2015

Assessment Center Structure and Construct Validity: A New Hope

Christopher Wiese
University of Central Florida

ASSESSMENT CENTER STRUCTURE AND CONSTRUCT VALIDITY: A NEW HOPE

by

CHRISTOPHER W. WIESE
B.S., University of Central Florida, 2008

A dissertation submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy
in the Department of Psychology
in the College of Sciences
at the University of Central Florida
Orlando, Florida

Summer Term
2015

Major Professor: Kimberly Smith-Jentsch

ABSTRACT

Assessment Centers (ACs) are a fantastic method to measure behavioral indicators of job performance in multiple diverse scenarios. Based upon a thorough job analysis, ACs have traditionally demonstrated very strong content and criterion-related validity. However, researchers have been puzzled for over three decades with the lack of evidence concerning construct validity. ACs are designed to measure critical job dimensions throughout multiple situational exercises. However, research has consistently revealed that different behavioral ratings within these scenarios are more strongly related to one another (exercise effects) than the same dimension rating across scenarios (dimension effects). That is, results from ACs suggest that we are unsure of what these behavioral measures represent.

Over the last three decades, researchers have sought to illuminate why same dimension ratings are inconsistent across scenarios. However, these investigations have been limited to changes influencing the source of the ratings (e.g., assessors, trained raters). No approach has been taken to change the structure of the AC. This study breaks with tradition and introduces a structurally different AC: A Day-In-The Life AC (DITLAC). A DITLAC structure is designed to mimic that of a normal day on the job. In the present study, the construct validity between a DITLAC and a traditionally structured AC is compared with the argument that the DITLAC will demonstrate stronger construct validity evidence. In several cases, this was found to be true.

This dissertation is dedicated to my family, both immediate and extended. My extended family (Julia, Andrew, Ed, Jessica, Marissa, Billy) has supported me along the way and have always been there when I needed guidance or constructive criticism. My immediate family (Nancy, William, Aimée, Robert) helped build the base of who I am today. I could not be where I am without their help and support. Lastly, I need to thank two catalyst that changed my worldview in many different ways: My Wife, Christine and My Son, Liam. By completely changing my perspective on life, they have inspired me to become a better person and reach further than I ever thought I could.

I love you all.

ACKNOWLEDGMENTS

I would like to acknowledge the work of all of my research assistants (all 30 of you) that helped me gather, code and analyze these data. Throughout all the technical goofs and odd participants, you were all steadfast in your efforts. I sincerely appreciate that.

I would also like to acknowledge a great deal of the IO program here at UCF for their help in discussing my dissertation and for putting up with the term DITLAC.

Lastly, I would like to acknowledge and thank my committee members for their invaluable insight and feedback on this project.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1: INTRODUCTION

Statement of Problem

In the realm of selection, Assessment Centers (ACs) are known to be one of the most prolific methodologies. Assessment Centers are characterized by the assessment of several different dimensions of a participant's behavior in multiple situational exercises as rated by several different individuals (International Task Force on Assessment Center Guidelines, 2009). Though ACs commonly produce both content and criterion-related validity evidence (e.g., Arthur, Day, McNelly, & Edens, 2003; Gaugler, Rosenthal, Thornton & Bentson, 1987; Sackett, 1987), proof of construct validity has eluded researchers. In fact, the construct validity of ACs has been a point of debate amongst researchers and practitioners alike for over three decades. Sackett and Dreher (1982) shed light on the construct validity problem by demonstrating that dimension ratings gleaned from ACs do not necessarily measure the dimensions they purport to assess. That is, they found that individuals behaved more consistently within a situation than across situations. This evidence goes against the original purpose of ACs, which was to assess job relevant dimensions in multiple exercises. If measures of these dimensions were construct valid than one would expect all measures of a dimension to be strongly related to one another across situations. Contrastingly, Sackett and Dreher (1982) demonstrated that these relationships were not strong and some neared zero. Furthermore, their results revealed that behaviors were more likely to be consistent within situations. That is, behavioral ratings on conceptually different dimensions were strongly related to each other within a given exercise. Their article spurred a stream of research investigating the newly dubbed 'construct validity problem' of ACs.

Often, researchers attributed this problem to measurement error. Simply put, they posited that the assessors making the behavioral ratings were to blame. Thus, most of the literature

investigating the construct validity problem focused on 'corrections' to the assessors. This took the form of using different types of assessor training, reducing the number of dimensions per assessor, providing behavioral checklists, among others (e.g., Donahue, Truxillo, Cornwell, & Gerrity, 1997; Gaugler & Thornton, 1989; Hennessy, Mabey, & Warr, 1998; Lievens, 2001; Reilly, Henry, & Smither, 1990; Schleicher, Day, Mayes, & Riggio, 2002; Schneider & Schmitt, 1992). Though the results of these 'design fixes' did demonstrate improvements in construct validity evidence (e.g., Arthur, Woehr, & Maldegen, 2000), relationships between different dimensions within exercises were still stronger than relationships between ratings of the same dimensions across exercises. Recently, there has been a call for the abandonment of the proposed 'design fixes' (Lance, 2008). Those in favor of abandoning this line of research typically take the opinion of either (a) dimension ratings need to be dropped all together and the focus should be on performance within the exercises or (b) variance in AC ratings can be attributed to both dimension and exercises as well as a general ability factor – though exercise variance will still likely explain the most.

These approaches, however, are still relatively new. As such, it can be asserted that after over 30 years of investigation there is still no clear consensus on what ACs are measuring. This is troublesome for a multitude of reasons. First, ACs are notoriously expensive to create and develop (Bray & Grant, 1966; Hinrichs, 1978). Even though research has often demonstrated the criterion-related validity, it may become increasingly harder to convince organizations to adopt ACs if it cannot be specified as to what is actually being measured. Another issue lies in the use of developmental ACs. Developmental ACs evaluate employees current level on specific dimensions in order to provide feedback that will improve behaviors associated with those specific dimensions. If an AC is unable to consistently measure a dimension across situations,

there is little point creating developmental ACs as there will be no clear way to provide feedback on a specific dimension. In light of this, the current study argues that the abandonment of the 'design fixes' was premature and proposes a novel way to fixing the construct validity problem of ACs.

As mentioned previously, the common focus of researchers seeking to investigate 'design fixes' was the assessor. However, there is one rather large assumption researchers are making by solely conceptualizing this issue as simply 'measurement error.' Compliance with this assumption infers that the behaviors being assessed in the AC are representative of those being performed on the job. This is not an illogical assumption to make. ACs are the result of a thorough job analysis in which several behavioral dimensions are identified and exercises are created in order to assess these dimensions. Given the high fidelity nature of this approach and the consistent evidence of content and criterion validity, it is easy to assume that AC behaviors match job behaviors. However, I argue that this is not the case.

A relatively uncontested proposition in psychology is that behavior is both a function of the person and the situation (Cooper & Withey, 2009;Endler & Parker, 1992). In fact, research has demonstrated that even the smallest changes within one's environment could create drastic differences in behavior (Cooper & Withey, 2009). Thus, using this logic, it can be argued that the situation of an AC should mirror that of a common day on the job. The high fidelity nature of Traditional Assessment Centers (TAC) is often attributed to the exercises and dimensions being assessed. That is, the tasks that one engages in during a TAC often represent the same tasks as one would perform on the job. It follows then that the behaviors one expresses within these tasks would be representative of the dimensions identified in the job analysis. It is easy to see then why ACs often have strong face validity (Chan & Schmitt, 1997; Howard, 1974; Klimoski &

Brickner, 1987; Macan, Avedon, Paese, & Smith, 1994). However, there are other situational characteristics of a TAC that make them unlike a typical day on the job.

A typical experience of an assessee going through a TAC may involve the following: Arrive at the location, fill out paperwork, receive instructions for an exercise, complete that independent exercise, be moved to a different location, start training for the next exercise, complete the next independent exercise, and so on. Each exercise is commonly seen as its own distinct situation (Neidig & Neidig, 1984), most likely due to the lack of a consistent context throughout the exercises. The structure in which TACs are organized does not represent the same structure on the job. Employees (a) typically do not receive instructions immediately prior to completing their work tasks, (b) are interrupted during the performance of one task by a conflicting responsibility, (c) do not have clear breaks after the completion of a task, and (d) work with the same core group of individuals throughout the day. Thus, it should hold that ACs designed to mimic this structure would better assess behaviors typically performed on the job. I propose to demonstrate this by constructing a day-in-the-life assessment center (DITLAC) and argue that the construct validity evidence will be stronger in DITLACs compared to that of TACs.

Purpose of Current Study

The purpose of the current study is to demonstrate the construct validity superiority of DITLACs over that of TACs. I argue that by manipulating four structural components (exercise integration, context consistency, instructional characteristics, and breaks) it is possible to create an AC that is more representative of the situations experienced on the job than a TAC. Specifically, I use situational strength theory (Mischel, 1973;Mischel, 1977) to argue that the

situations are too strong in TACs to allow for person factors (i.e., personality) to influence behaviors in AC exercises.

Situational strength theory argues that there are cues within a situation that contain information on how one should behave (Chatman, 1989; Mischel, 1977; Meyer et al., 2010; Smithikrai, 2008). When a situation is strong, these situational cues are transparent and easily interpreted, thus reducing behavioral variance between individuals. However when the situation is weak, the situational cues are more ambiguous and less clear, leading to different interpretations of the situation and, consequently, increased behavioral variance between individuals. Moreover, it is argued that DITLACs present a relatively weaker situation than TACs, therefore allowing person factors (i.e., personality) to have a greater degree of influence on behavior.

In respect to the construct validity, the results using this design will improve construct validity evidence in two ways. First, as dimension ratings taken from DITLAC will be influenced by the same determinant (i.e., personality) across exercises, same dimension correlations across exercises should be stronger in DITLAC than in TAC (evidence for convergent validity). Additionally, this also implies that the relationship between different dimension ratings within each exercise will be weaker in DITLAC than in TAC (evidence for discriminant validity). Second, using a nomological network approach to construct validity, results should demonstrate that theoretically similar measures should be more strongly related to one another in the DITLAC than in the TAC and that theoretically unrelated measures will be more strongly related to one another in the TAC than in the DITLAC. Thus, these approaches should demonstrate the construct validity superiority of DITLACs over TACs in two different ways.

This paper will unfold as follows. First, I will provide an overview of the 'construct validity' problem in ACs. Second, I will introduce the four structural components of ACs I intend to manipulate and elaborate on how variations on these components may affect an assessee's perceptions of the situation. Also, I will specify how differences in these structural components results in two different types of ACs: TACs and DITLACs. Next, I will present an integrated review of situational behavior and tie how the four aforementioned structural components affect the situational strength of the AC. Following this, I will lay out two sets of hypotheses that propose to demonstrate the construct validity superiority of DITLACs over TACs. This is followed by a detailed explanation of the methodology surrounding the current study.

CHAPTER 2: LITERATURE REVIEW

The Construct Validity Problem

There was a span of time from the early 1970s to the 1980s where ACs were treasured among researchers and practitioners alike. By 1973, Douglas Bray and William Byham had partnered up to create Development Dimensions, Inc. (DDI) to produce situational exercises that could be purchased by organizations – and they were. It was reported that hundreds of organizations were running some form of AC by the early 1970s (Byham, 1977; Highhouse & Nolan, 2012). However, the proverbial party ended with the publication of Sackett and Dreher's (1982) seminal investigation into the construct validity issues rampant in ACs. Their analysis was the catalyst that prompted almost 30 years of research probing the construct validity of ACs. In the following subsection, there will be a brief review of the three types of validity evidence demonstrated by researchers in ACs up until 1982, a dissection of the 'Construct Validity Problem', the strides researchers have taken to correct for this problem, and an analysis of the current state of AC construct validity research.

Assessment Center Construct Validity Procedures

In order to fully understand AC's 'Construct Validity Problem,' one must first recognize the three different types of validity typically analyzed in the ACs: Criterion, content, and construct validity. First, criterion-related validity is simply concerned with the degree to which there is an empirical relationship between a measure and a criterion an experimenter wishes to measure (American Psychological Association, 1954; Cronbach & Meehl, 1955). Criterion validity is typically segmented into two different types of validity depending on if the predictor and criterion are measured at relatively the same time (i.e., concurrent validity) or at different times (i.e., predictive validity). The criterion-related validity studies in the 1960s and 1970s were

primarily concerned with predictive validity (e.g., Anstey, 1966; Anstey, 1971; Bray & Campbell, 1968; Bray & Grant, 1966; Campbell & Bray, 1967; Himrichs, 1978; Kraut & Scott, 1972; Moses & Wall, 1975). For instance, Bray and Campbell (1968) demonstrated that ratings gleaned from an AC strongly related to the performance of salesmen after a few months of being on the job. Additionally, Hinrichs (1978) demonstrated that assessment ratings still predicted the job performance of managers in a manufacturing organization after eight years. Thus, one of the factors contributing to the popularity of using AC methodology was the strong criterion-related validity evidence.

Next, content validity refers to the degree to which the content of what's being measured adequately samples from the domain of situations currently being inspected (Cronbach, 1971). In other words, content validity concerns whether the test captures all aspects of the construct an experimenter intends it to measure. As the 'exercise' can be conceptualized as the measure of ACs, content validity in ACs concerns the degree to which the tasks performed in ACs represent those on the job. Content validity evidence was established through a thorough job analysis where important performance constructs are identified (Dreher & Sackett, 1981; Sackett & Dreher, 1982). Once these constructs were identified, organizations would either develop or purchase exercises specifically designed to assess them. Some argued that further validation was unnecessary and ACs could be implemented after the content-validation process (Norton, 1977). That is, at that time, the construct validity of ACs was simply assumed to be established through the content-validity process. Intuitively, this argument contains a degree of rationality as it has also been argued that content validity evidence is simply one type of construct validity evidence. In fact, paired with criterion-related validity evidence, many argued that ACs were valid based

on this evidence alone (Byham, 1980a; Byham, 1980b; Jaffee & Sefcik, 1980). However, Sackett and Dreher (1982) argued that the current validation process simply was not enough.

Lastly, construct validity concerns the degree to which a performance on a test represents the psychological qualities of the construct it asserts to assess (American Psychological Association, 1954; Cronbach, 1971; Cronbach & Meehl, 1955). That is, it answers the question: Does the test measure what it is supposed to measure? Thus, construct validity is concerned with the meaning of test scores. Up until 1982, there had been very few construct validations efforts as some believed it was not necessary. However, Sackett and Dreher (1982) provided a plethora of evidence contrasting this belief.

Sackett and Dreher (1982) argued that the content-validation strategies used to justify the validity of AC methods were simply deficient and that additional validation evidence was needed. The additional validation process took form under Campbell and Fisk's (1959) multitrait-multimethod (MTMM) matrix. The MTMM matrix method provides two different types of construct validity evidence. Convergent validity evidence is determined by analyzing the degree to which scores on the same trait are related to one another across different methods. The greater the strength of this relationship the more evidence one has that they are measuring the same construct. Additionally, discriminant validity evidence is demonstrated when the relationship between different traits within the same method produce relatively low correlations. If the relationships between different traits within the same method are low, it suggests that one is measuring conceptually different constructs. Applied to ACs, the multiple traits are synonymous with AC dimensions and the methods are synonymous with AC exercises. This is best evidenced by reviewing Table 1.

Table 1. *Example of construct validity evidence within ACs.*

		Exercise 1			Exercise 2			Exercise 3		
		D1	D2	D3	D1	D2	D3	D1	D2	D3
Exercise 1	D1	-								
	D2	*DDSE*	-							
	D3	*DDSE*	*DDSE*	-						
Exercise 2	D1	**SDDE**	~~DDDE~~	~~DDDE~~	-					
	D2	*DDSE*	**SDDE**	~~DDDE~~	*DDSE*	-				
	D3	*DDSE*	*DDSE*	**SDDE**	*DDSE*	*DDSE*	-			
Exercise 3	D1	**SDDE**	~~DDDE~~	~~DDDE~~	**SDDE**	~~DDDE~~	~~DDDE~~	-		
	D2	*DDSE*	**SDDE**	~~DDDE~~	~~DDDE~~	**SDDE**	~~DDDE~~	*DDSE*	-	
	D3	*DDSE*	*DDSE*	**SDDE**	~~DDDE~~	~~DDDE~~	**SDDE**	*DDSE*	*DDSE*	-

Note. D* = Dimension; *DDSE* = Different Dimension Same Exercise; **SDDE** = Same Dimension Different Exercise; ~~DDDE~~ = Different Dimension Different Exercise.

Imagine Table 1 as a correlation matrix. Convergent validity is demonstrated when same-dimension different-exercise (SDDE) correlations are high. When SDDE correlations are strongly related to one another this supports the notion that the scores on the measures across different exercises represent the same construct. Discriminant validity evidence is observed when different-dimension same-exercise (DDSE) correlations are low. That is, when DDSE correlations demonstrate poor relationships with one another, it can be suggested that scores on each of the dimension measures represent different constructs. For instance, if one were to factor analyze ratings from a construct valid AC, ratings of the same dimension would group together across different exercises. This is the type of construct validity evidence expected of ACs, however, it is not what Sackett and Dreher (1982) found.

Keeping the MTMM matrix framework in mind, Sackett and Dreher (1982) factor analyzed AC results from three different organizations. Their results showed that ratings were grouping according to the exercise (an exercise effect), rather than the dimensions as one would

expect. This pattern of results suggested that individuals were performing more consistently within exercises than across exercises. In fact, in some cases same dimension correlations across different exercises neared zero. In other words, this evidence suggested that ACs were not measuring the dimensions that they purported to measure. If they were, one would expect to find results indicating that an individual's performance on a specific behavioral dimension within a given exercise would be strongly related to ratings of that same behavioral dimension in another exercise. These findings sparked a stream of research seeking to discover exactly what ACs were measuring in hopes to fix the 'construct validity problem.'

'Fixing' the Construct Validity Problem

In light of the results unveiled by Sackett and Dreher (1982), a good portion of the research on ACs would focus on fixing the construct validity problem over the course of the next three decades (Thornton & Gibbons, 2009). Specifically, much of this research asserted that changes to certain 'design characteristics' would yield better convergent and discriminant validity evidence (Lance, 2008). Early blame for the poor construct validity evidence fell upon the assessors (Turnage & Muchinskiy, 1982). As such, several 'design fixes' took the form of incorporating behavioral checklists to alleviate assessor cognitive strain (e.g., Donahue et al., 1997; Reilly et al., 1990), reducing the number of dimensions each assessor was responsible for rating (e.g., Gaugler & Thornton, 1989), using different types of assessors (e.g., Sagie & Magnezy, 1997), and providing assessors with longer and/or different training (e.g., Dugan, 1988). Though many of these 'design fixes' demonstrated improved construct validity evidence, exercise effects remained dominate in explaining the meaning behind scores (Lance, 2008).

