b>	<b>T</b>	<b>R</b>	<b>C</b>
1b	Introduced self					L C
2c	Expressed empathy, considered another's feelings or related own feelings to others		L			C C
2d	Agreed with or showed understanding for a claim, statement, concern, or opinion		L		L	C
3f	Explained the details of the assistance that is needed		L	L	L	C
4d	Explained current actions being taken by self or others		L			C
4e	Explained future actions – next steps		L	L	L	C
7b	Discussed value or importance		C	L		L C
7e	Discussed the goal, purpose, mission or intention of it		L	L		L C
8a	Showed praise		L	L	C	C C
8b	Discussed value or importance		L	L	C	L C
<b>Additional Behaviors Linked to <i>Communication</i> Using Liberal Criteria</b>		<b>M</b>	<b>S</b>	<b>T</b>	<b>R</b>	<b>C</b>
1a	Provided greeting					C L
1c	Expressed appreciation or thankfulness					C L
1d	Provided salutation at the end of the email					L L
1e	Provided name at the end of the email					L L
1g	Used information or unprofessional language / abbreviations					L
2a	Expressed satisfaction, optimism, or personal dedication to something		L			L L
2b	Expressed dissatisfaction, regret, disappointment, or apologized		L			L
3a	Asked another to reconsider or wait on a decision					L L
3b	Asked another for a meeting / phone call / or discussion		L	L		C L
3c	Asked another to provide information, verify information, give advice / direction / instruction or made an inquiry		L	L	L	L
3d	Asked another to respond, expressed intent to follow-up, or expectation/hope of future contact		L	L	L	L
3e	Asked another to take action / delegated actions or express that help or assistance is needed		L	C	L	L
4a	Explained the current situation		L			L
4b	Provided or offered information or explanations to maintain communication		L	L	L	L
4c	Explained completed actions		L			L
4f	Sent someone information merely as a point of interest or to reference		L			L
5a	Communicated personal time constraints or implied a sense of urgency		L			L

Table 16. Behaviors linked to *Communication* for the high-level managers' exercise (cont.)

<b>Additional Behaviors Linked to <i>Communication</i> Using Liberal Criteria</b>		<b>M</b>	<b>S</b>	<b>T</b>	<b>R</b>	<b>C</b>
5b	Communicated current schedule / timeline / roadmap	L	L			L
6a	Implied or stated another had negative intentions or acted inappropriately or provided incomplete, false, or inappropriate information		L	L	L	L
6b	Implied or stated false, misleading, or deceitful information		C		C	L
7a	Showed praise	L	L	L	C	L
7c	Discussed the investment, commitment, time, or effort that has already been put in	L	L	L		L
7d	Discussed the potential future success or the trust, faith or confidence in it, dedication to it	L	L	L		L
8d	Discussed the potential future success or the trust, faith or confidence in it, dedication to it	L	L	C	L	L
9a	Had an issue with the spelling or grammar used in the email					L
9b	Addressed or reference someone using the wrong, incomplete, or misspelled name/title		L		C	L

Table 17. Cross-tabulation of participants for each AC exercise by gender

	<b>Males</b>	<b>Females</b>	<i>N</i>
High-Level Managers' Exercise	17	28	45
Mid-Level Managers' Exercise	29	55	84

Table 18. Correlations between gender, word count, original scores and automated scores for the mid-level managers' exercise

	1.	2.	Original Score							Conservative Score						Liberal Score							
	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.	15.	16.	17.	18.	19.	20.	21.	22.	23.
1. G	-	-																					
2. WC	.06	--																					
3. M	-.18	.04	--																				
4. S	-.05	.26*	.27*	--																			
5. T	-.03	.28*	.02	.42*	--																		
6. R	.10	.35*	.31*	.50*	.56*	--																	
7. C	.12	.31*	.25*	.58*	.54*	.74*	--																
8. Ch	-.02	.12	.31*	.42*	.32*	.61*	.51*	--															
9. OS	.05	.31*	.35*	.71*	.61*	.86*	.95*	.64*	--														
10. M	.08	.49*	<b>.01</b>	.30*	.23*	.41*	.38*	.20	.39*	--													
11. S	-.18	.35*	-.01	<b>.22*</b>	.22*	.19	.13	.03	.14	.11	--												
12. T	.05	.48	-.10	.11	<b>.17</b>	.26*	.28*	.11	.24*	.46*	.07	--											
13. R	-.13	.22*	.27*	.12	.26*	<b>.44*</b>	.26*	.36*	.34*	.14	-.03	.18	--										
14. C	.05	-.14	-.07	.22*	<-.01	.01	<b>.13</b>	-.05	.11	.47*	<.01	-.09	-.30*	--									
15. Ch	.02	.52*	.13	.10	.24*	.40*	.31*	<b>.25*</b>	.32*	.44*	.14	.38*	.35*	-.39*	--								
16. OS	-.04	.58*	.15	.36*	.36*	.57*	.49*	.35*	<b>.52*</b>	.84*	.32*	.47*	.54*	0.21	.62*	--							
17. M	.02	.59*	<b>.14</b>	.39*	.37*	.62*	.52*	.37*	.55*	<b>.78*</b>	.37*	.46*	.42*	.19	.58*	.91*	--						
18. S	-.11	.43*	.12	<b>.44*</b>	.36*	.59*	.49*	.33*	.54*	.77*	<b>.38*</b>	.50*	.26*	.37*	.41*	.82*	.88*	--					
19. T	-.11	.37*	.15	.28*	<b>.31*</b>	.51*	.44*	.34*	.44*	.38*	.30*	<b>.50*</b>	.53*	-.19	.50*	.65*	.74*	.61*	--				
20. R	-.06	.34*	.20	.16	.28*	<b>.54*</b>	.44*	.41*	.45*	.24*	.19	.45*	<b>.68*</b>	-.37*	.51*	.58*	.64*	.48*	.89*	--			
21. C	-.03	.40*	.17	.35*	.34*	.61*	<b>.55*</b>	.37*	.56*	.72*	.26*	.42*	.49*	<b>.21</b>	.47*	.86*	.93*	.83*	.83*	.75*	--		
22. Ch	-.01	.48*	.08	.29*	.33*	.56*	.46*	<b>.39*</b>	.48*	.75*	.32*	.52*	.40*	.20	<b>.60*</b>	.87*	.88*	.82*	.72*	.65*	.87*	--	
23. OS	-.02	.51*	.20	.34*	.37*	.62*	.54*	.41*	<b>.56*</b>	.71*	.30*	.43*	.55*	.08	.59*	<b>.90*</b>	.95*	.81*	.82*	.78*	.97*	.87*	--

Note. (N = 84) Gender: female = 0, male = 1. G = Gender, WC = Word Count, M = Mission and Values Leadership, S = Strategy and Results, T = Talent, R = Relationships, C = Communication, Ch = Change, OS = Overall Score, \*p<.05. Values for convergent validity are bolded along the diagonal.



Table 19. Correlations between gender, word count, original scores and automated scores for the high-level managers' exercise

	1.	2.	Original Score					Conservative Score					Liberal Score							
	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.	15.	16.	17.	18.	19.	20.
1. G	--																			
2. WC	-.02	--																		
Original	3. M	-.21	.27	--																
	4. S	-.09	.08	.17	--															
	5. T	-.09	.24	.36*	.47*	--														
	6. R	-.21	.18	.28	.45*	.54*	--													
	7. C	-.20	.08	.26	.33*	.41*	.67*	--												
	8. OS	-.06	.16	.43*	.62*	.52*	.67*	.59*	--											
Conservative	9. M	-.05	-.02	<b>.25</b>	.06	.10	.06	<.01	-.10	--										
	10. S	.02	.29	.03	<b>.10</b>	.12	-.01	.01	.09	-.02	--									
	11. T	-.39*	.28	.21	.05	<b>.30*</b>	.29	.37*	.34*	-.15	.11	--								
	12. R	-.04	.17	.33*	-.77	.27	<b>.34*</b>	.31*	.15	.18	-.10	.42*	--							
	13. C	-.41*	.64*	.30*	.20	.27	.31*	<b>.32*</b>	.25	.12	.28	.53*	.34*	--						
	14. OS	-.32*	.65*	.42*	.15	.39*	.38*	.31*	<b>.28</b>	.22	.32*	.53*	.48*	.86*	--					
Liberal	15. M	-.41*	.58*	<b>.16</b>	.16	.29	.14	.15	.12	<b>.21</b>	.32*	.56*	.24	.76*	.81*	--				
	16. S	-.28	.70*	.27	<b>.09</b>	.44*	.30*	.30*	.20	.05	<b>.33*</b>	.58*	.39*	.80*	.84*	.79*	--			
	17. T	-.44*	.62*	.14	.08	<b>.26</b>	.15	.18	.13	-.08	.30*	<b>.72*</b>	.32*	.71*	.73*	.84*	.81*	--		
	18. R	-.18	.64*	.15	.19	.38*	<b>.22</b>	.25	.22	-.03	.11	.54*	<b>.36*</b>	.46*	.60*	.55*	.70*	.71*	--	
	19. C	-.32*	.59*	.31*	.23	.47*	.32*	<b>.36*</b>	.28	-.09	.32*	.62*	.39*	<b>.70*</b>	.74*	.71*	.88*	.82*	.75*	--
	20. OS	-.27	.58*	.31*	.23	.47*	.28	.30*	<b>.28</b>	-.09	.36*	.62*	.39*	.64*	<b>.74*</b>	.71*	.85*	.80*	.78*	.97*

Note. (N = 45) Gender: female = 0, male = 1. G = Gender, WC = Word Count, M = Mission and Values Leadership, S = Strategy and Results, T = Talent, R = Relationships, C = Communication, OS = Overall Score, \*p<.05. Values for convergent validity are bolded along the diagonal. The competency Change was not evaluated in the high-level managers' exercise.

Table 20. Descriptive statistics for the mid-level managers' exercise for word count and competency scores

	Original Scores				Conservative Scores				Liberal Scores			
	<i>M</i>	<i>SD</i>	<i>Min</i>	<i>Max</i>	<i>M</i>	<i>SD</i>	<i>Min</i>	<i>Max</i>	<i>M</i>	<i>SD</i>	<i>Min</i>	<i>Max</i>
Word Count	90.80	128.68	3.00	570	--	--	--	--	--	--	--	--
Mission and Values Leadership	1.43	0.61	1.00	3.00	1.46	1.04	0.00	5.25	0.97	0.49	0.07	2.46
Strategy and Results	1.72	0.50	1.00	3.00	0.43	0.70	0.00	2.81	0.95	0.49	0.08	2.57
Talent	2.68	0.52	1.00	3.00	1.06	1.89	0.00	7.50	1.06	0.53	0.16	3.01
Relationships	2.15	0.55	1.00	3.00	1.49	1.08	-0.10	4.30	0.85	0.54	-0.02	2.63
Communication	2.06	0.39	1.00	3.00	1.27	2.12	-0.88	11.69	0.95	0.42	0.21	2.02
Change	1.88	0.78	1.00	3.00	1.34	1.51	0.00	5.81	1.24	0.70	0.16	3.38
Overall Score	2.01	0.38	1.00	2.79	1.17	0.63	0.19	3.26	0.94	0.42	0.26	2.10

*Note.*  $N = 84$ . Automated scores were the sum of the behaviors linked to that competency weighted by the SME effectiveness rating for each behavior. Higher scores indicate that more effective behaviors occurred or that the behaviors that occurred received higher effectiveness weights. Substantially higher scores appear when some behaviors occurred multiple times in a single response.

Table 21. Descriptive statistics for the high-level managers' exercise for word count and competency scores

	Original Scores				Conservative Scores				Liberal Scores			
	<i>M</i>	<i>SD</i>	<i>Min</i>	<i>Max</i>	<i>M</i>	<i>SD</i>	<i>Min</i>	<i>Max</i>	<i>M</i>	<i>SD</i>	<i>Min</i>	<i>Max</i>
Word Count	234.84	116.40	48	494	--	--	--	--	--	--	--	--
Mission and Values Leadership	1.62	0.78	1.00	3.00	1.00	2.02	0.00	5.00	1.73	1.19	0.00	4.00
Strategy and Results	1.71	0.55	1.00	3.00	0.73	1.54	-2.00	6.00	1.69	0.95	0.00	4.00
Talent	1.89	0.57	0.00	3.00	1.38	1.71	0.00	6.00	1.84	1.09	0.00	4.00
Relationships	2.04	0.60	1.00	3.00	0.58	0.87	-1.00	2.00	1.60	0.84	0.00	4.00
Communication	2.11	0.44	1.00	3.00	1.89	1.37	0.00	5.00	1.82	0.86	0.00	4.00
Overall Score	1.98	0.45	1.00	3.00	1.27	0.94	0.00	3.00	1.82	0.86	0.00	4.00

*Note.* *N* = 45. Automated scores were the sum of the behaviors linked to that competency weighted by the SME effectiveness rating for each behavior. Higher scores indicate that more effective behaviors occurred or that the behaviors that occurred received higher effectiveness weights. Substantially higher scores appear when some behaviors occurred multiple times in a single response.

Table 22. Linear regression of automated scores predicting original scores for the mid-level managers' exercise.

	<b>Conservative Scores</b>							<b>Liberal Scores</b>						
	<i>B</i>	<i>SE</i>	$\beta$	$R^2$	<i>95% CI</i>	<i>F</i>	<i>p</i>	<i>B</i>	<i>SE</i>	$\beta$	$R^2$	<i>95% CI</i>	<i>F</i>	<i>p</i>
Mission and Values Leadership	0.02	0.19	.01	.00	.00, .03	0.01	.91	0.11	0.09	.14	.02	.00, .11	1.57	.21
Strategy and Results	0.30	0.15	.22	.05	.00, .16	4.04	<.05	0.43	0.10	.44	.20	.06, .34	19.93	<.01
Talent	0.63	0.40	.17	.03	.00, .13	2.51	.12	0.32	0.11	.31	.10	.01, .23	8.87	<.01
Relationships	0.87	0.19	.44	.20	.06, .34	20.06	<.01	0.52	0.09	.54	.30	.14, .43	33.79	<.01
Communication	0.70	0.60	.13	.02	.00, .10	1.36	.25	0.60	0.10	.55	.30	.14, .44	35.33	<.01
Change	0.48	0.21	.25	.06	.00, .18	5.44	.02	0.35	0.09	.39	.14	.04, .29	14.53	<.01
Overall Score	0.86	0.16	.52	.27	.12, .41	30.02	<.01	0.61	0.10	.56	.32	.16, .45	37.99	<.01

Table 23. Linear regression of automated scores predicting original scores for the high-level managers' exercise.

	<b>Conservative Scores</b>							<b>Liberal Scores</b>						
	<i>B</i>	<i>SE</i>	$\beta$	$R^2$	<i>95% CI</i>	<i>F</i>	<i>p</i>	<i>B</i>	<i>SE</i>	$\beta$	$R^2$	<i>95% CI</i>	<i>F</i>	<i>p</i>
Mission and Values Leadership	0.64	0.39	.25	.06	.00, .23	2.77	.10	0.70	0.28	.36	.13	.00, .31	6.20	.02
Strategy and Results	0.27	0.43	.10	.01	.00, .13	0.39	.54	0.30	0.21	.22	.05	.00, .21	2.12	.15
Talent	0.89	0.43	.30	.09	.00, .27	4.22	<.05	0.50	0.28	.26	.07	.00, .24	3.22	.08
Relationships	0.49	0.21	.34	.12	.00, .30	5.71	.02	0.15	0.26	.09	.01	.00, .12	0.32	.58
Communication	1.01	0.45	.32	.11	.00, .29	5.05	.03	0.24	0.23	.16	.03	.00, .17	1.11	.30
Overall Score	0.59	0.30	.28	.08	.00, .25	3.72	.06	0.54	0.28	.28	.08	.00, .25	3.71	.06

Table 24. Linear regression of word count predicting Overall Scores

	<b>Mid-Level Managers' Exercise</b>							<b>High-Level Managers' Exercise</b>						
	<i>B</i>	<i>SE</i>	$\beta$	$R^2$	<i>95% CI</i>	<i>F</i>	<i>p</i>	<i>B</i>	<i>SE</i>	$\beta$	$R^2$	<i>95% CI</i>	<i>F</i>	<i>p</i>
Original Overall Score	104.99	35.15	.31	.10	.01, .23	8.92	<.01	41.53	38.78	.16	.03	.00, .17	1.15	.29
Conservative Overall Score	117.41	18.26	.58	.34	.18, .47	41.35	<.01	81.03	14.31	.65	.43	.20, .58	32.08	<.01
Liberal Overall Score	156.24	29.53	.51	.26	.11, .39	28.01	<.01	78.21	16.83	.58	.33	.12, .51	21.58	<.01

Table 25. Hierarchical regression of the automated score, word count and interaction term predicting the original score for the mid-level managers' exercise.

Model	Conservative Scores						Liberal Scores					
	<i>B</i>	<i>SE</i>	$\beta$	$\Delta R^2$	$\Delta F$	$\Delta p$	<i>B</i>	<i>SE</i>	$\beta$	$\Delta R^2$	$\Delta F$	$\Delta p$
1. Automated Score	0.31	0.06	.52	.27	30.02	<.01	0.52	0.08	.56	.32	37.99	<.01
2. Automated Score + Word Count	0.31 0.01	0.07 0.00	.51 .02	.00	0.03	.86	0.50 0.00	0.10 0.00	.54 .04	.00	0.14	.71
3. Automated Score + Word Count +Automated Score*Word Count	0.33 0.00 0.00	0.07 0.00 0.00	.52 .29 -.43	.09	11.77	<.01	0.51 0.00 0.00	0.09 0.00 0.00	.55 .40 -.53	.15	22.05	<.01

Table 26. Hierarchical regression of the automated score, word count and interaction term predicting the original core for the high-level managers' exercise.