Researchers were not blind to the lack of empirical findings. During the last decade, several authors have suggested that rather than conceptualize the exercise effects as artifacts of measurement error, this variance could represent meaningful performance differences between

exercises and might be related to significant performance factors on the job (Lance, 2008; Lance et al., 2000; Lievens, 2000). In fact, during the 2008 meeting of the Society for Industrial and Organizational Psychology (SIOP), some of the most prominent AC researchers (Brian Hoffman, Winfred Arthur, Charles Lance, Filip Lievens, Craig Russell, and David Woehr) suggested a complete moratorium on construct validity research using the MTMM model (Hoffman, 2008). Further, Lance (2008) placed the proverbial nail in the MTMM coffin in his commentary on the current state of AC validation by stating: "…[D]espite various design fixes… "construct validity" as conceived traditionally by transporting the MTMM methodological platform to the study of the structure of AC [ratings] probably cannot be salvaged or concocted at least on any kind of regular basis (p. 92)."

One criticism of the 'design' fixes, however, is that the majority do not specifically focus on the design or structure of the AC. Most of the manipulations involve changes to cognitive workload or training of the assessor (Lievens, Chasteen, Day, & Christiansen, 2006). This implies that the lack of construct validity evidence using the MTMM approach is a result of measurement error (Thornton & Gibbon, 2008). In turn, this suggests that the situations in which the participants are operating in are assumed to represent the same situations they would face on the job – therefore eliciting the same behaviors for assessors to rate. Thus, the research on the 'design' fixes did not focus on changing the situations or the participant's perceptions of the situation – with one exception.

Kleinmann (1993) argued that individuals differed in their ability to determine which dimensions were being assessed in AC exercises. If they recognized which dimensions they were being assessed on, they would perform more consistently on these dimensions across exercises. He found this to be the case. Participants who identified the same dimension across two exercises

performed more consistently than those who did not. In other words, those who identified the same dimension across exercises had more construct valid ratings using the MTMM approach. This finding served as the catalyst for the debate on whether to make targeted dimensions in ACs transparent or not. Though the results have been mixed (e.g., Kleinmann, Kuptsch, & Koller, 1996; Kolk, Born, & van der Flier, 2003; Smith-Jentsch, 2007; Wiese & Smith-Jentsch, 2009), these results represent the only meaningful attempt to change the assessee's perception of the situation within an AC in hopes of fixing the construct validity problem.

Current State of Assessment Center Research

The relative desertion of investigations in search of 'dimension effects' signaled the need for new streams of research to come about. Two streams of AC research have been emerging over the course of the past decade. The first embraces the lack of construct validity evidence using dimension ratings with full fervor: The task-based assessment centers (TBAC). Proponents of the TBAC argue for the removal of dimensions rating all together and suggest that AC research and design concentrate entirely on exercises. Simply put, TBAC use a rating methodology that focuses on general performance within exercises, not performance on specific dimensions. There is some validity to this approach as research has shown that ratings from TBAC demonstrate similar criterion validity evidence as dimension-based ACs (Jackson, 2007; Lance, Foster, Nemeth, Gentry, & Drollinger, 2007). To be noted, however, is that this approach does not change the design of the AC – simply the conceptualization of what constitutes performance.

The second stream of research is more inclusive than the TBAC approach. Supporters of the mixed-model assessment center (MMAC) approach believe that both dimension effects and exercise effects are meaningful components of ACs. The mixed-model approach does not necessarily focus on the design of the AC, rather how the data gleaned from ACs is modeled or

interpreted. Using a mixed-model approach allows for variance to be attributed to not only exercises, but to dimensions as well. However, the dimension factors are not as narrow as typical AC dimensions. For instance, Hoffman, Melchers, Blair, Kleinmann, and Ladd (2011) found that a model with exercise factors, two broad dimension factors, and one general performance factor best fit the AC data compared to alternative models tested.

There is a great deal of debate surrounding which of these new approaches will reveal the greatest understanding of ACs (see Jackson, Lance, & Hoffman, 2012 for a detailed review), however, these two approaches share something in common with the preceding investigations into AC construct validity. That is to say, these two new approaches do not attempt to change, manipulate, or investigate how changes to the structure of the AC could produce changes in participant behavior. The assumption is that ACs provide good behavioral data that represents the same or similar behaviors that participants will engage in on the job. In other words, these approaches assert that the problem is not a part of AC methodology, rather, our ineptitude in correctly analyzing or interpreting it.

In considering all of the research attempting to fix the construct validity problem, I argue that the call for the moratorium on research manipulating design features of ACs is shortsighted. Research has shown that AC design changes focused on enabling assessor to make better ratings are critically important (Lievens & Conway, 2001), however these design changes are clearly not enough to fix the problem (Lance et al., 2000; Lance, Foster, Gentry, & Thoresen, 2004; Lievens, 2001, 2002). Thus, if we can assume that when raters are (a) trained properly, (b) assigned the appropriate number of dimensions and participants, and (c) have clear instructions on how to rate each dimension that they are correctly and consistently interpreting a participant's behavior, poor construct validity evidence is clearly not a result of measurement error. Thus, a

logical next step would be to investigate the design of the ACs itself in hopes of producing more construct valid results. This is not to say that variance in scores attributed to exercises is 'meaningless' variance as some might argue. It is hard to suggest that such a large effect, which has been replicated numerous times, does not contain any valuable information. However, the potential implication of changing the design of ACs on construct validity is far too fruitful an avenue of research to desert so concretely. In order to fully appreciate the prospective value changes in structure may hold for the validity of ACs, it is first necessary to identify the different structures of ACs and their respective components.

Assessment Center Meta-Structure

The following details concerning AC design elements will herein be referred to using the term 'meta-structure' instead of 'design' in order to distinguish this line of design research from the construct validity 'design fixes' research. More specifically, the term meta-structure refers to changes concerning the AC as a whole: From when the assessee walks through the door until they complete the final exercise. Research investigating the meta-structure of ACs is not necessarily concerned with the AC's purpose (e.g., developmental ACs, selection ACs, diagnostic ACs), how it is being rated (e.g., 'design fixes'), what is being rated (e.g., TBAC), or how it is being analyzed or interpreted (e.g., MMAC). The meta-structure of an AC concerns the manner in which exercises within an AC are organized, the degree to which they are related to one another and the manner in which they are presented temporally. These factors directly affect the assessees' experience while in the AC but have been largely overlooked in AC research. In the following section, I elaborate upon the major components of AC meta-structure, and provide an analysis on how these structural features may impact assessee behavior.

Defining the meta-structure of an Assessment Center

There are no two AC designs that are exactly the same. Each AC differs on multiple aspects including the purpose of the AC, the number of dimensions assessed, the number of exercises, training of assessors, status of assessors, how ratings are made, how ratings are integrated, among others (Eurich, Krause, Cigularov, & Thornton, 2009; Spychalski, Quiñones, Gaugler, & Pohley, 1997). Another area where ACs vary is in their meta-structure. As I elaborate further on below, variations in the meta-structure of ACs can affect the assessee's perception of the situation – a major determinant of assessee behavior. Thus, research should be investigating components of the AC that impact or influence an individual's perception of the situation. These types of factors I call *structural components*.

It is important to specify that the structural components of an AC extend to factors occurring outside each exercise. Structural components include *any* feature of the AC design that impacts the participant's behavior within exercises from the moment that they arrive to the moment when the last exercise ends. When only considering the structural factors that influence an individual's behavior within an exercise, the result would be a deficient analysis of the situation which might lead to incorrect assumptions concerning the determinants of behavior. A careful distinction must be made here. Though structural components influence the assessee's perception of the situation, they are not the only part of the AC that contributes the assessee's viewpoint. Within the AC exercises, there are certain situational cues built into the content of the exercise that will affect an assessee's behavior. These content-related cues are not structural components. Structural components only concern factors within the design of the AC – Not necessarily the content itself. That is, the decisions made concerning structural components are mostly logistical in nature and are often not a part of the initial job analysis. With this in mind, I

propose that there are four different structural components researchers should consider: Exercise integration, context consistency, instructional characteristics, and breaks.

Exercise Integration. When constructing an AC, designers are recommended to include several distinct exercises so that a wide range of behaviors may be assessed (International Task Force on Assessment Center Guidelines, 2009). Considering that the use of AC methodology is applied to a wide range of jobs, it is rational that there is also a wide range of exercises – though some are more common than others. Some of the more popular types of exercises are the role play exercise, oral presentation exercise, and in-box exercise. A role play exercise is characterized by an assessee assuming the role of a position within the organization and subsequently interacting with role-players (often confederates) to create typical situations of that particular position. An oral presentation exercise can require the assessee to review a packet of information and then make a short oral presentation concerning their opinion on the packet of information provided to them. Lastly, an in-box exercise (formally called an in-basket exercise) typically entails an assessee reading and responding to a series of e-mails which someone in the targeted position would likely receive.

Exercise integration refers to how these different exercises are presented to the assessee. It is easiest to conceptualize exercise integration by describing the two extremes. If exercises are not at all integrated then the participant realizes that there is a clear beginning and end of the exercise. In other words, they know that they will be performing a new task shortly as the previous one has ended. If exercises are integrated, it is more difficult for the participant to make that determination. They are not aware that the random assortment of tasks they are performing is really an amalgamation of several different exercises.

More specifically, at one end of the continuum, the participants are specifically told (a) they will be performing several distinct situational exercises, (b) they are told when a certain exercise begins, and (c) they are told specifically when an exercise ends. At the other end of the continuum, the task requirements of each exercise is intertwined such that the assessee may have to choose between a task related to an in-basket exercise and a task related to a role-play exercise. In fact, when completing a task for one exercise, it is plausible that the assessee may be interrupted to perform a task from a different exercise. In other words, tasks from each exercise are spread out across the entire assessment period. For example, an assessee may have to answer an e-mail from a coworker (task from inbox exercise), but then is interrupted by an angry customer (task from role play exercise). Once they are finished with the angry customer, they notice they have a voicemail from their supervisor asking them to give a presentation in fifteen minutes (task from oral presentation exercise). At this point, they will have to decide whether to spend all of the fifteen minutes preparing for the presentation or decide to work on the e-mails for a portion of the time. When exercises are fully integrated, the tasks, responsibilities, and priorities of the assesse may become unclear to them – resulting in a greater degree of ambiguity. However, these are the extremes. It may be possible for distinct exercises to be completed in full, but the assessees are simply not told when one exercise ends or another begins. It also could be the case where participants are informed about what tasks belong to which exercise, but the actual performance of these tasks are not blocked off into distinct exercises. Nevertheless, more exercise integration often results in more ambiguity on how to behave. Thus, assessees will have to use their own judgment in deciding which behaviors will be more effective. As elaborated upon in the following section, I argue that this ambiguity will result in more typical behaviors.

Context Consistency. Simply put, context consistency refers to whether contextual job information across the entire AC remains the same. The degree of context consistency, however, may vary greatly. That is, context consistency exists on a continuum. For instance, if an AC has no context consistency each exercise is (a) composed of its own distinct narrative (e.g., each exercise is a different 'position'), (b) makes no mention of information from other exercises (e.g., working for different organizations for each exercise), and (c) does not require any information from previous exercises in order to complete the current exercise (e.g., interacting with a completely different set of 'coworkers'). In other words, there are no purposeful cues within an exercise or across exercises that are designed to trigger the recollection of any information from other exercises. However, the exact opposite is true for ACs designed to be context consistent. These ACs are designed such that information from one exercise may influence the way an individual behaves in another exercise. This can take the form of interacting with the same supervisor (digital or real life role player), having decisions made in one exercise impact the decisions made in subsequent exercises, or having the different exercises take place under the same narrative.

It is important to distinguish context consistency from exercise integration. Exercise integration is simply concerned with *how* the exercises are presented to the assessee, whereas context consistency focuses on *what* information is presented to the assessee. However, the logistical decisions made for one of these components will obviously affect the other. For example, when designing an AC that has decided to fully integrate the exercises, it would make more sense that there is, at least a degree of, context consistency. As this section progresses, it should become clear that decisions on one of the structural components influences the decisions made concerning other components.

To best illustrate how context consistency affects an assessee's perception of the situation the following example is provided. First, imagine an assessee going through multiple exercises. During a role-play exercise, an employee (confederate) is late for work and asks the assessee to lie to their supervisor so that they will not get in trouble. In most cases, an assessee will promptly refuse their request as it is obviously unethical. Later, in an inbox (or in-basket) exercise, the assessee needs to make a determination about whether or not to grant a different employee paid time off – a decision which is entirely up to the assessee with no specific guidance on how to make the decision. In this scenario, the encounter with the late employee should have no bearing on the decisions made with the second employee. That is, there is little context consistency. However, imagine that the first employee and the second employee are the same person. In this second scenario, the information concerning the employee's lateness may affect the assessee's decision on whether to allow them paid time off from work. Though this is only one example, it provides a picture on how context consistency may change the assessee's perception of the situation.

Instructional Characteristics. When making logistical decisions about the structure of an AC, designers must make decisions about the frequency, length, and content of instructions. First, instruction frequency refers to how often the assessee is exposed to instructions. For instance, instructions can take place immediately prior to each exercise or may only occur in one bulk instructional session at the beginning of the AC. Second, instruction length refers to how much time is spent getting instructions. Instructions can last for only 20 minutes or could be as long as half a day. Lastly, instruction content refers to what the assessee is actually learning in the instructional sessions. They can simply be learning how to use the technological mechanisms they need to utilize during the exercise, they could be told about the objectives they need to

accomplish, or they could be explicitly made aware of the dimensions they will be assessed on (i.e., skill transparency; Kleinnman, 1993). All three of these, of course, are not mutually exclusive. For instance, if the designers decide to provide a great deal of information to the assessee, the length of instructions will obviously increase if they wish for their instructions to be effective.

Furthermore, decisions on the previous structural components will influence the decisions concerning the instructional characteristics. One example would be if a designer wishes to have fully integrated exercises it would make no sense to provide instructions on different occasions, which would in turn affect the length of the instructions. Thus, the assessee would be exposed to one long bulk session of instructions. Additionally, if the designer wishes to have a consistent context throughout the AC, they may decide to provide instructional content concerning the values of a fictitious organization that pertains to all exercises.

There are several ways that choices concerning the instructional characteristics may impact the assessee's perception of the situation. If information concerning certain organizational norm or rules is contained within instructional sessions it may guide an assessee's behavior in some situations. That is, a situation where it is unclear how to behave might be less ambiguous if the instructional session contains information about the rules and regulations of the organization. Furthermore, if instructions take place immediately before each exercise, an assessee might be better able to recall and, thus, adhere to the information presented in the exercise. In other words, the information that may influence the assesee's perception of the situation will be in the forefronts of their minds.

Breaks. Much like instructions, decisions concerning breaks will involve the frequency and length of breaks, but also whether breaks will be transparent or natural. Transparent breaks

are when the assessees are clearly told that they are on a resting period. An example of a transparent break would be if they explicitly told the assessees they will take their lunch break from 12:00 PM to 1:00 PM. Conversely, an example of a natural break would be if the assessee is not explicitly told there will be a resting period. Natural breaks are self selected and not determined by any concrete rules or guidelines. Like other structural components, breaks are not excluded from the impact of other decisions. One example of this would be that if exercises are not integrated, breaks could take the form of the transition between different exercises.

In the same vein as exercise integration, breaks will influence an assesse's perspective of the situation through segmentation. The simple inclusion of transparent breaks could allow the assessee to recover cognitive resources. Thus, if they were presented with the same dilemma as an assessee that did not have any breaks, they may view the situation differently and, consequently, make different decisions. An individual's behavior will partially depend on the information they perceive (Galton, 1883;1965). If an assessee has more cognitive resources to use during an exercise, they may perceive more or different information, resulting in different behavior.

The case can clearly be made that these structural components affect an assessee's perception of the situation (and subsequent behavior) within ACs. Thus, the question begs: Why is there no research on this topic? The answer is relatively simple – It is impractical to conduct research that manipulates these components outside of the laboratory. ACs are notoriously expensive (Bray & Grant, 1966; Hinrichs, 1978). There are costs associated with conducting the job analysis, hiring and training experimenters (e.g., assessors, role players), and designing and implementing the AC. The manipulation of one of these characteristics would, in essence, create an entirely different AC. Even when the creation of two different ACs is plausible, the likelihood

of an organization wishing to take a risk on research of this nature is inarguably low. If a certain type of AC is known to produce consistent criterion validity results, there is no logical incentive for an organization to take part in an experimental investigation. Thus, research of this kind should be conducted on a smaller scale within a laboratory. Through experimental manipulation in a controlled setting, it is possible to maximize the manipulation of these characteristics in order to compare results to how ACs are typically structured. Additionally, even if there were several studies investigating the potential implications of the structural components, these studies would likely vary on multiple factors (e.g., type of participant, operationalization of dimension) which could impact the causal assertions concerning the structural components. Conducting these studies in a laboratory environment permits the research to control for such factors. Laboratory research of this nature can then be used to inform real-world assessment centers, which is exactly what I propose to do.

Traditional vs. Day-In-The-Life Assessment Centers

As these structural components are rarely described in method sections, it is difficult to argue for or against any type of typical AC structure. However, some inferences can be made through reviewing the methods sections of experimental AC literature and guidelines for developing ACs. For instance, several assumptions can be made if the AC uses Post-Exercise Dimension Ratings (PEDRs). Raters make PEDRs following the conclusion of an exercise; however, these ratings take time. Raters need to review their handwritten notes, behavioral checklists, or other behavioral cataloging approaches and make a determination of the score each assessee should get on each dimension. Two things can be inferred from this. First, when PEDRs are used, there are likely separate exercises. Second, as assessors need time to do these ratings, there is at least a small cognitive break for assessees between exercises. Though not all ACs use PEDRs, it is a very common practice (Thornton & Rupp, 2012). It should be noted, however,

that more contemporary ACs are recording these ratings and sending them off-site for dimension ratings. Thus, these two inferences cannot be made for more recent ACs. Furthermore, survey research suggests that only some ACs use partially integrated exercises and even fewer fully integrate their exercises (Eurich, Lraise, Cigularov, & Thornton, 2009). When exercises are moderately or barely related to one another, assessees are likely to receive training immediately prior to each exercise. It is also likely that there is little context consistency between exercises. Additionally, in several texts focusing on the development and design of ACs, authors suggest that assessees are provided with at least some information between each exercise (Thornton, 1992; Thornton & Byham, 1982; Thornton & Rupp, 2003). Therefore, a Traditional Assessment Center (TAC) design might be characterized by a lack of exercise integration, judicious context consistency, instructions immediately prior to exercises, and breaks in between exercises.

Conversely, a Day-In-The-Life Assessment Center (DITLAC) design is almost the exact opposite in terms of structural components. DITLAC designs are meant to imitate a typical day on the job for the assessee (Eurich et al., 2009). It makes sense then that their exercises are fully integrated (Eurich et al., 2009; Thornton & Rupp, 2009). Additionally, though I differentiate between *how* information is presented to the assessee (exercise integration) and *what* information is presented to the assessee (context consistency), other researchers do not necessarily make this distinction. For instance, the term 'integrated day-in-the-life assessment center' has been used to described AC in which participants performed several clearly distinct exercises, however, the context was consistent across these exercises (e.g., Hoover, Giamatista, Sorenson, & Bommer, 2010; Rode et al., 2007). Furthermore, DITLAC designs can contain a massed instructional session where the assessee is provided with background information that is pertinent to all exercises (Thornton & Rupp, 2009). Thus, rather than receiving multiple instructional sessions,

assessees in DITLAC designs receive one instructional session for all exercises or tasks. Finally, though there is little clear text in the literature, it can be assumed that if exercises are fully integrated to where there is not a definable beginning or end then there are unlikely to be traditional transparent breaks immediately following the completion of a series of tasks.