Model	Conservative Scores						Liberal Scores					
	<i>B</i>	<i>SE</i>	$\beta$	$\Delta R^2$	$\Delta F$	$\Delta p$	<i>B</i>	<i>SE</i>	$\beta$	$\Delta R^2$	$\Delta F$	$\Delta p$
1. Automated Score	0.14	0.70	.28	.27	3.72	.06	0.15	0.08	.28	.08	0.00	.25
2. Automated Score + Word Count	0.15 0.00	0.09 0.00	.31 -.04	.00	0.04	.84	0.15 0.00	0.10 0.00	.28 .00	.00	0.00	<.01
3. Automated Score + Word Count +Automated Score*Word Count	0.15 0.00 0.00	0.10 0.00 0.00	.32 -.04 -.06	.09	0.14	.71	0.15 0.00 0.00	0.10 0.00 0.00	.28 .00 .02	.00	0.00	.06



Table 27. Factor loadings from the exploratory factor analysis (EFA)

Behaviors	Factor Loadings		
	1	2	3
1. Provided professional response	--	--	--
1a. Provided greeting	.14	.12	<b>.46*</b>
1b. Introduced self	<b>.46*</b>	.27	-.13
1c. Expressed appreciation or thankfulness	<b>.45*</b>	.18	-.03
1d. Provided salutation at the end of the email	.23	<b>.38*</b>	.06
1e. Provided name at the end of email	.26*	<b>.41*</b>	.21
1f. Provided position and/or company information at the end of the email	.20	<b>.47*</b>	.06
<i>1g. Used information or unprofessional language / abbreviations</i>	-.11	-.02	-.07
1h. Email title was inappropriate or potentially offensive to recipient	<b>.45*</b>	.05	-.01
1i. Forwarded inappropriate or potentially offensive content to recipient	<b>.33*</b>	-.02	-.004
2. Recognized feelings / emotions	--	--	--
2a. Expressed satisfaction, optimism, hope, positive anticipation, or personal dedication to something	<b>.44*</b>	-.07	.26*
2b. Expressed dissatisfaction, regret, disappointment, or apologized	<b>.27*</b>	-.01	.24*
<i>2c. Expressed empathy, considered another's feelings or related own feelings to others</i>	.18	.14	-.08
2d. Agreed with or showed understanding for a claim, statement, concern, or opinion	.15	<b>.35*</b>	-.13
3. Made a request	--	--	--
3a. Asked another to consider something a different way or reframe thinking	.11	.26	<b>.22*</b>
<i>3b. Asked another for a meeting / phone call / or discussion or implied it would be a good idea</i>	-.10	.20	.13
<i>3c. Asked another to provide information, verify information, give advice / direction / instruction or made an inquiry</i>	-.10	.17	.04
<i>3d. Asked another to respond</i>	-.15	.12	-.14
3e. Asked another to take action / delegated actions or express that help or assistance is needed	<b>-.50*</b>	-.20	.05
3f. Explained the details of the actions they should take or of the assistance that is needed	<b>-.45*</b>	-.24*	-.05
3g. Asked for feedback, suggestions, concerns, or additional input or implied this was welcome	-.02	<b>.68*</b>	-.32*
4. Provided information or an update	--	--	--
4a. Explained the current situation	.10	<b>.59*</b>	.31*
4b. Provided or offered information or explanations to maintain communication	<b>.44*</b>	.21	-.03
4c. Provided or offered another general advice, direction, or information about something	<b>.75*</b>	-.02	.22
4d. Provided or offered another detailed advice, direction, or information about something	<b>.26*</b>	.22	-.08
4e. Provided specific examples of something as part of an explanation	-.04	<b>.26*</b>	-.11

Table 27. Factor loadings from the exploratory factor analysis (EFA) (cont.)

Behaviors	Factor Loadings		
	1	2	3
4f. Sent someone information or offered information merely as a point of interest or to reference	.08	<b>.35*</b>	-.18
4g. Informed another that they would handle a situation or personally take action	.20	<b>.39*</b>	.06
<i>4h. Explained one's or another's current or desired role, responsibility, or expectation</i>	-.05	.20	-.15
<i>4i. Expressed intent to follow-up or expectation/hope of future contact</i>	-.01	.10	.14
4j. Suggested or brainstormed a strategic idea or potential plan or change that could be implemented in the future	-.08	<b>.35*</b>	-.01
5. Established a timeframe / timeline	--	--	--
5a. Communicated personal time constraints or implied a sense of urgency	-.17	-.03	<b>.44*</b>
<i>5b. Communicated current schedule / timeline / roadmap</i>	-.13	.09	-.14
6. Evaluated the situation	--	--	--
<i>6a. Implied or stated another had negative intentions, acted inappropriately, acted impulsively, did not have adequate information, or provided incomplete, false, or inappropriate information</i>	-.14	.18	-.13
6c. Was generally critical, negative, or pessimistic regarding something	-.18	<b>.56*</b>	.001
7. Showed positive recognition (for a company, project, program, product, or work)	--	--	--
<i>7a. Showed praise</i>	.15	-.05	.003
7b. Discussed value or importance	<b>.35*</b>	.01	.001
7c. Discussed the investment, commitment, time, or effort that has already been put in	-.02	.39	<b>.97*</b>
7d. Discussed the potential future success or trust, faith or confidence in it, dedication to it	.09	.15	<b>.29*</b>
7e. Discussed the goal, purpose, mission or intention of it	.27*	<b>.61*</b>	-.04
8. Showed positive recognition (for a person, relationship, or collaboration)	--	--	--
8a. Showed praise	<b>-.36*</b>	-.09	-.09
8b. Discussed value or importance	.26*	<b>.29*</b>	.03
8d. Discussed the potential future success or trust, faith or confidence in it, dedication to it	-.16	.05	<b>.30*</b>
9. Spelling or grammar issue	--	--	--
9a. Had an issue with the spelling or grammar used in the email	-.18	<b>.58*</b>	-.003
9b. Addressed or referenced someone using the wrong, incomplete, or misspelled name/title	-.04	<b>.45*</b>	.04

Note. Items 6b and 8c were dropped from the EFA for having no variance. Italicized values did not significantly load on any of the factors in the 3-factor model. Bolded values indicate the factor the item loaded best on

## REFERENCES

- Aguinis, H., Henle, C. A., & Beaty, J. C. (2001). Virtual reality technology: A new tool for personnel selection. *The International Journal of Selection and Assessment*, 9(1/2), 70-83.
- Anderson, N. (2003). Applicant and recruiter reactions to new technology in selection: A critical review and agenda for future research. *International Journal of Selection and Assessment*, 11(2-3), 121-136. doi:1.1111/1468-2389.00235
- Arthur, W. R., Day, E., McNelly, T. L., & Edens, P. S. (2003). A meta-analysis of the criterion-related validity of assessment center dimensions. *Personnel Psychology*, 56(1), 125-154. doi:1.1111/j.1744-657.2003.tb00146.x
- Balogh, J., Bernstein, J., Cheng, J., Van Moere, A., Townshend, B., & Suzuki, M. (2012). Validation of automated scoring of oral reading. *Educational and Psychological Measurement*, 72(3), 435-452. doi:10.1177/0013164411412590
- Bandalos, D. L., & Boehm-Kaufman, M. R. (2009). Four common misconceptions in exploratory factor analysis. In C. E. Lance & R. J. Vandenberg (Eds.), *Statistical and Methodological Myths and Urban Legends*, 61-87.
- Barrios, M., Villarroya, A., Borrego, Á., & Ollé, C. (2011). Response rates and data quality in web and mail surveys administered to PhD holders. *Social Science Computer Review*, 29(2), 208-22. doi:1.1177/0894439310368031
- Bennett, R. E., & Bejar, I. I. (1998). Validity and automated scoring: It's not only the scoring. *Educational Measurement: Issues and Practice*, 17(4), 9-17.
- Black, I. (2006). The presentation of interpretivist research. *Qualitative Market Research: An International Journal*, 9(4), 319-324.
- Boyce, T. (2013). IGNITE lightning round: Simulations and personnel selection. *Presentation at the 28<sup>th</sup> Annual Society for Industrial and Organizational Psychology Conference*. Houston, TX.
- Breevaart, K., Bakker, A., Hetland, J., Demerouti, E., Olsen, O. K., & Espevik, R. (2014). Daily transactional and transformational leadership and daily employee engagement. *Journal of Occupational and Organizational Psychology*, 87(1), 138-157. doi:10.1111/joop.12041
- Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 56, 81-105. doi:10.1037/h0046016
- Cash, A. H., Hamre, B. K., Pianta, R. C., & Myers, S. S. (2012). Rater calibration when observational assessment occurs at large scale: Degree of calibration and characteristics

- of raters associated with calibration. *Early Childhood Research Quarterly*, 27(3), 529-542. doi:10.1016/j.ecresq.2011.12.006
- Cherniack, E. (2011). Not just fun and games: Applications of virtual reality in the identification and rehabilitation of cognitive disorders of the elderly. *Disability and Rehabilitation: Assistive Technology*, 6(4), 283-289. doi:10.3109/17483107.2010.542570
- Clauser, B. E., Margolis, M. J., Clyman, S. G., & Ross, L. P. (1997). Development of automated scoring algorithms for complex performance assessments: A comparison of two approaches. *Journal of Educational Measurement*, 34(2), 141-161.
- Colwell, K., Hiscock, C. K., & Memon, A. (2002). Interviewing techniques and the assessment of statement credibility. *Applied Cognitive Psychology*, 16(3), 287-300. doi:10.1002/acp.788
- Cook, K. E. (2012). Reliability assessments in qualitative health promotion research. *Health Promotion International*, 27(1), 90-101. doi:1.1093/heapro/dar027
- Creswell, J. W. (1998). *Qualitative inquiry and research design: Choosing among five traditions*. Thousand Oaks, CA: Sage Publishing.
- DeMaria, S. Jr., Bryson, E. O., Mooney, T. J., Silverstein, J. H., Reich, D. L., Bodian C., & Levine, A. I. (2010). Adding emotional stressors to training in simulated cardiopulmonary arrest enhances participant performance. *Medical Education*, 44, 1006-1015. doi:1.1111/j.1365-2923.201.03775.x
- Denby, B. & Schofield, D. (1999). Role of virtual reality in safety training of mine personnel. *Mining Engineering*, 51, 59-64.
- Egri, C. P., & Herman, S. (2000). Leadership in the North American environmental sector: Values, leadership styles, and contexts of environmental leaders and their organizations. *Academy of Management Journal*, 43(4), 571-604. doi:10.2307/1556356
- Elo, S., & Kyngas, H. (2008). The qualitative content analysis process. *Journal of Advanced Nursing*, 62, 107-115.
- Ericson, R. V. (2006). Ten uncertainties of risk-management approaches to security. *Canadian Journal of Criminology and Criminal Justice*, 48(3), 345-357. doi:1.1353/ccj.2006.0028
- Francis, J. J., Johnston, M., Robertson, C., Glidewell, L., Entwistle, V., Eccles, M. P., & Grimshaw, J. M. (2010). What is an adequate sample size? Operationalizing data saturation for theory-based interview studies. *Psychological Health*, 25(10), 1229-1245.
- Gibbons, A. M., Hughes, D., Riley, P., Thornton, G. C., & Sanchez, D. S. (2013). Is the future here? Assessment Center technology use and benefits. *Paper Submitted to the Academy of Management Annual Conference 2013*.

- Gordon, C., & Buckley, T. (2009). The effect of high-fidelity simulation training on medical-surgical graduate nurses' perceived ability to respond to patient clinical emergencies. *The Journal of Continuing Education In Nursing, 40*(11), 491-498. doi:10.3928/00220124-20091023-06
- Haig, B. D. (2005). Exploratory factor analysis, theory generation, and scientific method. *Multivariate Behavioral Research, 40*, 303-329.
- Handler, C. A. (2013). IGNITE lightning round: Simulations and personnel selection. *Presentation at the 28<sup>th</sup> Annual Society for Industrial and Organizational Psychology Conference*. Houston, TX.
- Hatfield, J. (2013). High definition animations: Enhancing realistic job preview and organization perceptions. *Paper presented at the 28<sup>th</sup> Annual Society for Industrial and Organizational Psychology Conference*. Houston, TX.
- Hawkes, B. (2013). Test-Takers' Empathy for animated humans in SJTs. *Paper presented at the 28<sup>th</sup> Annual Society for Industrial and Organizational Psychology Conference*. Houston, TX.
- Hoffman, C. C., & Thornton, G. (1997). Examining selection utility where competing predictors differ in adverse impact. *Personnel Psychology, 50*(2), 455-470. doi:10.1111/j.1744-6570.1997.tb00916.x
- Howell, J. M., & Avolio, B. J. (1993). Transformational leadership, transactional leadership, locus of control, and support for innovation: Key predictors of consolidated-business-unit performance. *Journal Of Applied Psychology, 78*(6), 891-902. doi:10.1037/0021-9010.78.6.891
- Hsieh, H. F., & Shannon, S. E. (2005). Three approaches to qualitative content analysis. *Quality Health Research, 15*, 1277-1288.
- International Task Force on Assessment Center Guidelines. (2009). Guidelines and ethical considerations for assessment center operations. *International Journal of Selection and Assessment, 17*, 243-253. doi: 1.1111/j.1468-2389.2009.00467.x
- Jordan, S. (2012). Student engagement with assessment and feedback: Some lessons from short-answer free-text e-assessment questions. *Computers & Education, 58*(2), 818-834. doi:10.1016/j.compedu.2011.10.007
- Kapoulas, A. & Mitic, M. (2012). Understanding challenges of qualitative research: Rhetorical issues and reality traps. *Qualitative Market Research: An International Journal, 15*(4), 354-368. doi:1.1108/13522751211257051
- Katz, M. L. & Shapiro, C. (1986). Technology adoption in the presence of network externalities. *Journal of Political Economy, 94*(4), 822-841.

- Kaufman, J. C., Baer, J., Cropley, D. H., Reiter-Palmon, R., & Sinnett, S. (2013). Furious activity vs. understanding: How much expertise is needed to evaluate creative work? *Psychology of Aesthetics, Creativity, and the Arts*, 7(4), 332-340. doi:10.1037/a0034809
- Klassen, S. (2006). Contextual Assessment in Science Education: Background, Issues, and Policy. *Science Education*, 90(5), 820-851. doi:10.1002/sce.20150
- König, C. J., Klehe, U., Berchtold, M., & Kleinmann, M. (2010). Reasons for being selective when choosing personnel selection procedures. *International Journal of Selection & Assessment*, 18(1), 17-27. doi:1.1111/j.1468-2389.201.00485.x
- Krause, D. E., & Thornton, G. (2009). A cross-cultural look at assessment center practices: Survey results from Western Europe and North America. *Applied Psychology: An International Review*, 58(4), 557-585. doi:1.1111/j.1464-0597.2008.00371.x
- Latorre, J. (2013). IGNITE lightning round: Simulations and personnel selection. *Presentation at the 28<sup>th</sup> Annual Society for Industrial and Organizational Psychology Conference*. Houston, TX.
- Laupacis, A., Feeny, D., Detsky, A. S., & Tugwell, P. X. (1992). How attractive does a new technology have to be to warrant adoption and utilization? Tentative guidelines for using clinical and economic evaluations. *Canadian Medical Association Journal*, 15(4), 473-481.
- Leopold, T. (2011). The death and life of a great American bookstore. *CNN*. Retrieved from <http://www.cnn.com/2011/US/09/12/first.borders.bookstore.closing/index.html>
- Lievens, F., & De Soete, B. (2012). Simulations. In N. Schmitt (Ed.), *The Oxford handbook of personnel assessment and selection* (pp. 383-410). New York, NY: Oxford University Press. doi:1.1093/oxfordhb/9780199732579.013.0017
- Lievens, F., & Patterson, F. (2011). The validity and incremental validity of knowledge tests, low-fidelity simulations, and high-fidelity simulations for predicting job performance in advanced-level high-stakes selection. *Journal of Applied Psychology*, 96(5), 927-94. doi:1.1037/a0023496
- Lievens, F., Van Keer, E., & Volckaert, E. (2010). Gathering behavioral samples through a computerized and standardized assessment center exercise: Yes, it is possible. *Journal of Personnel Psychology*, 9(2), 94-98. doi:1.1027/1866-5888/a000010
- Liu, O., Lee, H., & Linn, M. C. (2011). An investigation of explanation multiple-choice items in science assessment. *Educational Assessment*, 16(3), 164-184. doi:10.1080/10627197.2011.611702
- Lord, R. G. (1985). Accuracy in behavioral measurement: An alternative definition based on raters' cognitive schema and signal detection theory. *Journal of Applied Psychology*, 70(1), 66-71. doi:10.1037/0021-9010.70.1.66