A review of the extant literature on ACs reveals that DITLACs are seldom used in experimental studies. In fact, most of the research on DITLACs uses one particular AC: The Iliad Assessment Center. The Iliad AC is targeted to assess MBA students on a variety of dimensions (e.g., leadership, communication, and decision-making) in multiple exercises (e.g., in-basket, team meeting, and individual speech). The AC is presented in what the authors describe as a 'day-in-the-life' format, where the assessee assumes the role of a manager and participates in multiple exercises. Though these studies were not specifically concerned with the validity of DITLAC, results indicated that life satisfaction predicts AC performance, emotional intelligence explains unique variance in public speaking effectiveness, and emotion recognition predicts AC performance over general mental ability and conscientiousness (Bommer, Pesta, & Storrud-Barnes, 2011; Rode, et al., 2005; Rode et al., 2007). However, the term 'day-in-the-life' is used liberally in these studies. Some aspects of the Iliad AC do mirror the definition of DITLAC provided above. For instance, in Rode et al., (2005), assessees were given information pertaining to the leaderless group discussion exercise and persuasive speech exercise within the in-basket exercise, which implies there is a degree of context consistency. However, some features of the Iliad AC detract from the concept of representing a typical day on the job. For example, assessees are explicitly told that they will be participating in multiple exercises, which implies that there is little exercise integration. Further, the goal of the research was not to

compare the validity of DITLACs to that of a traditional AC, something which the present study addresses.

As alluded to earlier, compared to TAC designs, very few studies utilize DITLAC designs (e.g., Hoover et al., 2010; Bommer, et al., 2010; Rode et al., 2005; Rode et al., 2007). This is surprising given the assertions that DITLAC designs contain strong predictive and face validity (Development Dimensions International, 2009; Thornton & Mueller-Hanson, 2004). Moreover, ACs were originally conceived to assess typical, every day performance. Thus, the high fidelity nature of the situations inherent in DITLAC designs serves as a fruitful platform to assess typical behavior. However, this implies that behaviors expressed in TAC designs are less reflective of typical work behaviors and, thus, could be considered a less valid indicator of typical performance. In other words, the specific situational factors that result from the differences in the structural components within DITLAC and TAC designs may be strong enough to change the assessee's behavior. This could have implications concerning the construct validity of the ratings gleaned from either design. The specific rationale and expected patterns from these two designs are elucidated below.

An Integrated Model of Situational Behavior

Several of the early ACs were designed to assess typical performance. Specifically, they were designed to be highly representative of a common day on the job. However, some have argued that ACs do not necessarily measure typical performance, but, instead, measure maximum performance (e.g., Highhouse & Harris, 1993; Ployhart, Lim, & Chan, 2001). In maximum performance situations, an individual's behavior is said to be a function of their *ability*, whereas *personality* and *motivation* are better determinants of behavior in typical performance situations (Sackett, 2007). Commonly, typical performance is represented by what

an individual 'will do' and maximum performance is characterized by what a person 'can do' (Cronbach, 1960; DuBois, Sackett, Zedeck, & Fogli, 1993; Klehe & Latham , 2006), however, this is a simplified conceptualization. Specifically, maximum and typical performance are both representations of what an individual 'will do,' though they are measured under different situational circumstances. Maximum performance is assessed in the presence of strong situational cues (e.g., presence of observers, instructions to perform their best, and short performance periods), whereas typical performance is assessed in the absence of these cues (Beus & Whitman, 2012; Sackett, Zedeck, & Fogli, 1988). In general, the design and structure of ACs produce situations representative of maximum performance (e.g., Highhouse & Harris, 1993; Ployhart, et al., 2001). However, the three aforementioned situational cues are not the only aspects of the situation that affect behavior.

In the following section, I will explain how the structural components of an AC can be manipulated in order to create situations that are more representative of the typical workplace. To do this, one must first understand the determinants of human behavior. Specifically, the following section uses the interactionist perspective of behavior (Lewin, 1943) and situational strength theory (Mischel, 1973, Mischel, 1977) to illuminate in what circumstances typical performance is best assessed. Next, I will explain how each of the four structural components changes the situation to create better opportunities to assess behaviors in an environment more characteristic of a normal day on the job.

Interactionist Perspective of Behavior

The idea that both the situation and person play a role in the expression of behavior has many origins, thus, making it difficult to isolate a definitive starting point. Sir Francis Galton inferred that our behavior was determined in part by what our senses could detect, (Galton, 1883/1965) suggesting the environment affected our behavior to the degree to which we could

touch, see, smell, taste, and hear. However, the general idea that behavior is a function of both the situation and personality is largely credited to Kurt Lewin (Cooper & Withey, 2009; Endler & Parker, 1992; Weiss & Adler, 1984). Specifically, Lewin created a simple heuristic formula ($B = f(P,E)$ that represents that behavior (B) is a function (f) of interaction between the person (P) and the environment (E; Lewin 1943) – commonly called the interactionist perspective. Thus, in order to understand behavior, one must first understand both the person and the environment in which that behavior takes place.

Situational Strength. One of the major avenues in investigating the effect the environment has on human behavior is the work on situational strength. Situational strength concerns the idea that the interpretation of environmental cues can influence the expression of individual differences (Meyer et al., 2010; Meyer, Dalal, & Bonaccio, 2009; Mischel, 1977). The strength of the situation falls on a continuum with which strong situations reduces the likelihood of behavior being representative of the person (P) and weak situation decreasing the influence of situation on behavior. Formulaically, this concept might be represented as:

$$B = f\left(\frac{1}{SS} * P, SS * E\right)$$

where, as the situational strength (SS) increases, so does the situation or environment (E)'s influence on behavior. Furthermore, as the situational strength increases, the behavior is less of a function of the person. This implies that in strong situations, individual differences in behavior will be minimized. In other words, everyone will behave similarly.

Certain characteristics contrast strong and weak situations. In general, situations are laden with 'cues' and the perception and interpretation of these cues will dictate behavior (Meyer, Dalal, & Hermida, 2010). In order for these cues to have an effect on behavior they must first be perceived or recognized. Then, following the cue's recognition, the cue must be interpreted. In

strong situations, most cues are very clear (i.e., easily perceived) and unambiguous (i.e., easily interpreted). In other words, the cues will be uniformly perceived and interpreted which will result in relatively homogenous behavior between individuals (Lievens, Chasteen, Day, & Christiansen 2006; Mischel, 1973). When behavioral expectations are clear, the statistical result will be an attenuation between the correlations of person factors (e.g., personality) and behaviors.

On the contrary, the cues on how to behave in weak situations are less clear and more ambiguous than the cues in strong situations (Beaty, Cleveland, & Murphy, 2001; Cooper & Withey, 2009). This invites the opportunity for the 'person' to play more of a role in determining behavior. Specifically, individual differences will influence the degree to which some cues are recognized and others are ignored (Meyer et al., 2010; Rogers, 1981). This alone would lead to more variations in behaviors as a result of individual differences; however, even the cues that are recognized will need to be processed by each individual. As the cues are more ambiguous in weak situations, the same cue may be differentially interpreted depending on one's standing on common individual difference variables (e.g., personality). In fact, research has demonstrated that personality is a better predictor of performance in autonomous situations (i.e., weak situations) compared to situations where the individual has little control (i.e., strong situations; Barrick & Mount, 1993). Thus, an argument can be made that behaviors that individuals are likely to engage in (i.e., typical behaviors) are better assessed in weak situations. However, ACs are not seen as 'weak situations.'

Situational Strength in ACs. Researchers have stated that ACs do not present a good opportunity to assess typical behaviors (Highhouse & Harris, 1993; Ployhart, et al., 2001). The situational constraints placed upon assessees force their behaviors to be more homogenous.

However, there is room for variability. Specifically, situational strength is continuous in the sense that there is, theoretically, an infinite amount of variability between two different levels of situational strength. Thus, though the situational strength of ACs may be strong, it is plausible that variability in situational strength can exist between two different assessment centers. In fact, very slight changes to the situation can result in drastic differences in behavior (Cooper & Withey, 2009; Johns, 2006).

Furthermore, behaviors on the job do not necessarily take place in very weak situations. Some employees are extensively trained, constantly watched over, and are reminded constantly of how they are expected to behave (Driskell, Willis, & Cooper, 1992; Robie, Brown, & Bly, 2008; Song, Tsui, & Law, 2009). This does not mean, however, that the situations in all jobs negate the influence of individual differences on behavior. For instance, one of the situational factors for assessing maximum performance concerns the amount of time performance is assessed. Maximum performance is best assessed during short durations where individuals are able to 'give it their all' (Sackett, 2007). Considering this, employees are unlikely to be constrained by this situational factor during a normal work day. That is, much like ACs, it is possible for individual difference factors to influence behavior on the job. The goal of ACs then should be to mimic the situational strength of the job. I propose to do this though manipulating the four aforementioned structural components.

Situational Strength and AC Structure.

The general premise of why it may be beneficial to simulate the situational strength of the job in an assessment context can be further explained through examining the factors which commonly explain AC performance. Recent efforts have suggested that performance in ACs can be attributed to three main factors: exercises, dimensions, and a general performance factor (Hoffman et al., 2011). These three factors can be conceptualized as being features of the

situation (exercise factors) or the person (dimension factors, general performance factor). Out of these three factors, exercise factors account for the most variance in AC performance. This implies that the exercises of ACs explain the behaviors of an assessee to a greater degree than the dimensions and a general ability to perform. In other words, situation factors are influencing behavior more so than the person factors. Considering this, one of the goals of the DITLAC is to reduce the degree which AC performance is explained by exercise variance. This is not to say, however, that all exercise variance is erroneous variance. As mentioned earlier, behavior is a function of both the situation and the person. A considerable amount of research has demonstrated that situation factors can explain meaningful variance in performance (Lance, 2008; Lance et al., 2000; Lievens, 2000). Thus, the four manipulated structural components are not designed to arbitrarily reduce exercise variance in ACs. Instead, the goal is to more accurately represent the degree to which 'exercise' variance explains performance on the job in the AC.

Though the exercises utilized by TACs are clear reflections of tasks performed on the job, the structure in which they are presented to assessees is not. In a TAC, an assessee is exposed to various independent exercises that may not share similar contexts, which they are trained for immediately prior to performing, and are given breaks in between each. This is not representative of a typical work day. During a typical work day, an employee may be interrupted several times while performing a task, and interacts with mostly the same group of core individuals throughout the day. Instructions (if they are given) are more typically communicated en mass such as in one overarching morning meeting, and, aside from potentially lunch, breaks are typically up dependent upon the environment (e.g., how busy it is, how many employees are present, how many tasks the individual has to perform, etc.). As even trivial changes in the

environment may denote differences in situational strength, I argue that there are clear situational strength differences between TAC and the typical work environment. Further, I believe that the situational strength of DITLAC is more compatible to that of a normal work environment. I elaborate specifically on how using the four structural components as a framework creates these specific situational strength differences below.

Exercise integration. Exercise integration concerns *how* the exercises are presented to the assessee. In a TAC, an assessee performs each exercise (e.g., in-box, leaderless group discussion, role-play) separately. There is a clear beginning and end to each exercise. However, this is not reflective of a typical work environment. Specifically, this approach does not take into consideration the frequency and types of interruptions. Interruptions are a natural part of any job (e.g., Bailey & Iqbal, 2005; Ratwani, Andrews, Sousk, & Trafton, 2008) and can take several forms. They can range from a simple social interruption concerning a coffee order to an e-mail informing the employee of an impromptu deadline to a self-imposed cognitive recognition that another task needs to be completed. These are the types of situational conditions DITLACs incorporate through exercise integration.

When exercises are not integrated, like in a TAC, assessees have a better understanding of the tasks they need to perform than if the exercises were integrated. That is, an assessee knows that they will not have to choose between tasks associated with an inbox exercise while they are performing the leaderless group discussion. This is less obvious using the DITLAC approach. While performing a task associated with an inbox exercise, it is possible to interrupt the assessee with a task from another exercise. This provides the assessee with a choice – to continue working on the previous task or begin work on the other task. When making the choice between tasks, an assessee will evaluate the two different tasks on various parameters (e.g., value of completing

each task, time it will take to complete each task), parameters which are differentially weighted between individuals (Eyrolle & Cellier, 2000). For instance, an individual high on conscientiousness may choose to complete a task they feel more confident in (e.g., organizing documents) compared to someone who is high in extraversion (e.g., calling a customer back). As individual differences will influence behavior more in these situations, it is suggested that the situation is 'weaker' when exercises are integrated compared to when they are not.

Further, it has been suggested that assessees view each exercise as independent situations (Highhouse & Harris, 1993; Neidig & Neidig, 1984). When exercises are not integrated, it reinforces this perception. When assessees' perceive each exercise as a unique situation, the cues within each exercise are only affecting behaviors for that particular exercise. For example, the informational cues imbedded in the inbox exercise will only affect an assessee's behavior within the inbox exercise. If these cues are only affecting behavior within their respective exercises, behaviors are more likely to be consistent within exercises. In other words, when exercises are not integrated, exercise variance has a greater probability influencing AC performance. As it is not typical for tasks to be as segmented on the job as they are in TACs, the integration of exercises should reduce the degree to which exercise variance is explaining AC performance. Thus, exercise integration should create a weaker situation that reduces the explanatory power of exercises in reference to AC performance.

Context Consistency. Another structural component is context consistency. Context consistency concerns *what* information is presented to an assessee. In TACs, exercises rarely connect in terms of the information they present. That is, each exercise exists in its own narrative-and-informational bubble. This can range from exercises taking place in different contexts (e.g., inbox exercise set in an advertising department, role play exercise set in a sales

department) to where decisions made in one exercise will not be influenced by any information from other exercises. Again, this is not typical of the common work environment. For instance, when exercises take place in different departments (i.e., different contexts) an assessee is likely to be influenced by their preconceived notions of what is appropriate behavior in these departments. Whereas, when the context is consistent, the context will not produce differential effects on assessee's behavior between exercises. An additional example, an administrator making decisions concerning an employee's leave request may be influenced by how rude that employee was to them earlier in the day. These types of situations are apparent when using the DITLAC approach.

As mentioned earlier, situations are filled with cues and these cues serve as indicators of how to behave within the situation. One source of these cues is the social environment of the job (Salancik & Pfeffer, 1978). Cues from the social environment provide information on how an individual should feel and behave, however, if the social environment changes, interpretation of new cues will be independent of older cues. In other words, if the context is not consistent, the cues from one exercise will not be applicable for other exercises. Specifically, TAC's lack of context consistency limits the cross-exercise impact of these cues. That is, all cues within an inbox exercise will only activate information concerned with the inbox exercise. However, when the context is consistent, cues within the inbox exercise could activate information from any other exercise. Keeping with the administrator example, the administrator may allow the fact that the employee was rude to them influence their leave request decision. The degree to which an individual allows this to occur varies with their idiosyncratic dispositional characteristics. That is to say, context consistency should weaken the situation as individual differences will be better determinants of behavior than when there is no context consistency.

Additionally, context consistency should reduce the degree to which AC performance will vary across exercises. When an individual observes cues from the environment, they are inevitably tied to the source of the cues (Festinger, 1954). If the source of the cues (e.g., a supervisor, a coworker) are restricted such they only appear within a single exercise, the result will be more 'exercise-specific' cues. That is, when the context is not consistent across exercises; the assessees'behavior will be differentiated as a function of 'exercise-specific' contextual cues. For example, a role play exercise will contain 'exercise-specific' cues relevant only to that particular exercise when it exists within its own narrative and informational bubble than if it was preceded by an inbox exercise that had a consistent narrative with the subsequent role play exercise.

An individual's interpretation of a cue is partially tied to who is presenting that information to them (Festinger, 1954). The source is judged on multiple aspects (e.g., similarity to the individual) and the information cues the source provides to the individual is subsequently filtered through these judgments. Thus, any future interactions with the same source will be influenced by these existing judgments. A consequence of this for AC context consistency is that when an assessee is presented with an 'exercise-specific' cue (e.g., an individual they have never encountered before), the only information used to decide what behaviors to engage in is the information apparent in that moment. There is no need for the assessee to recall any other information specific to that situation (e.g., previous interactions with that person) in determining how to behave. Consequently, this would result in more exercise specific behavior. In other words, when the context is not consistent, behaviors will be more consistent within exercises and less consistent across exercises. Conversely, when the context is consistent across exercises, the same set of contextual cues affect behavior in all exercises. For example, if the same "character"

presents an assessee with two different pieces of information in two different exercises, the assessee's response to that character will be affected by previous experiences with that character. Thus, behaviors across exercises will be partially determined by the same cues. In turn, through a decrease the amount of exercise-specific cues, exercise variance should account for less variance in DITLAC than in TAC.

Instructional Characteristics. Instructional characteristics concern the frequency, amount, and content presented in instructions. The difference between the instructions that occur in TAC and DITLAC is primarily concerned with the frequency of instructions. When using the TAC approach, often, assessees are provided with instructions right before they perform each exercise, which is not common during a normal job day. On the job, employees may receive a morning briefing concerning all the different tasks they are expected to accomplish during their shift. Once the briefing ends, employees are sent on their way to accomplish these goals. In other words, the information given to employees concerning their daily job duties typically occurs in one bulk session at the beginning of a shift. Translating this analogy to the AC environment, if an AC had three exercise, an assessee would get three different instructional session prior to each exercise concerning the tasks they are going to perform in the subsequent exercise. Conversely, in DITLAC, the assessee would have one single instructional session at the beginning of the assessment which covers all of the different tasks they will perform in the subsequent assessment period.

In contrast to the frequent instructions assessees get right before each exercise when using the TAC approach, the DITLAC approach uses one bulk instructional session. Instructional sessions can be thought of as any other type of situation, filled with cues. Specifically, instructions typically concerns cues that direct employees exactly on how to perform in certain

situations. For example, instructions may inform the assessee about organizational policy regarding the priorities of tasks they will be exposed to. In TAC, these cues remain relatively strong within an exercise as the assessee was just recently exposed to them. However, much like the workplace, the strength of cues provided in instructions may diminish overtime. Furthermore, the impact of cues embedded within the instructions may have less of an influence on future behavior in the DITLAC condition due stress placed on an assessee's working memory during the bulk instructional session. As most adults do not differ significantly in their working memory capacity (Newell & Simon, 1972), the degree to which an assessee is impacted by the cues should be directly related to the amount of information presented in the instructional session (Cooke & Fiore, 2009). As the number of cues stored in working memory increases, the impact of these cues should lessen. Thus, that same cue may not constrain behavior as effectively using the DITLAC approach compare to the TAC approach. Considering the weakened cue strength, the instructional characteristics inherent in DITLAC support the assertion that it is a weaker situation.

Furthermore, when instructions are presented in one bulk session at the beginning of the assessment, this should reduce the degree to which exercise variance explains AC performance. When instructions occur right before each exercise, an assessee knows that this information will be most relevant to the assigned tasks immediately ensuing the instructional session. In other words, it is clear that inbox instructions will be most relevant to the tasks assigned in the inbox exercise. This should result in behaviors being more constant within exercises. Conversely, when the instructions are one bulk session at the beginning of the AC, it is less clear which specific instruction is relevant to which exercise. It would be possible then for instructions from the role play exercises to influence a task associated with the interview exercise. As such, when

instructions are presented in one bulk session, exercise variance should explain less of AC performance than when the instructions are presented immediately prior to each exercise.

Breaks. Breaks are periods in time in which the assessee is removed from situations in which they are expected to perform. As a result of the lack of exercise integration, TACs are likely to have transparent breaks in between each exercise. In addition to the psychological or cognitive breaks these periods provide, assessees are often taken to a different physical location to perform the next exercise. Though breaks are innate to the workplace for some jobs (e.g., some fast food employees required to take 10 minute breaks every 3 hours), the nature of the breaks in TACs is atypical. Employees' breaks are determined by the amount of time that they work on the job. Breaks are not taken once an employee finishes a task. Furthermore, when employees take shorter breaks they are more natural and self-imposed. That is, the employee may find some downtime in the workload and choose to browse the internet or check their personal e-mail. Often, they do not leave their physical location during this period. These are the types of breaks emphasized in the DITLAC approach.

The DITLAC approach can have naturally built in 'rest moments' where the assessee can choose to take a break or not. However, this choice is only available to the assessee if they recognize the cues associated with it. This is not true of the breaks used in the TAC approach. It has been argued that psychological and physical removal from a situation allows an individual the opportunity to reset, refresh, and reframe their current mind set (Finstad, Bink, McDaniel, & Einstein, 2006; Kvavilashvili, 1987; Ury, 1991). Breaks can result in behavioral changes in as little as 3 minutes (Finstad et al., 1991). These transparent breaks cue the assessee that it is time to rest, allowing for at least some downtime. These small rest intervals have shown to improve the learning from instructions (Donovan & Radosevich, 1999), thus increasing the likelihood that

instructions specific to a subsequent exercise will have a strong effect on the assessee. Conversely, assessees in DITLACs are not forced to take breaks; rather their breaks are more natural and self-imposed. Whether or not they take a cognitive break will be dependent on whether they recognize and interpret the cues in the situation as a sign to take a break. As the break cues imbedded in DITLACs are not as clear cut as the break cues in TAC, it should follow that DITLAC creates weaker situations than TAC.

Furthermore, exercise factors should account for less variance in AC performance when breaks are not made transparent. In a TAC, transparent breaks commonly occur between each exercise. The assessee is told that they will need to stop performing, asked to get up, and are taken to a different location before being asked to perform again. These breaks provide an opportunity for the assessee to reset, allowing performance in the previous exercise to have less of an impact on how one will perform in the subsequent exercise. For example, if an assessee performs poorly in the interview exercise, a transparent break will allow the assessee to cognitively reset and refresh such that their previous poor performance will have a less of an impact on future performance in the ensuing inbox exercise. Likewise, when breaks are not transparent, two things may happen. First, breaks may or not be taken. Whether an assessee takes a break will depend on their desire and need to be cognitively removed from the situation. Therefore, some assessee may take breaks and some may not. Second, if an assessee decides to take a break, the time and duration of this break may not be the same as another assessee. For instance, one assessee may take a break halfway through the assessment, whereas another assessee may wait until three quarters through the assessment to take a break. Thus, break taking in these circumstances are by and large determined by person factors. When transparent breaks occur between exercises, this allows the assessee to refresh and reset before performing in the

next exercise. As such, transparent breaks should increase the degree to which AC performance is determined by exercise factors.

When all four structural components are considered in unison, it can be argued that DITLAC present a better platform to assess typical performance whereas TAC provide the opportunity to rate maximum performance. Specifically, TACs present cues that reinforce the three conditions of maximum performance: Knowledge of being evaluated, short performance periods, and acceptance of instructions to do their best (Sackett et al., 1988). First, in TACs relative to DITLACs, assesses are reminded more often that they are being assessed. For instance, even if the assessor is not present during the exercise, when the assessee is moved from one exercise location to another, the presence of the assessor is reinforce. In turn, this reminds the assessee they are being assessed during this period. Further, the cues associated with a changing context or transparent breaks allow an assessee to know when he/she is not being assessed and when they are being assessed. Transparent breaks mean that the period of time in which assesses must sustain effort is shorter. Lastly, at the end of each instructional session, and prior to the start of a new exercise, assesses recommit in a sense to the notion of doing their best. . As such, TACs should better facilitate the assessment of maximum performance relative to DITLACs, which should better gauge typical performance.

Four things should be clear now. First, due to the structural differences, the situational strength of a normal day on the job is more similar to DITLAC than TAC. Second, both a normal workday and the DITLAC represent weaker situations than that of the TAC. Third, performance in TACs should be better explained by exercise variance than DITLACs. Lastly, TACs are better apt to assess maximum performance, whereas DITLACs are better suited to assess typical performance. The consequence of these four points is that behaviors exhibited in DITLAC

should be a better indicator of typical performance on the job than should behaviors in a TAC . Further, because the behaviors expressed in DITLACs have a greater likelihood of being influenced by individual differences than in TACs, behaviors across situations (i.e., exercises) should be more consistent. This holds critical implications for the construct validity problem as discussed below.

Validity Comparison: DITLAC vs. TAC

The traditional approach to establishing construct validity in ACs has been the MTMM matrix. Briefly, construct validity using a MTMM matrix is inferred when same-dimension scores across exercises correlate strongly with one another (i.e., convergent validity) and when different dimensions scores within exercises have a weak to null correlation (i.e., discriminant validity). This approach relies on the assumption that an individual's behavior will be consistent across situations internal to the AC. However, this is not the only way to provide construct validity evidence. Through using Cronbach and Meehl's (1955) conceptualization of the nomological network, construct validity can also be established through using indicators of the dimensions measured within the AC which are external to the AC itself. For instance, if conscientiousness is related to the dimension of planning and organizing, evidence for convergent validity can take the form of strong correlations between a self-reported measure of conscientiousness and planning and organizing PEDRs. The following section describes the rationale behind these two sets of establishing construct validity approaches and the expected differences between the DITLAC and TAC designs.

Nomological Network Construct Validity

When the concept of construct validity was first coming to form, Cronbach and Meehl (1955) published an article both describing the concept and instructions on how to provide

evidence for construct validity. The basic idea on how to establish construct validity was to create a nomological network. In essence, the nomological net offers evidence for construct validity by showing that conceptually related measures (or observables) of a construct are related to one another while the relationship between two measures of theoretically unrelated constructs is admissible. The concept of establishing a nomological network is the basis for the forthcoming set of hypotheses.

One of the key aspects of the nomological network is establishing theoretical ties between the two constructs each measure purports to measure. In this vein, it is critical to consider the sources of behavior in both DITLAC and TAC. As alluded to earlier, situations are stronger in TACs as compared to DITLACs. Thus, trait-related individual differences should play less of a role in determining behavior in TACs than in DITLACs.

Behavioral Manifestations of Personality

Personality has been long studied as a motivational force in the manifestation of certain behaviors. Some early research suggested that personality reflects a series of needs an individual wishes to satisfy (e.g., Allport, 1937). The culmination of this need takes the form of behavioral expression. When these needs are satisfied, individuals experience a degree of pleasure (Murray, 1938), however when they are not satisfied one experiences a sense of anxiety (e.g., Bakan, 1966; Wiggins & Trapnell, 1996). In specific reference to ACs, research has demonstrated that behaviors exhibited in ACs are related to the assessee's personality (e.g., Furham, Crump, & Whelan, 1997; Haaland & Christiansen, 2002). Furthermore, cues within a situation may guide a person to behave in a certain way which is in line with their personality. Trait Activation Theory (TAT) posits that individuals will only exhibit behavioral representations of certain traits (e.g., personality facets) when the situation cues or activates these traits (Haaland & Christiansen, 2002; Lievens et al., 2006; Tett & Burnett, 2003). A situation's trait activation potential (TAP) is

a function of the extent it provides trait-relevant cues. In other words, one of the conditions for personality to manifest into behavior is that the situation has trait-relevant cues. These cues, however, will only stimulate trait-related individual differences in behavior weak situations. If situational cues are strong enough, individual differences in personality traits will have a negligible impact on behavior. In general, the cues in weaker situations are less clear and more ambiguous. Thus, this suggested that the perception and interpretation of cues can be attributed to trait-related factors (i.e., personality)

Both the perception and interpretation of cues in the situation are said to be a function of individual differences (Meyer et al., 2011; Rogers, 1981). In reference to personality, it can be argued both the detection and analysis of situational cues are the result of an individual's level on certain personality characteristics. For example, an individual high on extraversion will engage in behavior reflective of extraversion. This is the basic premise behind Implicit Trait Policy (ITP) theory. ITP states that people who are high on a personality trait will consider behavioral manifestation of that trait to be more effective than other behaviors (Motowidlo & Beier, 2010). Thus, they are more likely to express behaviors in line with that trait (Motowidlo, Hooper, & Jackson, 2006; Oostrom, Born, Serlie, & van der Molen, 2012). These ideas may be best illustrated in an example concerning football.

With this in mind, it should hold that personality will be more predictive of trait-relevant behaviors in a DITLAC than in a TAC. DITLACs produce situations similar to that of a normal work day through mimicking the meta-structure of a typical performance situation. This, in turn, reduces the strength of the situation on an assessee's behavior (i.e., creating a weaker situation). Conversely, because of the lack of exercise integration, the multiple contexts, segmentation of instructions, and presence of transparent breaks, TACs are argued to represent a stronger

situation than that of DITLACs. Due to the strong situations imposed upon the assessee in TACs, variance on behavioral measures due to personality will be minimized – reducing the relationship between personality and behavioral metrics. Conversely, assessees should be better able to express trait relevant behavior in DITLAC as it represents an overall weaker situation.

Additionally, it is also proposed that the DITLACs will provide more discriminant validity evidence than TACs. Since the situational characteristics of TACs attenuate the influence of traits on behavior, the relationship between a personality trait and trait-relevant PEDRs should be no different from the relationship between a personality trait and trait-irrelevant PEDRs. That is, as trait-related variance is removed from PEDRs in the TAC condition, PEDRs should no longer be predicted by relevant traits and, thus, should have a relationship similar to that of a personality trait and a trait-irrelevant PEDR. However, there should be a difference in these two relationships in DITLAC. In DITLAC, personality traits should have more of an impact on how individuals behave as it is a weaker situation. In weaker situations, person characteristics, in this case personality, should be more predictive of behavior. Therefore, the pattern of relationships between personality traits and PEDRs should be significantly stronger for matched pairs than for unmatched pairs. This represents evidence of external discriminant validity. In this regard, for stronger situations, correlations between matched and unmatched pairs of traits with dimension ratings should not demonstrate as clear a pattern.

The strength of the inferences one can make concerning a nomological network is directly related to the amount of indicators within the network (Cronbach & Meehl, 1955). Thus, additional indicators would strengthen the claims concerning construct validity of both TAC and DITLAC. A logical supplementary indicator would be observations of typical performance from

an outside observer. An outside observer's judgment of an assessee's performance lends insight as the information used to make ratings comes from outside the assessment context. Outside observers have firsthand accounts of behavioral manifestations of personality variables as they are around the assessee in a variety of situations. Thus, typical performance ratings will be used as an additional node in assessing the construct validity of both ACs. As such, the same pattern of results from the first two hypotheses is expected to occur using typical performance ratings in lieu of self-reported personality test. Considering this, the first two hypotheses state:

Hypothesis 1a: The relationship between conceptually-matched external factors and PEDRs will be significantly stronger in the DITLAC than in the TAC

Hypothesis 1b: External factors will explain significantly more variance in conceptually-matched PEDRs in the DITLAC than in the TAC

Hypothesis 2: External factors will explain significantly more variance in conceptually-mismatched PEDRs in the TAC than in the DITLAC.

This set of hypotheses asserts that construct validity evidence will be superior in DITLACs when compared to TACs. Specifically, through using the nomological network approach, construct validity evidence is demonstrated by showing that conceptually related measures are more strongly related to each other than they are to conceptually unrelated measures. If these hypotheses are supported, it will present evidence of the construct validity supremacy of DITLACs over TACs. However, it is also argued that DITLACs will provide better construct validity evidence when using the traditional method of validation.

MTMM Construct Validity

The conventional AC literature has had a consistent problem with using the MTMM matrix (Campbell & Fiske, 1959) approach to establishing construct validity (e.g., Lance, Noble, & Scullen, 2002; Lievens & Conway, 2001; Sackett & Dreher, 1982). The results consistently seem to demonstrate poor evidence of convergent validity (low SDDE correlations) and

discriminant validity (high DDSE correlations). Several approaches have been taken to help correct for this problem, with most taking the form of 'corrections' to the assessors. Aside from one stream of literature (i.e., skill transparency; Kleinmann, 1993), none of these approaches sought to change the conditions under which the assessee was performing. This infers that behaviors exhibited in ACs are the same behaviors an assessee will engage in on the job. However, the aforementioned theoretical rationale challenges this opinion.

Due to the situational weakness of DITLACs compared to TACs, behaviors across exercises should be more cross-situationally consistent. Research has demonstrated that similar situations elicit similar behaviors (Haaland & Chistiansen, 2002; Mischel & Shoda, 1998; Tett & Gutterman, 2000). However, this is not to say that similarly strong situations will result in behavioral consistency across exercises. Behavioral ratings between situationally strong exercises should only be related to one another if the strong cues signal the same behavior. In fact, research has shown that assessees perceive each exercise as being entirely different situations (Highhouse & Harris, 1993; Neidig & Neidig, 1984). Thus, in a TAC, there should be no reason to expect behavior to be consistent across exercises. Further, rated behaviors from the DITLAC should be more cross-situationally consistent. Unlike TAC, behaviors in DITLAC have the same dispositional determinant: personality. This should result in more behavioral consistency across exercises. Using the MTMM matrix framework, this implies that same dimension ratings taken from different exercises should be significantly stronger in DITLAC than in TAC. That is, PEDRs from DITLACs should provide more convergent validity evidence than TACs. Thus, the next hypothesis states:

Hypothesis 3: *Same-Dimension Different Exercise (SDDE) correlations will be significantly stronger in DITLAC than SDDE correlations from TACs.*

Moreover, as the assessee is more likely to view exercises as distinct situations (Highhouse & Harris, 1993; Neidig & Neidig, 1984), they should perform more consistently within exercises in a TAC. That is, the determinant of behavior in TAC is the situation rather than personality. Thus, relationships between behaviors within an exercise should be stronger than behavioral ratings across exercises for TACs. These would represent the same exercise effects prolific throughout the AC literature. As the situation is a weaker determinant of behavior in the DITLAC, behaviors should be less consistent within exercises as compared to the pattern of behaviors in TACs. In terms of the MTMM matrix framework, this would constitute evidence for discriminant validity.

Hypothesis 4: *Different-Dimension Same Exercise (DDSE) correlations will be significantly stronger in TAC than DDSE correlations from DITLAC.*

Summary

Each of these hypotheses provides information concerning the construct validity of two different types of ACs in different ways. Construct validity evidence has notoriously been difficult for ACs to demonstrate. As such, support for these hypotheses could steer AC research and practice in new directions. If the simple change in structural components can result in better construct validity evidence, then the impact of this research will be great. In order to do this, however, two different ACs will be designed that solely manipulates the structural components: one DITLAC and one TAC. That is, every participant in the proposed laboratory will generally receive the same information concerning task components – the only difference between conditions will be variation in the structural components. The details of how this will be done are explained further in the following section.

Table 2. *Overview of Hypotheses*

Hypothesis 1a	The relationship between conceptually-matched external factors and PEDRs will be significantly stronger in the DITLAC than in the TAC
Hypothesis 1b	External factors will explain significantly more variance in conceptually-matched PEDRs in the DITLAC than in the TAC
Hypothesis 2	External factors will explain significantly more variance in conceptually-mismatched PEDRs in the TAC than in the DITLAC.
Hypothesis 3	Same-Dimension Different Exercise (SDDE) correlations will be significantly stronger in DITLAC than SDDE correlations from TACs.
Hypothesis 4	Different-Dimension Same Exercise (DDSE) correlations will be significantly stronger in TAC than DDSE correlations from DITLAC.

Notes. PEDR = Post-Exercise Dimension Ratings
SDDE = Same Dimension Different Exercise
DDSE = Different Dimension Same Exercise
TAC = Traditional Assessment Center
DITLAC = Day-In-The-Life Assessment Center

CHAPTER 3: METHOD

Participants

Participants included 324 pairs of individuals from a large southeastern university. The participants were recruited via the University's SONA-Systems program. This program allows students to sign up for research studies. In return for their participation, students receive credit for a course they are taking. The average age of the focal participant was 19.13 years. Around 54% of the sample responded that they identify themselves as White, 22% as Hispanic, 13% as Black, 5% as Asian, and less than 1% as American Indian. Males consisted of 28% of the sample, 68% reported as female. 163 participants were assigned to the Traditional AC and 159 were assigned to the Day-In-The-Life AC.

Assessment Center Design

Assessment Center Development.

The first step in development was to identify the specific dimensions relevant to the job being assessed. The job is an Emergency Room Administrator, which is similar to a medical secretary. They are responsible for tasks that are more clerical in nature (e.g., planning schedules, answering e-mails, etc.) as well as tasks which require the interaction with different people (e.g., patients, coworkers). In order to identify the specific dimension the AC needs to assess, the job of Medical Secretary was used as a basis. The skills and abilities needed to perform this job were taken from O*NET and can be seen in Table 3. These abilities and skills were then translated into four commonly assessed dimensions as identified by (Arthur et al., 2003) and one additional dimension (also seen in Table 3). Specifically, the dimensions that will be assessed in this AC are communication, influencing others, consideration, planning and organizing, and aggressiveness. Aggressiveness was added as low scores on intuitively

associated dimensions (i.e., influencing others, consideration) represent the lack of aggressiveness – not necessarily the presence of aggressiveness. Furthermore, aggressiveness is a commonly measured dimension in customer-service oriented ACs (e.g., Hinrichs, 1978; Sackett & Harris, 1988).

Table 3. *Medical Secretary Job Duties to Assessment Center Dimensions*

O*Net Skills and Abilities	Assessment Center Dimensions
Speaking	Communication
Speaking Clearly	Communication
Written Comprehension	Communication
Complex Problem Solving	Influencing Others
Critical Thinking	Influencing Others
Deductive Reasoning	Influencing Others
Service Orientation	Aggressiveness
Social Perceptiveness	Aggressiveness
Oral Expression	Aggressiveness
Service Orientation	Consideration
Active Listening	Consideration
Social Perceptiveness	Consideration
Coordination	Planning and Organizing
Time Management	Planning and Organizing
Information Ordering	Planning and Organizing

The Assessment Center Description

The focal participants recruited for this study ran through a four hour assessment center that was set in the context of a hospital emergency room. Specifically, they played the role of a job recruit applying for the position of an Emergency Room Administrator. This assessment center was made up of three exercises: An inbox exercise, a role play exercise, and an interview exercise. Though elaborated upon in the following section, each participant was randomly assigned to either a TAC or a DITLAC condition. Though these two conditions differ in terms of

their structural components, each participant was exposed to the same exact tasks. That is, both ACs were designed in such a way that the only difference between them is their structure. All of the tasks from all of the exercises are exactly the same. Each exercise is described in fuller detail below.