- Luciana, M. (2003). Practitioner review: Computerized assessment of neuropsychological function in children: Clinical and research applications of the Cambridge Neuropsychological Testing Automated Battery (CANTAB). *Journal of Child Psychology & Psychiatry & Allied Disciplines*, 44(5), 649-663.
- Luck, J., Peabody, J. W., & Lewis, B. L. (2006). An automated scoring algorithm for computerized clinical vignettes: Evaluating physician performance against explicitly quality criteria. *International Journal of Medical Informatics*, 75, 701-707.
- Mania, K., & Robinson, A. (2005). An experimental exploration of the relationship between subjective impressions of illumination and physical fidelity. *Computers & Graphics*, 29(1), 49-56. doi:10.1016/j.cag.2004.11.007
- Marsch, L. A. (2011). Technology-based interventions targeting substance use disorders and related issues: An editorial. *Substance Use & Misuse*, 46(1), 1-3. doi:1.3109/10826084.2011.521037
- Marsh, K., & Fayek, A. (2010). Surety assist: Fuzzy expert system to assist surety underwriters in evaluating construction contractors for bonding. *Journal of Construction Engineering & Management*, 136(11), 1219-1226. doi:1.1061/(ASCE)CO.1943-7862.0000224
- Mayring P. (2000). Qualitative content analysis. In U. Flick, E. von Kardoff, I. Steinke (Eds.), *A Companion to Qualitative Research* (pp. 266-269). Thousand Oaks, CA: Sage Publications Inc.
- Moretti, F., van Vliet, L., Bensing, J., Deledda, G., Mazzi, M., Rimondini, M., Zimmermann, C. & Fletcher, I. (2011). A standardized approach to qualitative content analysis of focus group discussions from different countries. *Patient Education and Counseling*, 82(3), 420-428. doi:1.1016/j.pec.2011.01.005
- Morse, J. M. (1989). Strategies for sampling. In J. M. Morse (Ed.), *Qualitative nursing research: A con-temporary dialogue* (pp. 117- 131). Rockville, MD: Aspen.
- Mostyn, B. (1985). The content analysis of qualitative research data: A dynamic approach. In M. Brenner, J. Brown, & D. Cauter (Eds.), *The research interview* (pp. 115–145). London: Academic Press.
- Mouritsen, M. (2013). Is your organization managing or mangling its technology assets? *Strategic Finance*, 95(1), 35-41.
- Neimeyer, G. J., & Gemignani, M. (2003). Qualitative methods. In R. Fernández-Ballesteros (Ed.), *Encyclopedia of psychological assessment* (pp. 785–800). Sage: London.
- O’Connell, M. (2013). IGNITE lightning round: Simulations and personnel selection. *Presentation at the 28<sup>th</sup> Annual Society for Industrial and Organizational Psychology Conference*. Houston, TX.

- O'Donoghue, T., & Punch, K. (2003). *Qualitative educational research in action: Doing and reflecting*. New York, NY: Routledge.
- Ostrom, J. K., Born, M. H., Serlie, A. W., & van der Molen, H. T. (2012). Implicit trait policies in multimedia situational judgment tests for leadership skills: Can they predict leadership behavior? *Human Performance, 25*(4), 335-353. doi:1.1080/08959285.2012.703732
- Papadopoulos, E. (2013). The predictive validity of a technology enhanced assessment center method for personnel selection. *Dissertation Abstracts International, 73*.
- Ployhart, R. E. (2006). Staffing in the 21st Century: New Challenges and Strategic Opportunities. *Journal of Management, 32*(6), 868-897. doi: 10.1177/0149206306293625
- Powers, D. E., Burstein, J. C., Chodorow, M., Fowles, M. E., & Kukich, K. (2002). Stumping e-rater: Challenging the validity of automated essay scoring. *Computers in Human Behavior, 18*(2), 103-134. doi: 10.1016/s0747-5632(01)00052-8
- Roth, P., Bobko, P., & McFarland, L. A. (2005). A meta-analysis of work sample test validity: Updating and integrating some classic literature. *Personnel Psychology, 58*, 1009–1037. doi:1.1111/j.1744- 657.2005.00714.x
- Rupp, D. E., Gibbons, A. M., & Synder, L. A. (2008). The role of technology in enabling third-generation training and development. *Industrial and Organizational Psychology: Perspectives on Science and Practice, 1*(4), 496-50. doi:1.1111/j.1754-9434.2008.00095.x
- Ryan, A., & Greguras, G. J. (1998). Life is not multiple choice: Reactions to the alternatives. In M. D. Hakel (Ed.), *Beyond multiple choice: Evaluating alternatives to traditional testing for selection* (pp. 183-202). Mahwah, NJ: Lawrence Erlbaum Associates Publishers.
- Sackett, P. R., & Dreher, G. F. (1982). Constructs and assessment center dimensions: Some troubling empirical findings. *Journal of Applied Psychology, 67*, 401–410. doi:10.1037/0021-9010.67.4.401
- Sandelowski, M. (1995). Sample size in qualitative research. *Research in Nursing and Health, 18*, 179–183. doi: 10.1002/nur.4770180211
- Schilling, J. (2006). On the pragmatics of qualitative assessment: Designing the process for content analysis. *European Journal of Psychological Assessment, 22*(1), 28-37. doi:1.1027/1015-5759.22.1.28
- Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin, 124*(2), 262-274. doi:10.1037/0033-2909.124.2.262
- Scott, J. C., Morelli, N., Golubovich, J., Impleman, K., Lawrence, A., Illingworth, A. J., Gutierrez, S. L., & Mead, A. D. (2013). Mobile devices in talent assessment: Where are



we now? *Symposium presented at the 28<sup>th</sup> Annual Society for Industrial and Organizational Psychology Conference*. Houston, TX.

- Shermis, M. D., Shneyderman, A., & Attali, Y. (2008). How important is content in the ratings of essay assessments? *Assessment in Education: Principles, Policy & Practice*, 15(1), 91-105. doi:1.1080/09695940701876219
- Sinnott, S., Guinane, M., Whelton, H., & Byrne, S. (2013). Is 50 cent the price of the optimal copayment? A qualitative study of patient opinions and attitudes in response to a 50 cent charge on prescription drugs in a publicly funded health system in Ireland. *BMC Health Services Research*, 13(1), 1-9. doi:10.1186/1472-6963-13-16
- SIOP. (2013). 2013 Searchable Database. *Sessions for the 28<sup>th</sup> Annual Society for Industrial and Organizational Psychology Conference*. Houston, TX. Retrieved from <http://www.siop.org/programsearChdefault.aspx?conid=7>
- SIOP. (2014). 2014 Searchable Database. *Sessions for the 29<sup>th</sup> Annual Society for Industrial and Organizational Psychology Conference*. Honolulu, HI. Retrieved from <http://www.siop.org/programsearChdefault.aspx?conid=8>
- Sitzmann, T., Casper, W. J., Brown, K. G., Ely, K., & Zimmerman, R. D. (2008). A review and meta-analysis of the nomological network of trainee reactions. *Journal of Applied Psychology*, 93(2), 280-295.
- Sosik, J. J., & Megerian, L. E. (1999). Understanding leader emotional intelligence and performance: The role of self-other agreement on transformational leadership perceptions. *Group & Organization Management*, 24(3), 367-390. doi:10.1177/1059601199243006
- Spörrle, M., Gerber-Braun, B., & Fösterling, F. (2007). The influence of response lines on response behavior in the context of open-question formats. *Swiss Journal of Psychology / Schweizerische Zeitschrift für Psychologie / Revue Suisse de Psychologie*, 66(2), 103-107.
- Sulsky, L. M., & Day, D. V. (1992). Frame-of-reference training and cognitive categorization: An empirical investigation of rater memory issues. *Journal of Applied Psychology*, 77(4), 501-510. doi:10.1037/0021-9010.77.4.501
- Tannenbaum, S. I., Mathieu, J. E., Salas, E., & Cohen, D. (2012). Teams are changing: Are research and practice evolving fast enough? *Industrial and Organizational Psychology*, 5(1), 2-24. doi:1.1111/j.1754-9434.2011.01396.x
- Tekin-Koru, A. (2009). Technology transfers and optimal entry strategies for the multinational firm. *Journal of International Trade & Economic Development*, 18(4), 553-574. doi:10.1080/09638190903217503
- Thornton, G. C., & Potemra, M. J. (2010). Utility of assessment center for promotion of police sergeants. *Public Personnel Management*, 39(1), 59-69.

- Thornton, G. C., III & Rupp, D. (2006). *Assessment centers in human resource management: Strategies for prediction, diagnosis, and development*. Mahwah, NJ: Erlbaum.
- Tziner, A., Meir, E. I., Dahan, M., & Birati, A. (1994). An investigation of the predictive validity and economic utility of the assessment center for the high-management level. *Canadian Journal of Behavioural Science/Revue canadienne des sciences du comportement*, 26(2), 228.
- Ward, P., Williams, A., & Hancock, P. A. (2006). Simulation for performance and training. In K. Ericsson, N. Charness, P. J. Feltovich, & R. R. Hoffman (Eds.), *The Cambridge handbook of expertise and expert performance* (pp. 243-262). New York, NY: Cambridge University Press.
- Weekley, J. A., Ployhart, R. E., & Holtz, B. C. (2006). On the development of situational judgment tests: Issues in item development, scaling, and scoring. In J. A. Weekley & R. E. Ployhart (Eds.), *Situational judgment tests* (pp. 157–182). San Francisco, CA: Jossey-Bass.
- Williamson, D. M. (2012). The conceptual and scientific basis for automated scoring of performance items. In R. W. Lissitz, H. Jiao (Eds.), *Computers and their impact on state assessments: Recent history and predictions for the future* (pp. 157-193). Charlotte, NC: Information Age Publishing.
- Williamson, D. M., Xi, X., & Breyer, F. J. (2012). A framework for evaluation and use of automated scoring. *Educational Measurement: Issues and Practice*, 31(1), 2-13.
- Wood, A. & McPhee, C. (2011). Establishing a virtual learning environment: A nursing experience. *The Journal of Continuing Education in Nursing*, 42(11), 510-515. doi:1.3928/00220124-20110715-01
- Xi, X., Higgins, D., Zechner, K., & Williamson, D. (2012). A comparison of two scoring methods for an automated speech scoring system. *Language Testing*, 29(3), 371-394. doi:1.1177/0265532211425673
- Ya-Chen Tina, S., I-Wen, P., & Yi-Wen, T. (2009). Information technology facilitates cost-effectiveness analysis in developing countries: an observational study of breast cancer chemotherapy in Taiwan. *Pharmacoeconomics*, 27(11), 947-961.
- Yardley, S. (2011). Death is not the only harm: Psychological fidelity in simulation. *Medical Education*, 45(10). doi:1.1111/j.1365-2923.2011.04029.x
- Zajtchuk, R. & Satava, R. M. (1997). Medical applications of virtual reality. *Communications of the ACM*, 40, 63-64.
- Zhang, M., Williamson, D. M., Breyer, F., & Trapani, C. (2012). Comparison of e-rater<sup>®</sup> automated essay scoring model calibration methods based on distributional targets. *International Journal of Testing*, 12(4), 345-364. doi:1.1080/15305058.2011.64597

## APPENDIX I

## 1. Cover Statement

This brief survey is part of a research study being conducted by Diana Sanchez, M.S. and Dr. Alyssa M. Gibbons at Colorado State University. We are seeking the opinions of experienced assessors regarding the relevance of several common behaviors taken by participants in assessment center exercises. Your input will be used to help us determine the feasibility of various scoring methods within an assessment center context.

Specifically, we would like you to read over the exercise materials and a list of common behaviors participants take in response to this exercise. We first ask that participants link each of the behaviors to the most applicable competency from a provided list and then rate each behavior on effectiveness with regard to successful performance on the exercise. The exercise is fairly brief; we anticipate that the entire task should take about 15-20 minutes.

So that we can be confident that our results reflect the views of experienced assessors, we ask that you participate only if you have personally worked as an assessor in an assessment center and assessed at least 3 different individuals.

The survey is anonymous and we will not ask you to provide any information that could identify you personally. As a thank you for your time, we would like to send a \$5 Starbucks gift card to the first 10 respondents to complete the survey. You will be taken to a separate survey after your participation to enter your e-mail address to receive the Starbucks gift card. The surveys are not linked in any way, so your responses cannot be connected with your e-mail address. After the first 10 responses, the survey will be closed. Please note that the gift cards are only valid at Starbucks locations within the United States.

Your participation in this research is entirely voluntary, and you are free to withdraw at any time (though you must reach the end of the survey to be eligible for the Starbucks gift card). We anticipate no risks arising from participating in this study that are greater than the risks of daily life.

If you agree to participate in the survey, please select the “I agree to participate” button below. If you have any questions about this study or are interested in the results once they become available, please contact Dr. Alyssa Gibbons (970-491-4940; [alyssa.gibbons@colostate.edu](mailto:alyssa.gibbons@colostate.edu)) or Diana Sanchez, M.S. ([diana.sanchez@colostate.edu](mailto:diana.sanchez@colostate.edu)). If you have questions about your rights as a volunteer in this research, contact the Colorado State University IRB, at [RICRO\\_IRB@mail.colostate.edu](mailto:RICRO_IRB@mail.colostate.edu).

Thank you so much for contributing your valuable time to our research!

\_\_\_\_\_ I agree to participate

## 2. Screening Questions

1. Have you worked as an assessor in an assessment center program?

Yes

No

2. Approximately how many different individual participants have you assessed?

0-2

3+

*[Note: Participants who indicate that they have not worked in an assessment center or that have assessed fewer than 3 participants will be directed to the following message:]*

Thank you for your interest, but you do not meet the inclusion criteria to participate in this study.

*[All other participants will proceed to the instructions in the next sections and through random selection either complete 4a-7a or 4b-7b]*

#### **4a. Instructions – Mid-Level Manager Instructions:**

As an experienced assessor, you may have noticed that some behavioral responses to assessment center exercises can lead to more effective performance than others. The following is a survey regarding effective leadership behaviors in response to an assessment center exercise. We would like your opinions about the relevance of several commonly chosen behaviors to particular simulation exercises we are using as part of a research study at Colorado State University.

This survey is broken into two distinct tasks. In Task 1 you will be asked to link various behaviors to a list of competencies. In Task 2 you will be asked to rate the effectiveness of those same behaviors in response to the exercise scenario given below. Of course, the full assessment center process includes multiple exercises and other measures, but at present, we are only concerned with this particular exercise.

This exercise is designed for use with industry professionals in leadership positions who take part in a simulation where they are playing the role of a manager within a fictitious medical organization. The exercise asks participants to respond to a customer concern that they received in the following scenario.

*Scenario: Participants enter the assessment center playing the role of a manager on their first day of work with a new company. The participant is told they will be leaving for a week of business travel later this afternoon and must complete all tasks and resolve all issues within the next 3 hours before they leave for their trip.*

Early in the assessment center the participant completes a live role-play exercise where they are asked to complete a performance review with one of their subordinates. Later that day that same subordinate sends them an email asking for assistance in handling a customer complaint they have received. The customer's email shows that they are very dissatisfied and upset by a new program the company has recently rolled out.

You will first be asked to link the provided list of behaviors to the list of competencies. Then you will be providing effectiveness ratings (i.e., -5 Very Ineffective to +5 Very Effective) on those same behavioral responses that a participant might take in response to this exercise prompt. Please take a few minutes to consider what successful performance would look like for this exercise.

When you are ready, please click “Continue” to proceed with the survey.

\_\_\_\_\_ Continue

## 5a. Task 1 - Linking Competencies

Below are names and brief definitions for the 6 competencies used in development of the assessment center exercises.

As part of Task 1, you are asked to read each behavior and decide which competencies each behavior links to. Please link each behavior to at least one competency. If a competency doesn't link to a behavior, please leave the space blank.

Please link a competency by typing in a number that you believe best represents the relationship between the behavior and the competency (i.e., 1 strong and clear link, 2 somewhat link that may be situation dependent, or 3 weak or vague link but still worth noting).

Make sure to base your judgment on the definitions provided here, not just the name of the competency.

### **Competencies:**

**Mission and Values Leadership:** Leads commitment to district mission and values that motivate and engage the team and others.

**Strategy and Results:** Demonstrates broad thinking and decision-making focused on district goals and succeeds in accomplishing these goals.

**Talent:** Seeks opportunities for self, team, and overall district to learn and develop.

**Relationships:** Builds strong relationships with key internal and external customers and partners.

**Communication:** Influences others through effective communication about district priorities.

**Change:** Supports change initiatives within the district by seeking opportunities for continual improvement.

1 - strong and clear link

2 - somewhat link that may be situation dependent

3 - weak or vague link but still worth noting

	MVL	S&R	T	R	Com	Ch
1. Provided greeting (e.g., "Hi," "Good Morning," "Dear _____," and NOT just name "Taylor")	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2. Introduced self (e.g., "My name is _____, I am the Manager...")	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3. Expressed appreciation or thankfulness (NOT just "thanks" or "thank you")	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4. Provided salutation at the end of the email (e.g., "Sincerely," "Regards," "Thanks," "Thank you,")	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
5. Provided name at the end of the email	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
6. Provided position and/or company information at the end of the email (e.g., "Manager of ...")	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
7. Used informal or unprofessional language / abbreviations (e.g., "thru" "ASAP")	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
8. Email title was inappropriate or potentially offensive to recipient (e.g., "complaint" in an email response sent to a customer)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
9. Forwarded content that was inappropriate or potentially offensive to recipient (e.g., "she is pretty upset and demanding a response" forwarded to a customer in a response)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
10. Expressed satisfaction, optimism, hope, positive anticipation, or personal dedication to something	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
11. Expressed dissatisfaction, regret, disappointment, or apologized	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
12. Expressed empathy, considered another's feelings or related own feelings to others	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
13. Agreed with or showed understanding for a claim, statement, concern, or opinion	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
14. Asked another to consider something a different way or re-frame thinking	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
15. Asked another for a meeting / phone call / or discussion or implied it would be a good idea	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
16. Asked another to provide information, verify information, give advice/direction/instruction or made an inquiry (e.g., "Please copy me on the email")	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>