Inbox Exercise. For this exercise, the focal participant were told that a coworker's wife unexpectedly went into labor and had to leave for the day. However, the expecting father did not complete all of his work and the focal participant will need to complete these tasks. The focal participant was responsible for responding to both angry and calm e-mail complaints from former patients, digitally logging these complaints, creating a work schedule, logging paid-time-off employee requests, and entering patient records into a digital database. The first three tasks (answering patient e-mails, creating a schedule, logging paid time off requests) were specifically designed to tap personality characteristics. As such, pilot studies were conducted to see how long it took the average participant to complete these three tasks alone. Times ranged from 45 minutes to almost an hour. In order to ensure that everyone could complete the three tasks, the exercise length was set at an hour. Further, if each participant was not working for the same amount of time, it could confound with one of the structural components (e.g., breaks). Thus, a monotonous task of copying over physical patient packets into a digital form was added. Greater descriptions of the tasks are provided below.

First, the focal participant was expected to respond to and catalog each patient complaint e-mail. These e-mails are listed in Appendix A.1. In order to log each patient experience, the focal participant needed to enter basic information (e.g., name of patient, e-mail address), briefly describe the patient's experiences, rate if the patient's experience was positive or negative, and then lastly rate how positive or negative they believed the patient's experience was. An example

of this form is provided in Appendix A.2. Furthermore, the focal participant needed to respond to each patient directly. These e-mail responses were recorded so that they may be rated later. The specific process on how ratings were made is further elucidated upon in the Post-Exercise Dimension Ratings section.

The second task of the in-box was creating a schedule for the following week. In this task, the focal participant needed to read over the rules and regulations concerning the restrictions associated with making a schedule (e.g., two individuals must be on shift at all times, shifts are limited to 8 hours a day, no one can work more than 30 hours in one week), consolidate each employee's availability (sent in through e-mails) to an excel document, and finally create the schedule to send in to their supervisor. They were required to send in the final schedule both through e-mail and over a telephone call. After pilot testing this task, it was determined that there were multiple solutions given the availability of each employee. That is, it is possible for different focal participants to create different schedules with each different schedule being equally correct. In order to facilitate this process, each focal participant was given two excel documents. The first excel document was designed to help with the consolidation process, while the second excel document was designed to assist them with creating the final schedule. Examples of these forms can be seen in Appendix B.1 and B.2.

The third task in this exercise concerns the completion of paid-time-off requests. The focal participant received e-mails from their coworkers requesting paid-time-off for multiple reasons. The focal participant needed to enter these requests into an online form and they had a certain amount of discretion concerning whether or not these requests are approved or denied. The focal participant was provided with guidelines on how to make these decisions; however,

these guidelines are relatively clear. Examples of the requests and the online forms are provided in Appendix C.

Lastly, the focal participant was asked to transfer physical patient records into a digital database. The identifying information contained within these packets were based in fiction, however, the validity of the medical aspects were verified by a practicing physician in the State of Florida. At first, the focal participant was only given 10 of these forms to complete, though there will be additional forms to complete if necessary. These forms contain a great deal of information about a patient's visit including information about the patient's medical history, reason for coming in, and forms completed by the triage nurse and the assigned physician. Examples of these forms are provided in Appendix D.

Role Play Exercise. During this exercise, the focal participant had to directly interact with customers and coworkers in an emergency room setting. The focal participant believed that they are testing out new software that directly connects the focal participant to the hospital via 'The Virtual Office.' In actuality, the participant responded to pre-recorded videos that were designed such that no matter how the participant responds, the subsequent reaction from the video will not conflict with that response. From the perspective of the focal participant, the coworkers and customers will appear to be talking directly to them. When the focal participant was asked to respond, the software used to run this digital scenario will loop the video – such that it will appear that the customer or coworker is listening to the focal participant's response. The computer program detected when the focal participant starts and stops talking. Thus, once the focal participant stopped talking, the program moved onto the next video – seamlessly transitioning between the loops and the next 'scene.'

The tasks within the role play exercise concern the interpersonal interaction the focal participant had with customers/patients, coworkers, and supervisors. The events within the role play exercise are designed to tap a variety of behavioral responses. For instance, there are events within the role play exercise where (a) a coworker breaks some rule and asks the focal participant to lie for them; (b) the focal participant was falsely accused of breaking the rules; and (c) the focal participant needed to respond to both rude and friendly patients. Furthermore, the focal participant needed to make public address announcements, answer e-mails, and respond to voice messages. Specific events designed to tap targeted-dimensions are describe in more detail in the section on Post-Exercise Dimension Ratings below. This exercise lasted from 40 to 45 minutes.

Interview Exercise. For this exercise, the focal participant believed they were interviewed by a senior employee of the hospital. Before the interview began, the focal participant was asked to review a packet of information concerning (a) sexual harassment events that occurred in the past, (b) an employee disciplinary situation, and (c) material concerning the hospital's mission statement initiative. After allowing 20 minutes to review this material, the focal participant was asked a series of questions from an interviewer. These questions concerned their opinion on the sexual harassment situation and employee disciplinary situation, their articulation of the points the mission statement was designed to convey, and their perspective on what work still needs to be done. The interviewer also asked a series of distraction questions. The interview itself lasted around 10 minutes.

In order to keep with the high fidelity nature of ACs, the medical aspects of the exercises were reviewed by a practicing physician in the state of Florida. Thus, everything from the diagnosis of symptoms on the patient records to the patient complaints were reviewed for their

fidelity. Further, the experimenters for this experiment acted as if they are employees of the hospital, not research assistants in a lab. They were provided with name tags and dressed professionally for each session. Lastly, the entire environment in which the focal participant was in was designed to look like a hospital waiting room. This included environmental aspects such as medical signage (e.g., no smoking, medical information), hanging scrubs, and plant decorations. All of these aspects were implemented in hopes to further immerse the focal participant in the experiment.

Experimental Manipulation

There were two experimental conditions in this study: The Traditional Assessment Center (TAC) condition and the Day-In-The-Life Assessment Center (DITLAC) condition. As mentioned earlier, there are specific structural differences between the two types of AC. These structural components include exercise integration, context consistency, instructional characteristics, and breaks. The specific way each of these four structural components was manipulated is expanded upon below.

Traditional Assessment Center. In the TAC condition, the participant was explicitly told that they are participating in three different exercises. There were clear cues informing the focal participant of the beginning and end of each exercise. Thus, there was no exercise integration. Further, there was very little context consistency. Each exercise occurred in its own narrative and informational bubble. For example, no information from the inbox exercise cued any information in the role-play exercise or interview exercise. Within each exercise the focal participant interacted with the same set of characters and none of these characters appeared in any other exercise. However, there is a degree of context consistency in that they played the same role (i.e., job recruit for the position of emergency room administrator) the entire time – however, this is similar to how some TACs are conducted. Third, before each exercise, the focal participant

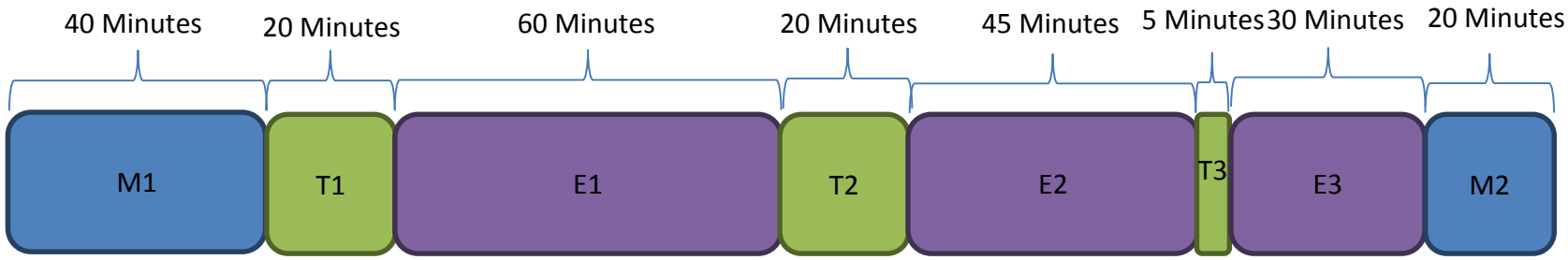
received instructions regarding the specific tools they would be using and organizational policies relevant to only that exercise. In other words, they went through three instructional sessions – one for each exercise. Lastly, there were transparent breaks that occur between each exercise. The focal participant was physically moved to a different location in between each exercise. These were short breaks (approximately 3 minutes in length), but noticeable.

In contrast to the TAC condition, focal participants in the DITLAC experienced fully integrated exercises. They were interrupted several times during their performance period and had the opportunity to perform a task associated with a different exercise. They were not be told that they would be experiencing three different exercises and the structure of the exercises should not make this fact apparent to them. Further, the context was consistent across exercises. That is, the focal participants interacted with the same core group of individuals throughout all three exercises. For instance, each of the customer complaints mentioned an employee from the role play exercise by name. Also the employees on the schedule from the inbox exercises were the same individuals involved in the sexual harassment scenario from the interview exercise. Third, focal participants in the DITLAC condition received one bulk instructional session that covered all of the same material as the three instructional sessions from the TAC condition, however, it occurred in one continuous block of time. Lastly, there were set periods within the DITLAC condition where the focal participant took quick breaks, however, these periods were not as transparent as they were in the TAC condition. Specifically, there were cues informing the participant that the completion of a certain task is not absolutely necessary and other employees will take care of it later.

Despite these structural differences, each condition took approximately three hours. That is, the amount of time that participants spend on instructions and performing tasks was exactly

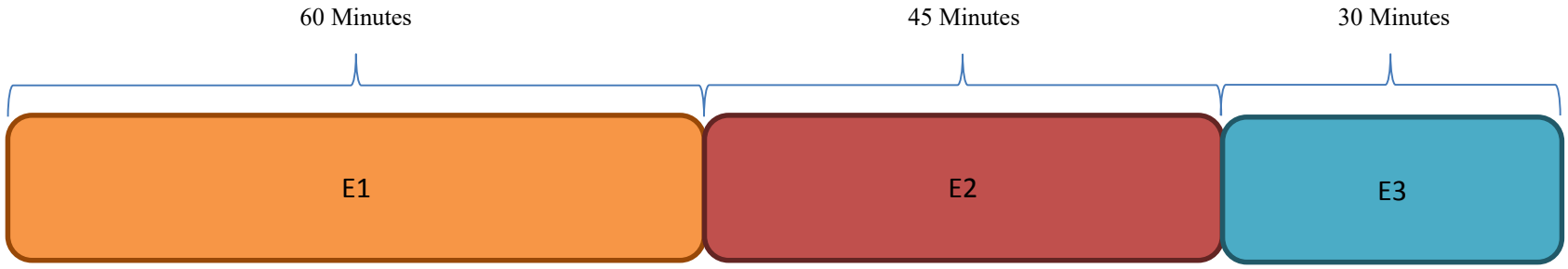
the same between conditions, however, the order in which participants experienced each of the components differed between conditions. Specifically, each participant received 20 minutes of instructions for the inbox exercise, 20 minutes of instructions for the role play exercise, 5 minutes of instructions for the interview exercise, 60 minutes to perform the inbox exercise, 45 minutes to perform the role play exercise, and 30 minutes to perform the interview exercise. An image that focuses on how these times are broken down can be seen in Figure 1.

Traditional Assessment Center Study Timeline

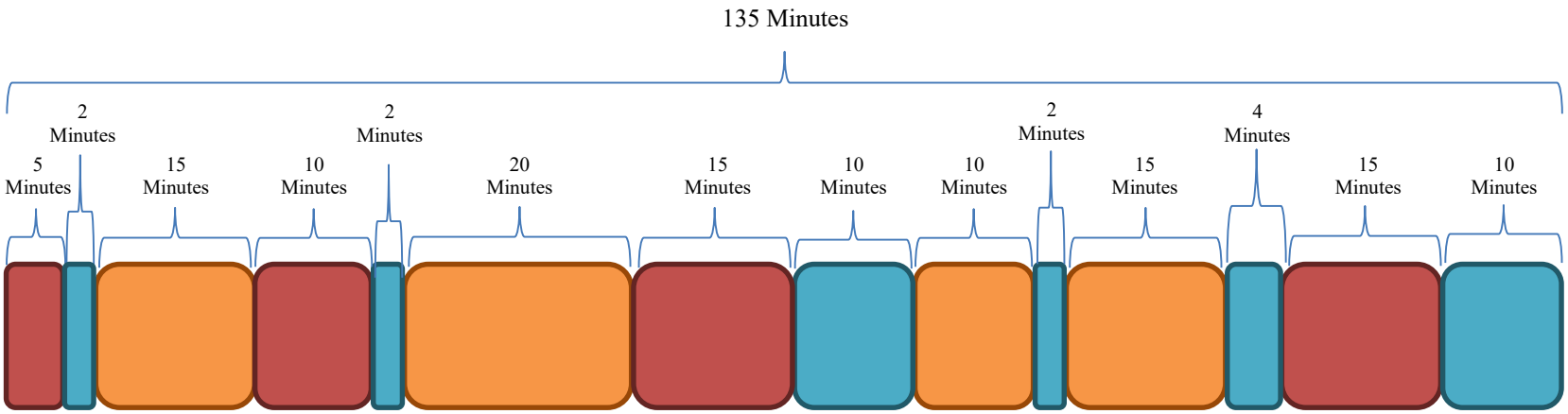


Day-In-The-Life Assessment Center Study Timeline

Traditional Assessment Center Exercise Organization



Day-In-The-Life Assessment Center



Note. M = Measurement Period; T = Instructional Period; E1/Orange Cells = Inbox Exercise; E2/Red Cells = Role Play Exercise; E3/Light Blue Cells = Interview Exercise

Figure 1. *Traditional and Day-In-The-Life Assessment Center Layouts*

Measures.

Personality. Self-rated personality was assessed using Form S and Form R of the NEO-FFI-3 (McCrae & Costa, 2008). Form S is designed to assess the personality of the individual filling out the measure. In other words, the items are phrased in the first person (e.g., I am not a worrier). Form R is designed to assess the personality of the acquaintance of the individual filling out the measure. That is, the items are phrased in the third person (e.g., she/he is a worrier). This measure evaluates four personality variables (Neuroticism, Extraversion, Openness to Experience, Agreeableness, and Conscientiousness) 12 items per dimension, on a scale from 1 (*Strongly Disagree*) to 5 (*Strongly Agree*). Sample items from each dimension are "When I am under a great deal of stress, I sometimes feel like going to pieces," "I really enjoy talking to people," "I generally try to be thoughtful and considerate," and "I keep my belongings neat and clean" representing Neuroticism, Extraversion, Agreeableness, and Conscientiousness respectively. Extraversion ($\alpha = .79$), Neuroticism ($\alpha = .82$), Agreeableness ($\alpha = .81$), and Conscientiousness ($\alpha = .89$) were sufficiently reliable.

Assertiveness. Assertiveness will be assessed using a measure designed by Lorr and Moore (1980). The measure contains 31 items on a 6-point Likert Scale (1 = *Strongly Disagree*; 6 = *Strongly Agree*). Example items from this measure include "I nearly always argue for my viewpoint if I think I am right," "It is easy for me to make "small talk" with people I have just met," and "I defend my point of view even if someone in authority disagrees with me." ($\alpha = .75$)

Post-Exercise Dimension Ratings. Post Exercise Dimension Ratings were made by 10 trained raters. Two raters will be assigned to a single dimension. That is, each rater did not rate more than a single dimension. This is designed to reduce the potential of rater errors. One of the issues underlying assessment center field data is the opportunity for trained raters to make errors

in their judgments of behaviors associated with targeted dimensions. These errors occur for multiple reasons. First, in cases where a rater is assigned to follow and rate a single applicant, they will need to have a thorough understanding of all of the dimensions being assessed. As the number of dimensions assessors are requested to assess can reach as many as twenty five (Sackett & Hakel, 1979), it is understandable that the assessor may experience cognitive strain. The human mind can only hold a few distinct concepts simultaneously in working memory Lachman, Lachman, & Butterfield, 1979). Researched has postulated that such a heavy cognitive load has led raters to make incorrect judgments of dimensions and recommend that raters are assigned less dimensions to rate (Gaugler & Thornton, 1989). Additionally, another issue in field assessment center ratings is that they are often made on the spot. That is, once the exercise has concluded, ratings will need to reflect on the behaviors they just witnessed and mark down the degree to which it represents a specific dimension. Given that assessment centers last all day and that exercises can be more than an hour long, it may be difficult for the rater not to engage in recency bias or result to committing halo errors.

In order to avoid these issues, raters were assigned only one dimension and all judgments were made on recorded responses. This approach allowed raters to experience less cognitive strain as they will only have to rate one dimension. Furthermore, the additional cognitive strain associated with understanding all manifestations of a dimension will not overload the rater. Lastly, if the rater felt like they missed something while the participant was responding, they had the opportunity to go back and review these responses because all responses were recorded.

These raters were trained in a manner similar to that of the partner participant. They were exposed to the definitions of their assigned dimension, provided examples of the extreme scale points, and then will be asked to make their judgments. In order to assess reliability, two raters

were assigned to each dimension. In order to establish reliability, rater-pairs used pilot data. During the initial training, they rated a series of participants, compared their answers for each event they rate, and then come to a consensus to the score to give that specific event. Both their initial ratings and their consensus ratings were recorded.

As mentioned previously, five dimensions were assessed in this assessment center and all dimensions were assessed in each of the exercises. However, some exercises presented more opportunities for the focal participant to exhibit dimension-specific behaviors. This is a factor of both the design of the assessment center as well as the implicit nature of the dimension itself. For instance, it is easy to design several events that target the dimension of consideration as the opportunity to act considerate can take place within one interaction (e.g., interaction between focal participant and customer, between focal participant and coworker). However, this is not true for the dimension of planning and organizing. In order to assess an individual on the dimension of planning and organizing, one must make judgments on a series of events, not just a single instance. As such, the number of instances a dimension was rated differed. Each dimension was rated on a 6-point Likert scale (1 = *Not at All*; 6 = *Definitely*) concerning the degree to which they believe the focal participant's behavior in the below scenarios represents the dimension. The definitions of each dimension and a general description of the events in which they will be assessed are reported below.

Communication is defined as 'the extent to which an individual conveys oral and written information and responds to questions and challenges' (P. 133, Arthur et al., 2003). In the inbox exercise, this dimension was assessed through rating an audio recording of the focal participant reporting a schedule to their supervisor. For the role play exercise, communication was assessed by rating the audio files of the participants making public address announcements concerning

critical patient updates. Lastly, for the interview exercise, participants needed to reiterate SUMMIT Hospital's mission statement in such a way that it will be clearly conveyed to patients of the hospital. ($\alpha = .76$)

Influencing Others is defined as 'the extent to which an individual persuades others to do something or adopt a point of view in order to produce desired results and takes action in which the dominant influence is one's owns convictions rather than the influence of others' opinions' (P. 134, Arthur et al., 2003). In the inbox exercise, participants were required to respond to two angry and two calm customers who are upset over their latest visit to SUMMIT Hospital. Specifically, they were rated on their ability to stay calm and persuade the patient that their experience was not a typical one. During the role play exercise, participants were confronted by both angry and calm coworkers and customers requesting that they break the rules. The participant were rated on their ability to stay calm, stand up for themselves, and resolve the situation. Finally, in the interview exercise, the participant were presented with a scenario where they have been falsely accused of breaking the rules. In order to rate influencing others, the participant were required to provide a verbal testimony of the events that transpired and any defense they have, if any. ($\alpha = .80$)

Aggressiveness is defined as the extent to which an individual uses coarse language and shrill tone of voices in workplace conversations as well as resigning to threats and manipulations in order to achieve their personal goals at the expense of others. As mentioned earlier, influencing others and aggressiveness are relatively similar dimensions. Thus, many of the opportunities to express aggression were the same as influencing others. Therefore, the scenarios used to gauge influencing others were also be the scenarios aggressiveness was rated on ($\alpha = .70$).

Consideration is defined as 'the extent to which an individual's action reflect a consideration for the feelings and needs of others as well as an awareness of the impact and implications of decisions relevant to other components both inside and outside the organization' (P. 133, Arthur et al., 2003). For the inbox exercise, consideration was rated using patient e-mails which detail their latest experience at SUMMIT Hospital. Specifically, these e-mails tapped the participant's willingness to provide sympathetic, empathetic, and supportive behaviors. During the course of the role play exercise, the participant were exposed to situations where two parents are searching for their child who has been in a school bus accident and an employee is confiding in the participant regarding a sexual harassment incident. Consideration for this exercise was assessed through the participant's articulation of sympathy, empathy, and their willingness to go above what is required of them. Lastly, the interview exercise presented the participant with a sexual harassment situation. They were specifically asked to provide an official testimony of the events they witnessed. Much like the previous exercise, they participant's response was judged on their degree of sympathy, empathy, and willingness to go about that which is required of them ($\alpha = .87$).

Lastly, Planning and organizing concerns the extent to which an individual systematically arranges his/her own work and resources as well as that of others for efficient task accomplishment; and the extent to which an individual anticipates and prepares for the future (P. 135, Arthur et al., 2003). This dimension was assessed in the inbox exercise through an assessment of approach to completing all of the tasks. Specifically, they were assessed on whether they complete the tasks in a logical order, the degree to which they switch between tasks, and whether or not they complete the three main tasks (patient e-mails, schedule, and paid time off requests). For the role play exercise, participants were asked whether or not employees

should take breaks during an extremely busy time. This decision required the participant to engage in foresight and access their time-management skills. For the interview exercise, the participants was asked to recall the events of the day for an employee coming in after them. They specifically needed to tell this employee what tasks are left to complete and the criticality of these tasks (α = .85)

Partner Ratings of Targeted Dimensions. Partners of the focal participant completed a measure asking them to rate the focal participant on all five targeted dimensions in both a context generic and context specific scenarios. After going through the initial training for each dimension described above, context-generic items simply asked the partner to rate the degree to which the focal participant represented each of the target dimension on a 6-point Likert scale (1 = *Extremely Poor*; 6 = *Extremely Well*). For the context-generic portion of the measure, there will be only one item per dimension. Following this, in order to complete the context-specific portion of the measure, partners went through 5 sets of ratings, each representing one of the targeted dimensions. Each section began by reinforcing the dimensions through a review of the definitions and behavioral examples. Next, the partner witnessed the same events as the focal participants which are designed to assess that specific dimension. After each event, the partner were asked the degree to which they believe the focal participant's behavior in these scenarios would represent the dimension they were rating on a 6-point Likert scale (1 = *Not at All*; 6 = *Definitely*). Reliabilities for Communication (α = .82), Influencing Others (α = .92), Aggressiveness (α = .94), Consideration (α = .89), and Planning and Organizing (α = .86) were all in acceptable ranges.

Cognitive Ability. Cognitive ability was be assessed using the Wonderlic Contemporary Cognitive Ability Test. Cognitive ability will be used as a control variable as it has been related

to general AC performance (Goldstein, et al., 1998; Spector, Schneider, Vance, & Hezlett, 2000), ATIC (König, Melchers, Kleinmann, Richter, & Klehe, 2007), and flexibility (Arthur et al., 2003). As this measure is proprietary, exact examples items are not available. However, the publisher of this measure has provided representative example items, which are included in Appendix E.

Demographic Information. Participants filled out a series of measuring asking them some basic demographic information. This includes information like age, gender, racial identification, GPA, and major. Most of these measures were primarily be used as descriptive indicators of the sample.

Procedure.

Though the procedures differed slightly between the two conditions; the first hour of the experiment was essentially be the same. Below the procedures for each condition are described. The layout of the experimental areas can be observed in Figures 2 and 3.

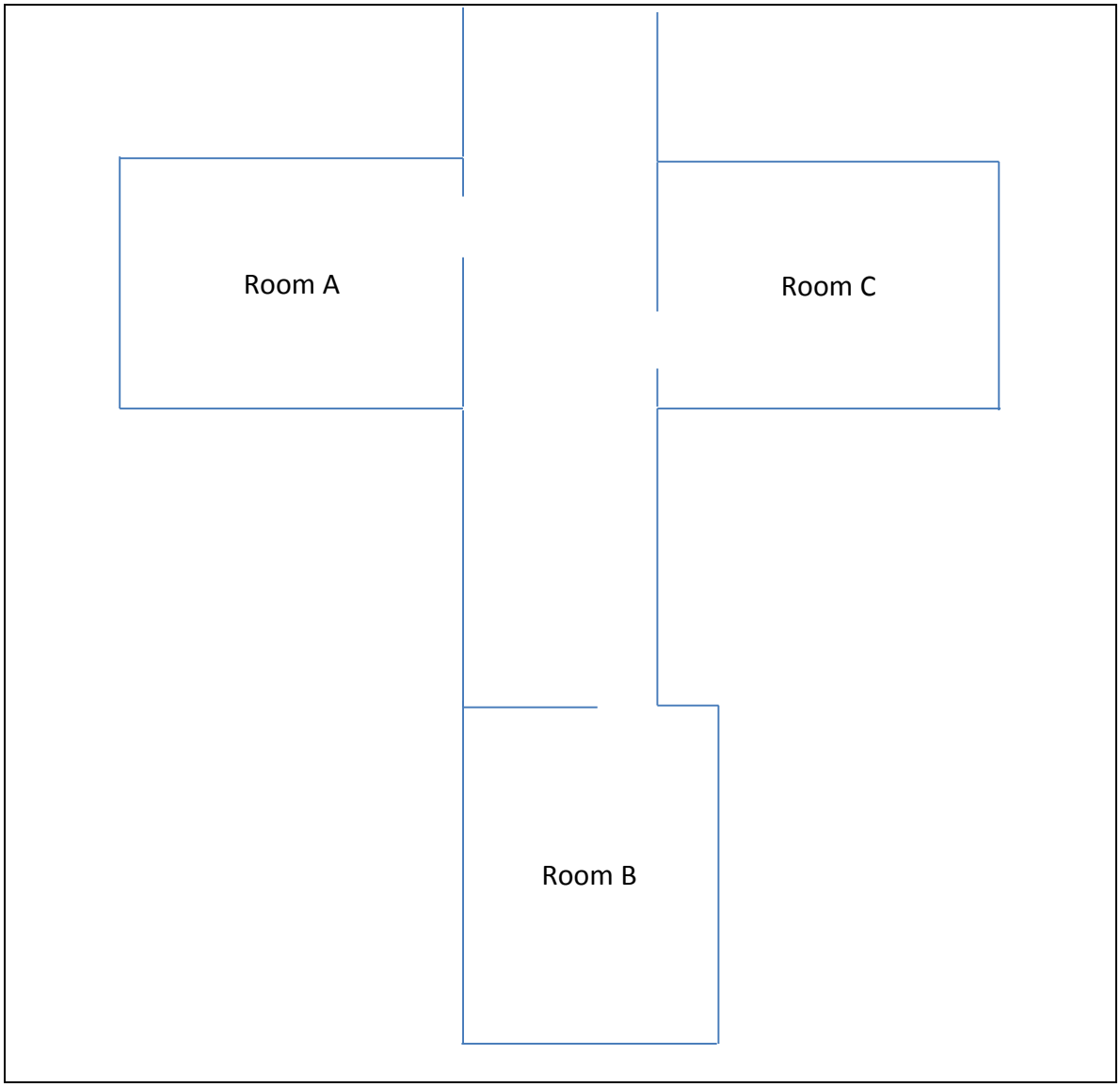


Figure 2. *Layout of experimental area.*

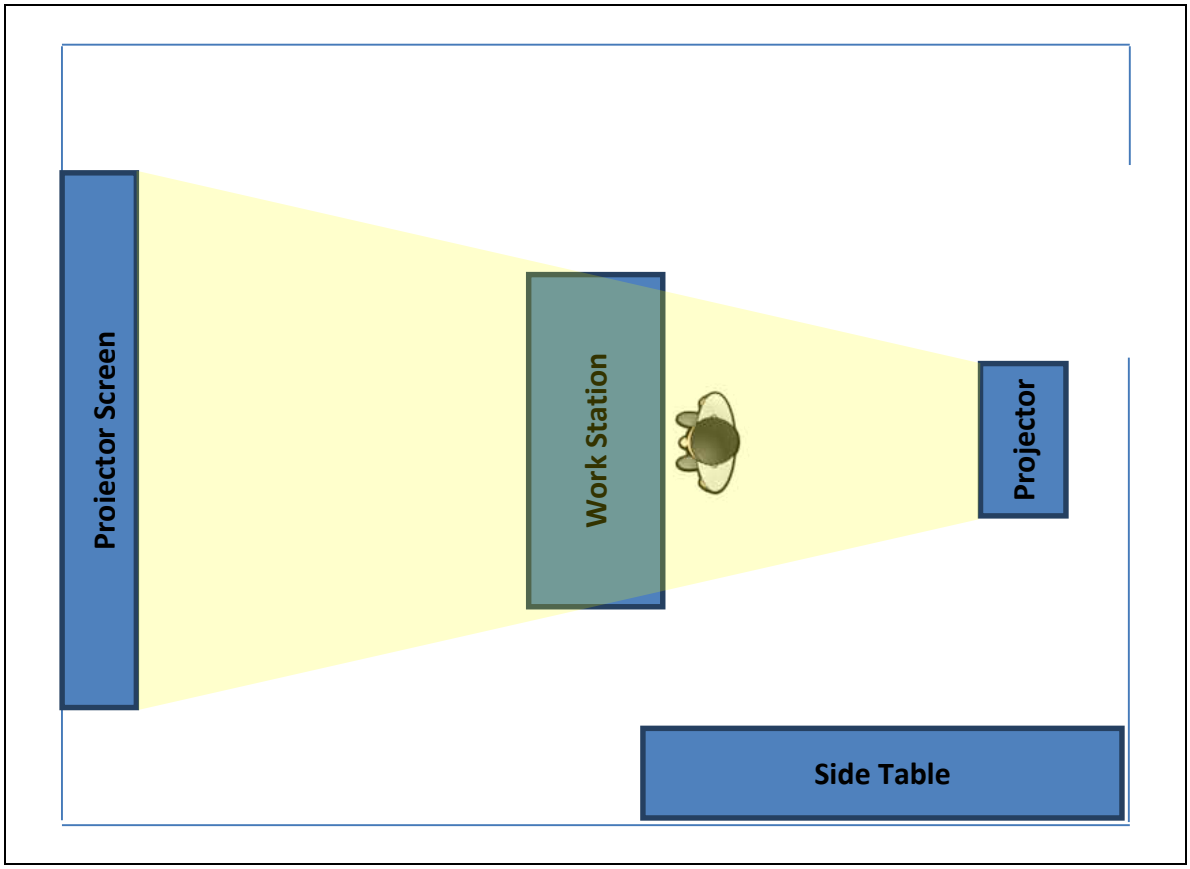


Figure 3. *Layout of Room A.*

TAC Condition. When the focal participant and his or her partner arrived at the study location, they were greeted by an experimenter acting as an employee of SUMMIT Hospital. The experimenter then escorted the focal participant to a specific room, Room A, where the experimenter sat the participant down at the side table and ask them to read over and sign an informed consent form. The focal participant was also asked to sit patiently and relax for two minutes after signing the form. The experimenter then left the room and escorted the partner to another room, Room B. The experimenter asked the partner to read over and sign the informed consent. After the partner signed the informed consent, the experimenter gave instructions to the partner to fill out a pre-hire paperwork packet, Packet #1, and then leave the room to monitor the participants in Room C. Packet #1 contains both the self-reported and partner-worded versions of

the personality measure, assertiveness measure, cognitive flexibility measure, and the social flexibility measure – it also contains a familiarity form. After the focal participant has relaxed for two minutes, the experimenter will escort the focal participant to the main workstation and provide instructions about completing the Wonderlic Cognitive Ability Measure and Packet #1. The experimenter will then monitor both participants' progress in Room C.

Since the focal participant has additional material to fill out (i.e., Wonderlic Cognitive Ability Measure), the partner is expected to finish their paperwork first. Once the partner finishes Packet #1, the experimenter will start them on a ninety minute presentation. This presentation will train the partner on the five targeted dimensions in the assessment center and will subsequently ask them questions about the degree to which the focal participant's behavior would reflect these behaviors in multiple scenarios. For the training, the partners will learn the difference between the five dimensions, be provided with a breakdown on the behaviors representing these dimensions, and listen to example responses for someone high and low on each dimension. They will also be provided with a dimension reference sheet, in case they need a reminder on the definitions of each dimension. They will first answer general questions, not in the hospital context, on how much they believe the focal participant represents each dimension. After answering these five questions, they will learn about the job the focal participant is applying for and the tasks they will go through. The partner will then rate the focal participant on each of the five dimensions within various scenarios taken directly from the AC. Specifically, the partner will go through five sets of ratings. Each set of ratings will focus on one of the five dimensions. Within each set, the will be asked to rate the focal participant on a specific dimension in the exact same scenarios the focal participant will experience. Before each set, the

partner will be told which dimension they will be assessing and the specific behaviors associated with that dimension.

After the partner starts on the presentation, the experimenter will monitor both of the participants' progress from Room C. At this point, the experimenter will wait for the focal participant to finish their paperwork. Once the focal participant finishes their paperwork, the experimenter will provide them a simple overview of the technological devices they will be using during the Assessment Center. These technological devices will be a wireless keyboard and mouse (which controls the screen on the projector), a wired keyboard and mouse (which controls the computer at the workstation), and wireless headphones. Once this overview is complete, the experimenter will begin the Welcome/Inbox Instructions. These instructions will provide an overall context of the hospital and provide participant instructions on how to perform the tasks within the inbox exercise. This instructional session will take approximately twenty minutes. After these instructions, the experimenter will start the inbox exercise, which will take exactly one hour.

Following the inbox exercise, the experimenter will move the focal participant to a different location for a short duration. The focal participant will then be asked to sit patiently as the experimenter sets up the next exercises. After approximately 3 minutes, the focal participant will begin the role play exercise instructions, which will take approximately twenty minutes. Once the instructions are completed, the focal participant will begin the role play exercise, which will take approximately forty-five minutes. During the role play exercise, the partner will likely finish their presentation. When this occurs, the experimenter will ask the partner to fill out Packet #2. Packet #2 contains the demographic measures. After Packet #2 is completed, the partner will be debriefed and released.

At this point, the focal participant should be wrapping up the role play exercise. Once the role play exercise is completed, the experimenter will move the participant to a different location again and will ask the participant to wait as they set up the next exercise. After the approximately 3 minute break, the focal participant will begin the interview exercise instructions, which will take approximately 5 minutes. After these instructions are completed, the focal participant will immediately begin the interview exercise, which will take approximately 30 minutes, 20 minutes of which will be allocated towards reviewing a packet of material. At the conclusion of the interview exercise, the participant will watch a presentation that reviews the three exercises they participated in and will be asked to fill out the Ability To Identify Criteria (ATIC) measure. After they complete the ATIC measure, they will fill out Packet #2. Once Packet #2 is completed they will be debriefed and free to go.

DITLAC Condition. The DITLAC condition will start off very similar to the TAC condition. In fact, the procedure for the partner will be exactly the same. Where the procedures begin to differ is after the focal participant completes the Packet #1. Instead of getting instructions for the inbox exercise, they will receive instructions for all three exercises at once. However, the instructions will not explicitly mention different exercises. The instructions will be presented in such a way that the focal participant will simply understand it as a series of tasks with no overarching group (i.e., exercise). This instructional session will last approximately 45 minutes. Once the focal participant completes these instructions, they will begin on a single 'exercise' that will last 135 minutes. This single exercise will consist of exactly the same tasks as the TAC condition, however, with a few minor adjustments as detailed in the experimental manipulation section.

After the focal participant completes the 135 minute performance period, the focal participant will watch a presentation that describes each of their tasks they just performed and sorts them into the three exercises from which they were pulled. After watching this presentation, the focal participant will fill out the ATIC measure and then Packet #2. They will be subsequently debriefed and released.

Structural Components Checks. In order to ensure that the casual mechanisms behind the results are, in fact, due to changes in structural components, the independent effects of each of the four structural components will be measured. First, in order check exercise integration, participants in both conditions will be asked to complete a measure asking them to report the dimensions they believed were being assessed in each exercise. Prior to filling out the measure, each condition will be reminded (TAC)/informed (DITLAC) which tasks belonged to which exercises. If the psychological mechanisms for exercise integration are working in the proposed ways, participants in the DITLAC condition should be more consistent in reporting dimensions across exercises than participants from the TAC condition. Second, the instructional characteristics manipulation check will take the form of blood pressure/heart rate checks immediately following. In the DITLAC condition, participants will have their blood pressure/heart rate taken after a brief resting period following the signing of the informed consent (base-line measures) and immediately after they complete their instructions. In addition to the base-line measurement, participants in the TAC condition will have their blood pressure measured immediately following each instructional session. It is expected that the difference between 'after-instructions' blood pressure/heart rate measures and the baseline measure will be larger in the DITALC condition than in the TAC condition. When individuals experience more stress in instructions, they are less likely to retain instructions into the work environment (Porras

& Hargis, 1982). Thus, the larger difference anticipated to be observed within the DITLAC condition will be an indicator that the cues within the instructions will have less of an influence on behaviors in the subsequent exercise.

Third, context consistency will be checked through comparing the relationship between same-task different-context (i.e., exercise) correlations and same-task same-context correlations. As mentioned earlier, the only difference between the two conditions are the manipulation of structural components. As such, no matter which condition, the participant will receive identical tasks. If context consistent is working in its intended way, the difference between same-task different-context correlations gleaned from the TAC condition will be weaker than the correlation between same-task same-context correlations found in the DITLAC condition. Lastly, the manipulation check for breaks will be determined through comparing baseline blood pressure/heart rate ratings to end-of-assessment center blood rate heart pressure. If breaks in the TAC condition create clear opportunities for the participants to rest and relax than the difference between their baseline and end-of-assessment physiological ratings should be smaller than the same difference found in the DITLAC condition.