	MVLS&R	T	R	Com	Ch
17. Asked another to respond (e.g., "Please email me later today")	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
18. Asked another to take action / delegated actions or express that help or assistance is needed	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
19. Explained the details of the actions they should take or of the assistance that is needed (e.g., provided an outline, gave bullet points, explained what should be done, said or explained, told someone how to do something)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
20. Asked for feedback, suggestions, concerns, or additional input or implied this was welcome	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
21. Explained the current situation (e.g., gave a summary, an overview explanation, or very general information as an update such as "Taylor received a customer complaint")	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
22. Provided or offered information or explanations to maintain communication (e.g., "If you have any questions I would be happy to address them")	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
23. Provided or offered another general advice, direction, or information about something (NOT delegating)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
24. Provided or offered another detailed advice, direction, or information about something (NOT delegating)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
25. Provided specific examples of something as part of an explanation (e.g., "For example...")	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
26. Sent someone information or offered information merely as a point of interest or to reference (e.g., forwarded an email as an FYI, "See below")	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
27. Informed another that they would handle a situation or personally take action (e.g., "I will respond to Suzanne directly" NOT general or detailed information on what would be done)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
28. Explained one's or another's current or desired role, responsibility, or expectation (e.g., "In the future I expect you to take action when a customer complaint is received")	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
29. Expressed intent to follow-up or expectation/hope of future contact (e.g., "I look forward future conversations" or "I will email you as soon as I...")	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
30. Suggested or brainstormed a strategic idea or potential plan or change that could be implemented in the future	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
31. Communicated personal time constraints or implied a sense of urgency (e.g., "right away" "later today" "before I leave today" "before our conference call")	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

	<b>MVL</b>	<b>S&amp;R</b>	<b>T</b>	<b>R</b>	<b>Com</b>	<b>Ch</b>
32. Communicated current schedule / timeline / road map	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
33. Implied or stated another had negative intentions, acted inappropriately, acted impulsively, did not have adequate information, or provided incomplete, false, or inappropriate information	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
34. Implied or stated false, misleading, or deceitful information	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
35. Was generally critical, negative, or pessimistic regarding something	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
36. Showed praise (for a company, project, program, product, or work completed)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
37. Discussed value or importance (for a company, project, program, product, or work completed)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
38. Discussed the investment, commitment, time, or effort that has already been put in (for a company, project, program, product, or work completed)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
39. Discussed the potential future success or the trust, faith or confidence in it, dedication to it (for a company, project, program, product, or work completed)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
40. Discussed the goal, purpose, mission or intention of it (for a company, project, program, product, or work completed)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
41. Showed praise (for a person, relationship, or collaboration)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
42. Discussed value or importance (for a person, relationship, or collaboration)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
43. Discussed the investment, commitment, time, or effort that has already been put in (for a person, relationship, or collaboration)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
44. Discussed potential future success or the trust, faith or confidence in another, dedication to it (for a person, relationship, or collaboration)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
45. Had an issue with the spelling or grammar used in the email	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
46. Addressed or referenced someone using the wrong, incomplete, or misspelled name/title	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>



Please consider and rate the following behaviors as though they were in a written response to the scenario. Please provide a rating on a scale of -5 to 5.

- 5 Ineffective: this action would be worse than doing nothing
- 0 Neutral: this action would have no meaningful impact
- 5 Effective: this action would be better than doing nothing

	-5	-4	-3	-2	-1	0	1	2	3	4	5
17. Asked another to respond (e.g., "Please email me later today")	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
18. Asked another to take action / delegated actions or express that help or assistance is needed	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
19. Explained the details of the actions they should take or of the assistance that is needed (e.g., provided an outline, gave bullet points, explained what should be done, said or explained, told someone how to do something)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2. Asked for feedback, suggestions, concerns, or additional input or implied this was welcome	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
21. Explained the current situation (e.g., gave a summary, an overview explanation, or very general information as an update such as "Taylor received a customer complaint")	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
22. Provided or offered information or explanations to maintain communication (e.g., "If you have any questions I would be happy to address them")	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
23. Provided or offered another general advice, direction, or information about something (NOT delegating)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
24. Provided or offered another detailed advice, direction, or information about something (NOT delegating)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
25. Provided specific examples of something as part of an explanation (e.g., "For example...")	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
26. Sent someone information or offered information merely as a point of interest or to reference (e.g., forwarded an email as an FYI, "See below")	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
27. Informed another that they would handle a situation or personally take action (e.g., "I will respond to Suzanne directly" NOT general or detailed information on what would be done)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
28. Explained one's or another's current or desired role, responsibility, or expectation (e.g., "In the future I expect you to take action when a customer complaint is received")	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
29. Expressed intent to follow-up or expectation/hope of future contact (e.g., "I look forward future conversations" or "I will email you as soon as I...")	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3. Suggested or brainstormed a strategic idea or potential plan or change that could be implemented in the future	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
31. Communicated personal time constraints or implied a sense of urgency (e.g., "right away" "later today" "before I leave today" "before our conference call")	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Please consider and rate the following behaviors as though they were in a written response to the scenario. Please provide a rating on a scale of -5 to 5.

- 5 Ineffective: this action would be worse than doing nothing
- 0 Neutral: this action would have no meaningful impact
- 5 Effective: this action would be better than doing nothing

	-5	-4	-3	-2	-1	0	1	2	3	4	5
32. Communicated current schedule / timeline / road map	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
33. Implied or stated another had negative intentions, acted inappropriately, acted impulsively, did not have adequate information, or provided incomplete, false, or inappropriate information	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
34. Implied or stated false, misleading, or deceitful information	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
35. Was generally critical, negative, or pessimistic regarding something	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
36. Showed praise (for a company, project, program, product, or work completed)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
37. Discussed value or importance (for a company, project, program, product, or work completed)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
38. Discussed the investment, commitment, time, or effort that has already been put in (for a company, project, program, product, or work completed)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
39. Discussed the potential future success or the trust, faith or confidence in it, dedication to it (for a company, project, program, product, or work completed)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4. Discussed the goal, purpose, mission or intention of it (for a company, project, program, product, or work completed)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
41. Showed praise (for a person, relationship, or collaboration)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
42. Discussed value or importance (for a person, relationship, or collaboration)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
43. Discussed the investment, commitment, time, or effort that has already been put in (for a person, relationship, or collaboration)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
44. Discussed potential future success or the trust, faith or confidence in another, dedication to it (for a person, relationship, or collaboration)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
45. Had an issue with the spelling or grammar used in the email	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
46. Addressed or referenced someone using the wrong, incomplete, or misspelled name/title	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

## **7a. End of Survey**

Thank you so much for your help! As a thank you for your time, we would like to offer you a \$5 Starbucks gift card (please note that these are only valid for Starbucks locations in the United States). In order to deliver your gift card, we need a valid email address for you. To preserve the anonymity of your survey responses, we will use a separate survey to collect your email address.

The two surveys are not linked in any way, and we will be unable to connect your email address with your survey responses.

To enter your email address, please click here:

<https://www.surveymonkey.com/s/8JVXBTR>

#### **4b. Instructions – High-Level Manager Instructions:**

As an experienced assessor, you may have noticed that some behavioral responses to assessment center exercises can lead to more effective performance than others. The following is a survey regarding effective leadership behaviors in response to an assessment center exercise. We would like your opinions about the relevance of several commonly chosen behaviors to particular simulation exercises we are using as part of a research study at Colorado State University. This survey is broken into two distinct tasks. In Task 1 you will be asked to link various behaviors to a list of competencies. In Task 2 you will be asked to rate the effectiveness of those same behaviors in response to the exercise scenario given below. Of course, the full assessment center process includes multiple exercises and other measures, but at present, we are only concerned with this particular exercise.

This exercise is designed for use with industry professionals in leadership positions who take part in a simulation where they are playing the role of a manager within a fictitious medical organization. The exercise asks participants to respond to a customer concern that they received in the following scenario.

*Scenario: Participants enter the assessment center playing the role of a manager on their first day of work with a new company. The participant is told they will be leaving for a week of business travel later this afternoon and must complete all tasks and resolve all issues within the next 3 hours before they leave for their trip.*

Early in the assessment center the participant is told they will be a part of an upcoming event where their client is the keynote speaker. Later that day proprietary information is leaked, deeply upsetting this same client and causing them to withdrawal as the keynote speaker at the upcoming event. The client's email shows that they are very upset with this leak and may seek legal action against the company.

You will first be asked to link the provided list of behaviors to the list of competencies. Then you will be providing effectiveness ratings (i.e., -5 Very Ineffective to +5 Very Effective) on those same behavioral responses that a participant might take in response to this exercise prompt. Please take a few minutes to consider what successful performance would look like for this exercise.

When you are ready, please click “Continue” to proceed with the survey.

\_\_\_\_\_ Continue

## 5b. Task 1 - Linking Competencies

Below are names and brief definitions for the 6 competencies used in development of the assessment center exercises.

As part of Task 1, you are asked to read each behavior and decide which competencies each behavior links to. Please link each behavior to at least one competency. If a competency doesn't link to a behavior, please leave the space blank.

Please link a competency by typing in a number that you believe best represents the relationship between the behavior and the competency (i.e., 1 strong and clear link, 2 somewhat link that may be situation dependent, or 3 weak or vague link but still worth noting).

Make sure to base your judgment on the definitions provided here, not just the name of the competency.

### **Competencies:**

**Mission and Values Leadership:** Leads commitment to district mission and values that motivate and engage the team and others.

**Strategy and Results:** Demonstrates broad thinking and decision-making focused on district goals and succeeds in accomplishing these goals.

**Talent:** Seeks opportunities for self, team, and overall district to learn and develop.

**Relationships:** Builds strong relationships with key internal and external customers and partners.

**Communication:** Influences others through effective communication about district priorities.

**Change:** Supports change initiatives within the district by seeking opportunities for continual improvement.

1 - strong and clear link

2 - somewhat link that may be situation dependent

3 - weak or vague link but still worth noting



	<b>MVL</b>	<b>S&amp;R</b>	<b>T</b>	<b>R</b>	<b>Com</b>
1. Provided greeting (“Hi,” “Good Morning,” “Dear” and NOT just name “Dr. Willis.”)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2. Introduced self (“My name is ____, I am the Director...”)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3. Expressed appreciation or thankfulness (NOT just “thanks” or “thank you”)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4. Provided salutation at the end of the email (“Sincerely”, “Regards”, “Thanks”, “Thank you”)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
5. Provided name at the end of the email	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
6. Provided position and/or company information at the end of the email (“Director of ...”)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
7. Used informal or unprofessional language / abbreviations (e.g., “thru,” “ASAP”)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
8. Expressed satisfaction, optimism, or personal dedication to something	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
9. Expressed dissatisfaction, regret, disappointment, or apologized	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
10. Expressed empathy, considered another’s feelings or related own feelings to others	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
11. Agreed with or showed understanding for a claim, statement, concern, or opinion	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
12. Asked another to reconsider or wait on a decision	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
13. Asked another for a meeting / phone call / or discussion	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

	MVL	S&R	T	R	Com
14. Asked another to provide information, verify information, give advice/direction/instruction or made an inquiry	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
15. Asked another to respond, expressed intent to follow-up, or expectation/hope of future contact ("Please email me later today" or "I look forward future conversations")	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
16. Asked another to take action / delegated actions or express that help or assistance is needed	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
17. Explained the details of the assistance that is needed	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
18. Explained the current situation (Gave a summary, an overview explanation, or very general information as an update)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
19. Provided or offered information or explanations to maintain communication (e.g., "I was told by Alex that..." "I will let you know of any updates,")	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
20. Explained completed actions (e.g., contact with another, info to another, info from another)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
21. Explained current actions being taken by self or others (e.g., things being currently acted upon, investigated, or looked at, scheduled, or planned)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
22. Explained future actions – next steps (e.g., needed actions, ideas, not yet but should be arranged, organized or set into action)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
23. Sent someone information merely as a point of interest or to reference (e.g., forwarded an email as an FYI, "See below")	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
24. Communicated personal time constraints or implied a sense of urgency (e.g., "right away", "later today," "before I leave today," "before our conference call")	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
25. Communicated current schedule / timeline / roadmap	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
26. Implied or stated another had negative intentions, acted inappropriately or provided incomplete, false, or inappropriate information	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

	MVL	S&R	T	R	Com
27. Implied or stated false, misleading, or deceitful information	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
28. Showed praise (for a company, project, program, product, or work completed)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
29. Discussed value or importance (for a company, project, program, product, or work completed)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
30. Discussed the investment, commitment, time, or effort that has already been put in (for a company, project, program, product, or work completed)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
31. Discussed the potential future success or the trust, faith or confidence in it, dedication to it (for a company, project, program, product, or work completed)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
32. Discussed the goal, purpose, mission or intention of it (for a company, project, program, product, or work completed)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
33. Showed praise (for a person, relationship, or collaboration)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
34. Discussed value or importance (for a person, relationship, or collaboration)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
35. Discussed the investment, commitment, time, or effort that has already been put in (for a person, relationship, or collaboration)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
36. Discussed potential future success or the trust, faith or confidence in another, dedication to it (for a person, relationship, or collaboration)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
37. Had an issue with the spelling or grammar used in the email	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
38. Addressed or reference someone using the wrong, incomplete, or misspelled name/title (e.g., "Mr. Willis" instead of "Dr. Willis" or "Mary Jane" instead of "Mary Jane Starkey")	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

### 6b. Task 2 - Effectiveness Ratings

When considering effective leadership behavior, how necessary is it for the manager to send an email in response to this scenario as opposed to ignoring the email entirely? (Please indicate your response on the dial below.)

Not Necessary at all  Very Necessary  
0    1    2    3    4    5    6    7    8    9    10

Please consider and rate the following behaviors as though they were in a written response to the scenario. Please provide a rating on a scale of -5 to 5.

- 5 Ineffective: this action would be worse than doing nothing
- 0 Neutral: this action would have no meaningful impact
- 5 Effective: this action would be better than doing nothing

	-5	-4	-3	-2	-1	0	1	2	3	4	5
1. Provided greeting ("Hi," "Good Morning," "Dear" and NOT just name "Dr. Willis,")	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2. Introduced self ("My name is ____, I am the Director...")	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3. Expressed appreciation or thankfulness (NOT just "thanks" or "thank you")	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4. Provided salutation at the end of the email ("Sincerely", "Regards", "Thanks", "Thank you")	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
5. Provided name at the end of the email	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
6. Provided position and/or company information at the end of the email ("Director of ...")	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
7. Used informal or unprofessional language / abbreviations (e.g., "thru," "ASAP")	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
8. Expressed satisfaction, optimism, or personal dedication to something	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
9. Expressed dissatisfaction, regret, disappointment, or apologized	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
1. Expressed empathy, considered another's feelings or related own feelings to others	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
11. Agreed with or showed understanding for a claim, statement, concern, or opinion	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
12. Asked another to reconsider or wait on a decision 13. Asked another for a meeting / phone call / or discussion	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
13. Asked another for a meeting / phone call / or discussion	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Please consider and rate the following behaviors as though they were in a written response to the scenario. Please provide a rating on a scale of -5 to 5.

- 5 Ineffective: this action would be worse than doing nothing
- 0 Neutral: this action would have no meaningful impact
- 5 Effective: this action would be better than doing nothing

	-5	-4	-3	-2	-1	0	1	2	3	4	5
14. Asked another to provide information, verify information, give advice/direction/instruction or made an inquiry	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
15. Asked another to respond, expressed intent to follow-up, or expectation/hope of future contact ("Please email me later today" or "I look forward future conversations")	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
16. Asked another to take action / delegated actions or express that help or assistance is needed	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
17. Explained the details of the assistance that is needed	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
18. Explained the current situation (Gave a summary, an overview explanation, or very general information as an update)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
19. Provided or offered information or explanations to maintain communication (e.g., "I was told by Alex that...", "I will let you know of any updates,")	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2. Explained completed actions (e.g., contact with another, info to another, info from another)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
21. Explained current actions being taken by self or others (e.g., things being currently acted upon, investigated, or looked at, scheduled, or planned)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
22. Explained future actions – next steps (e.g., needed actions, ideas, not yet but should be arranged, organized or set into action)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
23. Sent someone information merely as a point of interest or to reference (e.g., forwarded an email as an FYI, "See below")	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
24. Communicated personal time constraints or implied a sense of urgency (e.g., "right away," "later today," "before I leave today," "before our conference call")	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
25. Communicated current schedule / timeline / roadmap	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
26. Implied or stated another had negative intentions, acted inappropriately or provided incomplete, false, or inappropriate information	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Please consider and rate the following behaviors as though they were in a written response to the scenario. Please provide a rating on a scale of -5 to 5.

-5 Ineffective: this action would be worse than doing nothing

0 Neutral: this action would have no meaningful impact

5 Effective: this action would be better than doing nothing

	-5	-4	-3	-2	-1	0	1	2	3	4	5
27. Implied or stated false, misleading, or deceitful information	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
28. Showed praise (for a company, project, program, product, or work completed)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
29. Discussed value or importance (for a company, project, program, product, or work completed)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3. Discussed the investment, commitment, time, or effort that has already been put in (for a company, project, program, product, or work completed)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
31. Discussed the potential future success or the trust, faith or confidence in it, dedication to it (for a company, project, program, product, or work completed)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
32. Discussed the goal, purpose, mission or intention of it (for a company, project, program, product, or work completed)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
33. Showed praise (for a person, relationship, or collaboration)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
34. Discussed value or importance (for a person, relationship, or collaboration)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
35. Discussed the investment, commitment, time, or effort that has already been put in (for a person, relationship, or collaboration)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
36. Discussed potential future success or the trust, faith or confidence in another, dedication to it (for a person, relationship, or collaboration)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
37. Had an issue with the spelling or grammar used in the email	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
38. Addressed or reference someone using the wrong, incomplete, or misspelled name/title (e.g., "Mr. Willis" instead of "Dr. Willis" or "Mary Jane" instead of "Mary Jane Starkey")	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

## **7b. End of Survey**

Thank you so much for your help! As a thank-you for your time, we would like to offer you a \$5 Starbucks gift card (please note that these are only valid for Starbucks locations in the United States). In order to deliver your gift card, we need a valid email address for you. To preserve the anonymity of your survey responses, we will use a separate survey to collect your email address.

The two surveys are not linked in any way, and we will be unable to connect your email address with your survey responses.

To enter your email address, please click here:

<https://www.surveymonkey.com/s/8JVXBTR>