CHAPTER 4: RESULTS

All formal tests of hypotheses were conduct in IBM SPSS 22. Means, standard deviations, and correlations among study variables are reported in Table 4. Table 5 represents these same relationships for each type of AC

External Convergent Validity

The first hypothesis sought to demonstrate that the relationship between external indicators (i.e., self-reported personality and typical behavior) and conceptually-relevant PEDRs was significantly stronger in the DITLAC than in the TAC. In the present context, stronger denotes both strength (slope) and how well (variance explained) external indicators predicted conceptually-relevant PEDRs. Both techniques can provide evidence of convergent validity.

Strength – Simple Effects Test

The first method I used to test differences in convergent validity was a multiple regression approach. Specifically, each PEDR was regressed on the set of conceptually relevant personality traits, typical behavioral ratings, condition, the interaction term of the each external factor and condition as well as cognitive ability and familiarity. Five different multiple regression analyses were run for each of the PEDRs as displayed in Tables 6-10. The significance of the interaction terms in each regression model indicates whether the external factors predicts the PEDR in a significantly different manner in each type of AC. As such, a more appropriate indicator would be to interpret the effect size of the personality trait within each condition. That is, the significance of the interaction term does not uniquely provide evidence of convergent validity. It must be inferred via a simple effects test. As such, the regression analyses were run using Model 1 in the Hayes (2013) macro for IBM SPSS in order to estimate the of each external factor's

predictive strength in each condition. The simple effects test reveals the strength of the relationship between an external factor and conceptually relevant PEDR for each type of AC. It should be noted that these tests do not directly compare the predictive strength of the personality trait in each condition. Rather, the significance of the simple effects test is based upon the strength of a predictor *within* a condition. Thus, this is a more liberal test of hypothesis 1 as evidence of construct validity is established within condition, but the predictive strength cannot be compared between conditions.

Communication PEDRs. For this analysis, extraversion, agreeableness, typical ratings of communication, type of AC, cognitive ability, familiarity, the interaction term between extraversion and type of AC, the interaction term between agreeableness and type of AC, and the interaction term between typical ratings of communication and type of AC were included in the regression model to predict communication PEDRs (Table 6). In partial support of hypothesis 1, the simple effects analysis revealed that agreeableness significantly predicted PEDRs of communication PEDRs in the DITLAC ($B = .29$, $p < .05$) but not in the TAC ($B = .18$, $p > .10$). However, extraversion did not significantly predict PEDRs of communication for either type of AC. Interestingly, typical ratings of communication significantly predict ratings of communication in the TAC ($B = .13$, $p < .05$), but not in the DITLAC $B = .03$, $p > .10$).

Table 4. Means, Standard Deviations, and Correlations between study variables.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1. Condition[a]																		
2. Cognitive Ability	.03																	
3. Familiarity[b]	.04	-.10																
4. Extraversion	-.03	-.08	-.07															
5. Neuroticism	-.08	.04	-.03	-.28**														
6. Agreeableness	-.02	-.21*	.12*	.25**	-.30**													
7. Conscientiousness	.01	-.15*	.13*	.30**	-.44**	.31**												
8. Assertiveness	.01	.03	-.04	.10	-.09	-.22**	.06											
9. COM[d] PEDRs[c]	.32**	.30**	-.11	.05	-.07	.05	-.02	-.02										
10. IO[e] PEDRs	-.03	.11	-.04	-.01	.01	.08	-.09	.08	.16**									
11. AGG[f] PEDRs	-.03	.03	.11	-.06	.03	-.10	-.06	.05	-.12*	.25**								
12. CON[g] PEDRs	.14*	.03	-.01	.13*	-.03	.06	.05	.02	.35**	.35**	.02							
13. PO[h] PEDRs	.01	.23**	.06	-.10	-.14*	.05	.06	-.06	.18**	.31**	.28**	.22**						
14. Typical COM	.01	.05	.01	.05	-.17**	-.04	.06	.03	.11	.15*	.08	.09	.06					
15. Typical IO	.00	.15*	-.01	.00	-.11	-.04	.02	-.01	.12*	.11	.10	.06	.12*	.40**				
16. Typical AGG	-.09	-.10	-.06	-.02	.11	-.20**	-.09	.09	-.08	.04	.03	-.04	-.06	-.31**	-.36**			
17. Typical CON	.02	.08	.00	.03	-.09	.19**	.10	-.06	.11	.12*	.04	.16**	.10	.43**	.49**	-.49**		
18. Typical PO	.04	.02	.08	-.02	-.08	.08	.16**	-.04	.17**	.06	-.10	.04	.07	.47**	.29**	-.25**	.41**	
Means	.50	21.95	35.77	3.87	2.57	3.94	3.98	4.20	2.98	3.20	1.33	2.39	2.11	4.82	4.38	2.13	4.82	5.10
SD	.50	4.61	56.32	0.53	0.64	0.52	0.58	0.66	0.81	0.86	0.37	0.60	0.85	0.85	0.93	0.88	0.81	0.80

Notes. ** $p < .01$, * $p < .05$

[a] Condition – 0 = Traditional Assessment Center; 1 = Day-in-the-Life

[b] Familiarity in months

[c] PEDRs = Post Exercise Dimension Ratings

[d] COM = Communication

[e] IO = Influencing Others

[f] AGG = Aggressiveness

[g] CON = Consideration

[h] PO = Planning and Organizing

Table 5. Means, Standard Deviations, and Correlations between study variables by Condition (TAC on above top diagonal, DITLAC below diagonal)

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1. Cognitive Ability			-.22**	-.08	.00	-.21*	-.14	-.06	.29**	.02	.06	-.01	.19*	.12	.30**	-.11	.15	.10
2. Familiarity[b]		.00		.10	-.19*	.18*	.21*	.01	-.07	.06	.11	.08	.10	.12	-.04	-.04	.02	.14
3. Extraversion		-.08	-.19*		-.18*	.19*	.21**	.11	.12	.04	-.10	.25**	-.13	.08	.01	-.02	.04	.03
4. Neuroticism		.08	.13	-.39**		-.32**	-.44**	-.12	-.03	.10	.00	.02	-.12	-.19*	-.17*	.07	-.03	-.05
5. Agreeableness		-.21*	.07	.31**	-.30**		.31**	-.21*	.02	.08	-.08	-.03	.09	.01	-.03	-.23**	.14	.06
6. Conscientiousness		-.15	.06	.39**	-.44**	.31**		.08	-.01	-.12	-.01	.05	-.01	.13	.03	-.10	.15	.21**
7. Assertiveness		.11	-.07	.10	-.07	-.23**	.04		.03	-.05	.01	.07	-.12	.08	.08	.04	-.02	.00
8. COM[d] PEDRs[c]		.33**	-.19*	.00	-.06	.10	-.03	-.07		.11	-.15	.32**	.10	.17*	.17*	-.04	.12	.20*
9. IO[e] PEDRs		.19*	-.12	-.05	-.09	.08	-.06	.18*	.26**		.30**	.25**	.29**	.08	.00	.09	-.01	-.04
10. AGG[f] PEDRs		.01	.12	-.02	.05	-.11	-.11	.09	-.08	.21*		.04	.30**	.07	.04	.06	.01	-.08
11. CON[g] PEDRs		.06	-.09	.04	-.07	.15	.04	-.02	.34**	.46**	.01		.21*	.16	.09	-.05	.16	.12
12. PO[h] PEDRs		.26**	.04	-.08	-.16	.01	.13	.00	.27**	.33**	.26**	.23**		.05	.08	-.01	.01	.02
13. Typical COM		-.02	-.09	.01	-.15	-.09	-.04	-.01	.04	.21*	.09	.02	.06		.41**	-.30**	.46**	.48**
14. Typical IO		.01	.01	.00	-.05	-.04	.01	-.08	.08	.20*	.15	.04	.16	.39**		-.28**	.47**	.23**
15. Typical AGG		-.08	-.08	-.03	.15	-.18*	-.08	.15	-.07	-.01	-.02	.01	-.13	-.31**	-.44**		-.42**	-.18*
16. Typical CON		.01	-.01	.02	-.15	.23**	.06	-.10	.10	.25**	.08	.17*	.19*	.40**	.51**	-.56**		.43**
17. Typical PO		-.06	.02	-.06	-.11	.10	.10	-.08	.12	.17*	-.13	-.06	.12	.45**	.34**	-.34**	.39**	
DITLAC	Means	22.09	37.81	3.85	2.52	3.93	3.99	4.21	3.24	3.17	1.31	2.47	2.12	4.83	4.37	2.05	4.83	5.13
	SD	4.72	62.25	0.56	0.60	0.52	0.58	0.70	0.76	0.90	0.38	0.62	0.84	0.80	0.97	0.82	0.82	0.77
TAC	Means	21.83	33.77	3.89	2.62	3.95	3.97	4.19	2.73	3.23	1.34	2.31	2.10	4.81	4.38	2.21	4.80	5.08
	SD	4.52	49.97	0.51	0.67	0.53	0.59	0.62	0.77	0.83	0.35	0.56	0.85	0.90	0.89	0.93	0.80	0.84

Notes. ** $p < .01$, * $p < .05$

[b] Familiarity in months

[c] PEDRs = Post Exercise Dimension Ratings

[d] COM = Communication

[e] IO = Influencing Others

[f] AGG = Aggressiveness

[g] CON = Consideration

[h] PO = Planning and Organizing

Table 6. Multiple Regression Analysis of Communication Post-Exercise Dimension Ratings and Simple Effects Tests.

	B	*SE*	*LLCI*	*ULCI*
Cognitive Ability	.05**	.01	.03	.07
Familiarity	-.01**	.00	.00	.00
Condition	1.48	.95	-.38	3.35
Agreeableness	.18	.12	-.05	.40
Extraversion	.17	.12	-.07	.41
Typical COM	.14*	.07	.00	.27
Agreeableness x Condition	.12	.17	-.21	.44
Extraversion x Condition	-.23	.17	-.57	.10
Typical COM x Condition	-.11	.10	-.31	.10

	B	*SE*	*LLCI*	*ULCI*
Agreeableness (TAC)	.18	.12	-.05	.40
Agreeableness (DITLAC)	.29*	.13	.05	.54
Extraversion (TAC)	.17	.12	-.07	.41
Extraversion (DITLAC)	-.06	.12	-.29	.17
Typical COM (TAC)	.14*	.07	.01	.27
Typical COM (DITLAC)	.03	.08	-.13	.18

Notes. ** $p < .01$, * $p < .05$
Condition – 0 = Traditional Assessment Center; 1 = Day-in-the-Life
Familiarity in months
PEDRs = Post Exercise Dimension Ratings
COM = Communication

Influencing Others PEDRs. Table 7 shows the regression analysis where influencing others was regressed onto assertiveness, agreeableness, extraversion, neuroticism, typical ratings of influencing others, cognitive ability, type of AC, and the interaction term of each of the external factors by type of AC. In partial support of hypothesis 1, the simple effects analysis revealed that assertiveness ($B = .31$, $p < .05$), agreeableness ($B = .34$, $p < .05$), and typical ratings of influencing others ($B = .19$, $p < .05$) all significantly predicted PEDRs of influencing others in the DITLAC. Further, none of these external indicators significantly predicted PEDRs of influencing others in the TAC. However, neither extraversion ($B = -.23$, $p > .05$) nor neuroticism ($B = -.12$, $p > .10$) significantly predicted PEDRs of influencing others in the DITLAC.

Table 7. Multiple Regression Analysis of Influencing Others Post-Exercise Dimension Ratings and Simple Effects Tests

	B	*SE*	*LLCI*	*ULCI*
Cognitive Ability	.02	.01	-.01	.04
Familiarity	-.01	.01	-.01	.01
Condition	-.90	1.64	-4.11	2.32
Agreeableness	.18	.15	-.11	.48
Extraversion	.12	.14	-.16	.40
Assertiveness	-.05	.12	-.28	.19
Neuroticism	.16	.12	-.06	.39
Typical IO	.02	.08	-.15	.18
Agreeableness x Condition	.16	.22	-.27	.58
Extraversion x Condition	-.35	.20	-.75	.05
Assertiveness x Condition	.36*	.16	.04	.68
Neuroticism x Condition	-.28	.18	-.63	.07
Typical IO x Condition	.17	.11	-.04	.39

	B	*SE*	*LLCI*	*ULCI*
Agreeableness (TAC)	.18	.15	-.11	.48
Agreeableness (DITLAC)	.34*	.16	.03	.65
Extraversion (TAC)	.12	.14	-.16	.40
Extraversion (DITLAC)	-.23	.15	-.51	.06
Assertiveness (TAC)	-.05	.12	-.28	.19
Assertiveness (DITLAC)	.31**	.11	.10	.52
Neuroticism (TAC)	.16	.12	-.06	.39
Neuroticism (DITLAC)	-.12	.13	-.38	.15
Typical IO (TAC)	.02	.08	-.15	.18
Typical IO (DITLAC)	.19**	.07	.05	.34

Notes. ** $p < .01$, * $p < .05$
Condition – 0 = Traditional Assessment Center; 1 = Day-in-the-Life
Familiarity in months
PEDRs = Post Exercise Dimension Ratings
IO = Influencing Others

Aggression PEDRs. Aggression PEDRs were regressed unto the conceptually relevant predictors of typical ratings of aggression, neuroticism, assertiveness, and agreeableness, which, along with cognitive ability, type of AC, and each of their interaction terms with type of AC (Table 8). Unfortunately, none of the external factors predicted PEDRs of aggressiveness in either type of AC.

Table 8. Multiple Regression Analysis of Aggressiveness Post-Exercise Dimension Ratings and Simple Effects Tests.

	B	*SE*	*LLCI*	*ULCI*
Cognitive Ability	.05	.01	-.01	.02
Familiarity	.01*	.00	.00	.02
Condition	-.13	.66	-1.42	1.16
Agreeableness	-.07	.07	-.01	.15
Assertiveness	-.01	.05	-.11	.11
Neuroticism	-.01	.05	-.11	.10
Typical AGG	.02	.04	-.05	.09
Agreeableness x Condition	-.02	.10	-.22	.16
Assertiveness x Condition	.06	.07	-.09	.20
Neuroticism x Condition	.04	.08	-.12	.19
Typical AGG x Condition	-.06	.05	-.16	.05

	B	*SE*	*LLCI*	*ULCI*
Agreeableness (TAC)	-.07	.07	-.20	.07
Agreeableness (DITLAC)	-.09	.07	-.23	.05
Assertiveness (TAC)	-.01	.05	-.11	.11
Assertiveness (DITLAC)	.05	.05	-.04	.15
Neuroticism (TAC)	-.01	.05	-.11	.10
Neuroticism (DITLAC)	.03	.06	-.08	.15
Typical AGG (TAC)	.02	.04	-.05	.09
Typical AGG (DITLAC)	-.04	.04	-.12	.04

Notes. ** $p < .01$, * $p < .05$

Consideration PEDRs. Agreeableness, neuroticism, typical ratings of consideration, type of AC, cognitive ability, the interaction term between agreeableness and type of AC, the interaction term between neuroticism and type of AC, and the interaction term between typical ratings of consideration and type of AC were used to predict PEDRs of consideration (Table 9). Agreeableness significantly predicted PEDRs of consideration in the DITLAC ($B = .17$, $p < .05$) but not in the TAC ($B = -.04$, $p > .10$). However, neuroticism did not predict PEDRs of consideration in either condition. Unexpectedly, typical ratings of consideration predicted consideration PEDRs in both the DITLAC ($B = .11$, $p < .10$) and TAC ($B = .11$, $p < .10$).

Table 9. Multiple Regression Analysis of Consideration Post-Exercise Dimension Ratings and Simple Effects Tests

	B	*SE*	*LLCI*	*ULCI*
Cognitive Ability	.01	.01	-.01	.02
Familiarity	-.01	.01	-.01	.01
Condition	-.68	.77	-2.21	.84
Agreeableness	-.04	.10	-.24	.16
Neuroticism	.00	.08	-.15	.15
Typical CON	.11	.06	-.01	.23
Agreeableness x Condition	.21	.14	-.07	.48
Neuroticism x Condition	-.01	.12	-.23	.23
Typical CON x Condition	.01	.09	-.16	.18

	B	*SE*	*LLCI*	*ULCI*
Agreeableness (TAC)	-.04	.10	-.24	.16
Agreeableness (DITLAC)	.17*	.10	.03	.37
Neuroticism (TAC)	.01	.08	-.15	.15
Neuroticism (DITLAC)	.01	.09	-.17	.17
Typical CON (TAC)	.11	.06	-.01	.23
Typical CON (DITLAC)	.11	.06	-.01	.23

Notes. ** $p < .01$, * $p < .05$

Planning and Organizing PEDRs. PEDRs of planning and organizing were regressed on conscientiousness, neuroticism, typing ratings of planning and organizing, type of AC, cognitive ability, and the interaction terms of the external factors by type of AC. Neuroticism (B = -.20, p = .12), and typical ratings of planning and organizing (B = .12, p = .18) were trending towards significance but did not reach statistical significance in the DITLAC (Table 10). However, none of the external factors predicted planning and organizing PEDRs in the TAC.

Table 10. Multiple Regression Analysis of Planning and Organizing Post-Exercise Dimension Ratings and Simple Effects Tests.

	B	*SE*	*LLCI*	*ULCI*
Cognitive Ability	.05*	.01	.03	.07
Familiarity	.01	.01	-.01	.01
Condition	-1.24	1.12	-3.47	.98
Conscientiousness	-.04	.13	-.30	.22
Neuroticism	-.13	.11	-.35	.10
Typical PO	-.01	.08	-.17	.15
Conscientiousness x Condition	.18	.19	-.18	.54
Neuroticism x Condition	-.07	.17	-.41	.27
Typical PO x Condition	.13	.12	-.11	.37

	B	*SE*	*LLCI*	*ULCI*
Conscientiousness (TAC)	-.04	.13	-.30	.22
Conscientiousness (DITLAC)	.14	.13	-.12	.40
Neuroticism (TAC)	-.13	.11	-.35	.09
Neuroticism (DITLAC)	-.20	.13	-.45	.05
Typical PO (TAC)	-.01	.08	-.17	.15
Typical PO (DITLAC)	.12	.09	-.06	.30

Notes. ** $p < .01$, * $p < .05$

Variance – ΔR^2 Comparison

The next approach I used to test Hypothesis 1 was to establish whether conceptually relevant external factors explained significantly more variance in PEDRs in the DITLAC in the TAC. As such, this was tested by comparing the ΔR^2 from the DITLAC to the same regression model in the TAC after controlling for cognitive ability and familiarity. That is, separate regression analyses were run for each condition. Within each analysis, the PEDR of interest was first regressed onto cognitive ability and familiarity in model 1. In model 2, the conceptually-related external factors were added to the analysis. The ΔR^2 between model 1 and model 2 represents the amount of variance that the set of external indicators uniquely explains in the PEDRs after accounting for cognitive ability and familiarity. Cohen, Cohen, West, and Aiken (2003) specified a method of comparing two R^2 between samples. Specifically, this technique involves using the combined standard errors from each sample in order to build confidence intervals around the difference between the R^2 from each sample. For instance, if ΔR^2 for

DITLAC was .42 (n = 150) and the ΔR^2 was .21 (n = 147) for TAC, the first step would involve calculating the squared standard error for each sample using the following formula:

$$SE_{R^2}^2 = \frac{4R^2(1-R^2)^2(n-k-1)^2}{(n^2-1)(n+3)} \quad (1)$$

where k represents the number of predictors. In the case that there were two conceptually related personality traits, the $SE_{R^2}^2$ for DITLAC would be:

$$SE_{R_{DITLAC}^2}^2 = \frac{4(.42)(1-.42)^2(150-2-1)^2}{(150^2-1)(150+3)} \quad (2)$$

$$SE_{R_{DITLAC}^2}^2 = \frac{(1.68)(.58)^2(147)^2}{(22499)(153)} \quad (3)$$

$$SE_{R_{DITLAC}^2}^2 = \frac{1221.37}{(22499)(153)} \quad (4)$$

$$SE_{R_{DITLAC}^2}^2 = .0035 \quad (5)$$

Following this same procedure for the TAC, the $SE_{R_{TAC}^2}^2$ = .0033. The next step would be to calculate the difference in SE^2 using the following equation:

$$SE_{R_{DITLAC}^2 - R_{TAC}^2} = \sqrt{SE_{R_{DITLAC}^2}^2 + SE_{R_{TAC}^2}^2} \quad (6)$$

$$SE_{R^2_{DITLAC}-R^2_{TAC}} = \sqrt{.0035 + .0033} = \sqrt{.0068} = .083 \quad (7)$$

This value can then be used to create confidence internals *(CI*) around the difference in the R^2s by multiplying it be a constant factor, which will depend on the desired degree of inclusion. Cohen et al., (2003) suggest that in order to build a 95% *CI* it is recommended to use a factor of 2. However, in order to give an approximate test of $\alpha = .05$, some have argued that using a 95% confident internal is too conservative. In fact, Payton, Greenstone, & Schenker (2003) suggested that using a 83% or 84% size for the intervals should be used when the *SEs* in the two samples are relatively equal. As such, this paper used a 1.4 constant factor, which should equate to building an approximately 83%/84% *CI* around the estimates. Thus, in our example, we would take the difference in R^2 (.21) and use the difference in the *SE* to build a *CI*s around this estimate using (.083)(1.4) = .1163. In other words, the lower-bound estimate (LCI) would be .093(.21-.1163) and the upper-bound estimate (HCI) would be .33 (.21+.1163). Since this the *CI*s do not include 0, it can argued that the set of predictors in DITLAC explain significantly more variance in the PEDR than the same predictors in the TAC. The results from these analyses using the current data are reported in Tables 11-15 – Summarized in Table 16.

Communication PEDRs. The addition of the three external factors (agreeableness, extraversion, and typical ratings of communication) significantly explained unique variance in Communication PEDRs over that of cognitive ability and familiarity ($\Delta R^2 = .04$, $F = 2.22$, $p < .10$) in the DITLAC. This additional variance, however, was not significantly different than the variance these same set of predictors accounted for in communication PEDRs for the TAC ($\Delta R^2 = .05$, $F = 2.72$, $p < .05$).

Influencing Others PEDRs. The set of four external indicators (assertiveness, extraversion, neuroticism, agreeableness, and typical ratings of influencing others) explained significant variance in influencing others PEDRs ($\Delta R^2 = .11$, $F = 3.66$, $p < .01$) over that of cognitive ability and familiarity. Additionally, this additional variance was significantly greater ($\Delta R^2_{diff} = .09$, LCI = .01, HCI = .16) than the same analysis for the TAC ($\Delta R^2 = .03$, $F = .81$, $p > .10$).

Table 11. Multiple Regression Analysis of Communication Post-Exercise Dimension Ratings by Condition with control variables, conceptually matched, and unmatched external predictors.

	DITLAC												TAC											
	Model 1				Model 2				Model 3				Model 1				Model 2				Model 3			
	B	*SE*	*LLCI*	*ULCI*	*B*	*SE*	*LLCI*	*ULCI*	*B*	*SE*	*LLCI*	*ULCI*	*B*	*SE*	*LLCI*	*ULCI*	*B*	*SE*	*LLCI*	*ULCI*	*B*	*SE*	*LLCI*	*ULCI*
Cognitive Ability	.05**	.01	.02	.07	.06**	.01	.03	.08	.06**	.01	.03	.09	.04**	.01	.01	.07	.04**	.01	.02	.07	.04**	.02	.01	.07
Familiarity	.00**	.00	.00	.00	.00**	.00	-.01	.00	.00**	.00	-.01	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
Agreeableness					.33**	.13	.08	.58	.27†	.15	-.02	.56					.15	.12	-.08	.39	.17	.14	-.11	.45
Extraversion					-.08	.12	-.31	.16	-.07	.13	-.33	.19					.17	.13	-.08	.41	.19	.13	-.07	.44
Typical COM					.03	.08	-.12	.18	-.03	.10	-.22	.16					.13†	.07	-.01	.27	.05	.09	-.13	.23
Neuroticism									-.06	.12	-.31	.19									.00	.11	-.22	.23
Conscientiousness									.02	.13	-.23	.27									-.11	.13	-.36	.13
Assertiveness									-.09	.09	-.28	.09									.04	.11	-.17	.26
Typical IO									.05	.08	-.11	.21									.02	.09	-.15	.19
Typical AGG									.08	.10	-.11	.27									.03	.08	-.12	.19
Typical CON									.03	.10	-.18	.23									.04	.10	-.16	.24
Typical PO									.09	.10	-.10	.28									.17‡	.09	-.01	.35
ΔR^2	.13				.04				.02				.07				.05				.03			
F	10.50**				5.64**				2.54**				5.39**				3.87**				2.00*			
ΔF	10.50**				2.22*				.44				5.39**				2.72*				.71			

Notes.

Table 12. Multiple Regression Analysis of Influencing Others Post-Exercise Dimension Ratings by Condition with control variables, conceptually matched, and unmatched external predictors.

	DITLAC												TAC											
	Model 1				Model 2				Model 3				Model 1				Model 2				Model 3			
	B	*SE*	*LLCI*	*ULCI*	*B*	*SE*	*LLCI*	*ULCI*	*B*	*SE*	*LLCI*	*ULCI*	*B*	*SE*	*LLCI*	*ULCI*	*B*	*SE*	*LLCI*	*ULCI*	*B*	*SE*	*LLCI*	*ULCI*
Cognitive Ability	.05**	.01	.02	.07	.06**	.01	.03	.08	.06**	.01	.03	.09	.01	.02	-.02	.04	.01	.02	-.02	.05	.02	.02	-.02	.05
Familiarity	.00**	.00	.00	.00	.00**	.00	-.01	.00	.00**	.00	-.01	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
Agreeableness					.33**	.13	.08	.58	.27‡	.15	-.02	.56					.13	.15	-.17	.43	.25	.16	-.07	.56
Extraversion					-.08	.12	-.31	.16	-.07	.13	-.33	.19					.11	.14	-.17	.38	.10	.14	-.18	.38
Neuroticism					.03	.08	-.12	.18	-.03	.10	-.22	.16					.18	.11	-.04	.41	.18	.12	-.07	.42
Assertiveness									-.06	.12	-.31	.19					-.06	.12	-.29	.18	-.05	.12	-.29	.18
Typical IO									.02	.13	-.23	.27					.03	.08	-.13	.19	.02	.09	-.17	.21
Conscientiousness									-.09	.09	-.28	.09									-.15	.14	-.42	.12
Typical AGG									.05	.08	-.11	.21									.14	.09	-.03	.31
Typical CON									.08	.10	-.11	.27									.00	.11	-.22	.22
Typical PO									.03	.10	-.18	.23									-.09	.10	-.29	.10
Typical COM									.09	.10	-.10	.28									.17‡	.10	-.02	.36
ΔR^2	.04				.11				.08				.01				.03				.05			
F	3.20*				3.62**				3.31**				.53				.73				1.01			
ΔF	3.20*				3.66**				2.59*				.53				.81				1.39			

Notes. ** $p < .01$, * $p < .05$

Table 13. Multiple Regression Analysis of Aggressiveness Post-Exercise Dimension Ratings by Condition with control variables, conceptually matched, and unmatched external predictors.

	DITLAC												TAC											
	Model 1				Model 2				Model 3				Model 1				Model 2				Model 3			
	B	*SE*	*LLCI*	*ULCI*	*B*	*SE*	*LLCI*	*ULCI*	*B*	*SE*	*LLCI*	*ULCI*	*B*	*SE*	*LLCI*	*ULCI*	*B*	*SE*	*LLCI*	*ULCI*	*B*	*SE*	*LLCI*	*ULCI*
Cognitive Ability	.01	.01	-.01	.02	.00	.01	-.01	.02	.00	.01	-.01	.02	.01	.01	.00	.02	.01	.01	-.01	.02	.01	.01	-.01	.02
Familiarity	.00*	.00	.00	.00	.00*	.00	.00	.00	.00*	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
Agreeableness					-.10	.08	-.25	.06	-.04	.08	-.21	.13					-.06	.06	-.18	.07	-.04	.07	-.17	.09
Neuroticism					.03	.06	-.10	.15	.00	.07	-.14	.14					-.01	.05	-.10	.09	.00	.05	-.10	.10
Assertiveness					.06	.05	-.05	.16	.06	.05	-.04	.17					.00	.05	-.10	.10	.00	.05	-.10	.10
Typical AGG					-.04	.05	-.13	.05	.00	.06	-.11	.11					.02	.03	-.05	.08	.03	.04	-.04	.11
Extraversion									.02	.08	-.12	.17									-.08	.06	-.20	.04
Conscientiousness									-.13	.07	-.27	.02									.02	.06	-.10	.13
Typical IO									.06	.05	-.03	.15									.00	.04	-.08	.08
Typical CON									.02	.06	-.10	.13									.03	.05	-.06	.12
Typical PO									-.09	.05	-.20	.02									-.07	.04	-.15	.02
Typical COM									.06	.05	-.05	.17									.04	.04	-.04	.13
ΔR^2	.04				.03				.06				.00				.03				.06			
F	2.56[‡]				1.54				1.58				1.13				.64				.70			
ΔF	2.56[‡]				1.04				1.59				1.13				.41				.77			

Notes. ** $p < .01$, * $p < .05$

Table 14. Multiple Regression Analysis of Consideration Post-Exercise Dimension Ratings by Condition with control variables, conceptually matched, and unmatched external predictors.

	DITLAC												TAC											
	Model 1				Model 2				Model 3				Model 1				Model 2				Model 3			
	B	*SE*	*LLCI*	*ULCI*	*B*	*SE*	*LLCI*	*ULCI*	*B*	*SE*	*LLCI*	*ULCI*	*B*	*SE*	*LLCI*	*ULCI*	*B*	*SE*	*LLCI*	*ULCI*	*B*	*SE*	*LLCI*	*ULCI*
Cognitive Ability	.01	.01	-.02	.03	.01	.01	-.01	.03	.01	.01	-.01	.03	.00	.01	-.02	.02	.00	.01	-.03	.02	.00	.01	-.02	.02
Familiarity	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
Agreeableness					.17	.11	-.05	.38	.19	.12	-.05	.44					-.07	.10	-.27	.12	-.08	.11	-.30	.13
Neuroticism					.01	.09	-.17	.19	.00	.10	-.20	.21					.01	.07	-.14	.16	.05	.08	-.12	.21
Typical CON					.11[‡]	.06	-.02	.23	.19*	.09	.01	.36					.12*	.06	.00	.23	.10	.08	-.05	.25
Extraversion									-.05	.11	-.27	.17									.24*	.10	.05	.43
Conscientiousness									.04	.11	-.17	.25									-.01	.10	-.19	.18
Assertiveness									.00	.08	-.16	.15									.05	.08	-.11	.21
Typical AGG									.11	.08	-.05	.27									.00	.06	-.12	.11
Typical IO									.01	.07	-.12	.14									-.02	.06	-.15	.11
Typical PO									-.12	.08	-.27	.04									.01	.07	-.12	.14
Typical COM									.04	.08	-.12	.20									.03	.07	-.10	.16
ΔR^2	.01				.05				.03				.00				.03				.05			
F	.91				1.76				1.10				.04				.84				1.03			
ΔF	.91				2.31[‡]				.65				.04				1.37				1.16			

Notes. ** $p < .01$, * $p < .05$

Table 15. Multiple Regression Analysis of Planning and Organizing Post-Exercise Dimension Ratings by Condition with control variables, conceptually matched, and unmatched external predictors.

	DITLAC												TAC											
	Model 1				Model 2				Model 3				Model 1				Model 2				Model 3			
	B	*SE*	*LLCI*	*ULCI*	*B*	*SE*	*LLCI*	*ULCI*	*B*	*SE*	*LLCI*	*ULCI*	*B*	*SE*	*LLCI*	*ULCI*	*B*	*SE*	*LLCI*	*ULCI*	*B*	*SE*	*LLCI*	*ULCI*
Cognitive Ability	.05**	.01	.02	.08	.05**	.01	.02	.08	.05**	.01	.02	.08	.04**	.02	.01	.07	.04**	.02	.01	.07	.04**	.02	.01	.08
Familiarity	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00*	.00	.00	.01	.00*	.00	.00	.01	.00	.00	.00	.01
Conscientiousness					.15	.13	-.11	.41	.23	.14	-.05	.50					-.06	.14	-.33	.22	-.03	.14	-.31	.25
Neuroticism					-.19	.13	-.44	.06	-.24	.13	-.50	.03					-.12	.12	-.35	.12	-.08	.13	-.34	.17
Typical PO					.12	.09	-.06	.29	.04	.10	-.17	.24					-.02	.08	-.19	.15	-.04	.10	-.24	.15
Extraversion									-.28	.14	-.56	.01									-.20	.15	-.49	.09
Assertiveness									-.01	.10	-.21	.19									-.14	.12	-.38	.10
Agreeableness									.01	.16	-.31	.32									.18	.16	-.14	.49
Typical AGG									.04	.10	-.17	.25									.06	.09	-.11	.24
Typical IO									.09	.09	-.08	.26									.05	.10	-.14	.25
Typical COM									-.04	.10	-.25	.16									.05	.10	-.15	.25
Typical CON									.13	.11	-.10	.35									-.03	.12	-.26	.20
ΔR^2	.07				.06				.05				.07				.01				.04			
F	5.33**				4.00**				2.31**				5.30**				2.30*				1.41			
ΔF	5.33**				2.96*				1.09				5.30**				.35				.79			

Notes. ** $p < .01$, * $p < .05$

Aggression PEDRs. Adding the external factors of aggression to the model (neuroticism, agreeableness, assertiveness, typical ratings of aggression) did not account for unique variance over that of cognitive ability and familiarity in either the DITLAC (ΔR^2 = .03, F = 1.04, p > .10) or the TAC (ΔR^2 = .01, F = .01, p < .10).

Consideration PEDRs. When neuroticism, agreeableness, and typical ratings of consideration were added to the regression analysis, it did account for unique variance over cognitive ability and familiarity (ΔR^2 = .05, F = 2.31, p < .05) in the DITLAC. Additionally, these same external factors did not account for significant variance in the TAC (ΔR^2 = .03, F = 1.37, p > .10). This difference, however, was not significantly different (ΔR^2_{diff}= .02, LCI = -.03, HCI = .07).

Planning and Organizing PEDRs. Conscientiousness, neuroticism, and typical ratings of planning and organizing significantly explained additional variance in PEDRs of planning and organizing over cognitive ability and familiarity in the DITLAC (ΔR^2 = .06, F = 2.96, p < .05), but not in the TAC (ΔR^2 = .01, F = .35, p > .10). However, though close, the difference between the two ΔR^2 values was not statistically significant (ΔR^2_{diff}= .05, LCI = -.01, HCI = .10).

In sum, aside from aggression PEDRs, there was evidence for convergent validity within the DITLAC. However, statistical tests indicated that this evidence was significantly stronger for influencing others PEDRs. As such, hypothesis 1 is partially supported.

Table 16. ΔR^2 Convergent Validity Test between different types of ACs

	ΔR^2_{DITLAC}	$SE^2_{R^2_{DITLAC}}$	ΔR^2_{TAC}	$SE^2_{R^2_{TAC}}$	ΔR^2_{DIFF}	*LLCI*	*ULCI*
Communication	.04	.00	.05	.00	-.01	-.08	.05
Influencing Others	.11	.00	.03	.00	.09*	.01	.16
Aggressiveness	.03	.00	.01	.00	.02	-.03	.06
Consideration	.05	.00	.03	.00	.02	-.03	.07
Planning and Organizing	.06	.00	.01	.00	.05	-.01	.10

Notes. ** p < .01, * p < .05

External Discriminant Validity

Hypothesis 2 argued that the discriminant validity evidence in the DITLAC would be stronger than the evidence found in the TAC. This hypothesis was tested by comparing the amount of variance that conceptually relevant variables accounted for in PEDRs to that of conceptually irrelevant variables. Specifically, I tested to see if the conceptually irrelevant variables explained significant variance over the conceptually relevant variance both within and across condition using multiple regression. The test for this hypothesis was an extended analysis of hypothesis 1. Like in hypothesis 1, the first regression model contained cognitive ability and familiarity and the conceptually relevant variables were added in model 2. In the third model, the conceptually irrelevant variables were added. This was done five times (one for each PEDR) for each condition, resulting in ten regression analyses (See Tables 11-15. Evidence for discriminant validity was revealed in two ways. First, there would be evidence if the set of conceptually irrelevant variables did not explain significant variance in the PEDR over that of the conceptually relevant variables. This evidence would be established within types of AC. Second, this additional variance would be then be compared across ACs. If the conceptually irrelevant variables explained more variance in the PEDRs in the TAC than in the DITLAC, it can be argued that the discriminant validity evidence in DITLAC is stronger. For these analyses, conceptually irrelevant variables were defined as any personality characteristics and typical behavior measurements that were not conceptually relevant to the PEDR being analyzed. The analyses for this hypothesis are displayed on Table 17.

Table 17. ΔR^2 Discriminant Validity Test between different types of ACs

	ΔR^2_{DITLAC}	$SE^2_{R^2_{DITLAC}}$	ΔR^2_{TAC}	$SE^2_{R^2_{TAC}}$	ΔR^2_{DIFF}	*LLCI*	*ULCI*
Communication	.02	.00	.03	.00	-.01	-.06	.03
Influencing Others	.08	.00	.05	.00	.03	-.05	.10
Aggressiveness	.06	.00	.03	.00	.03	-.03	.10
Consideration	.03	.00	.05	.00	.01	-.05	.05
Planning and Organizing	.05	.00	.04	.00	.02	-.04	.09

Notes. ** $p < .01$, * $p < .05$

Communication PEDRs. Adding the conceptually irrelevant variables to the regression analysis predicting communication PEDRs did not significantly account for additional variance in either the DITLAC ($\Delta R^2 = .02$, $F = .44$, $p > .10$) or the TAC ($\Delta R^2 = .03$, $F = .71$, $p > .10$). In addition, these ΔR^2 values were not statistically different (ΔR^2_{diff}= -.01, LCI = -.06, HCI = .03).

Influencing Others PEDRs. The conceptually irrelevant variables did account for unique variance in PEDRs of influencing others ($\Delta R^2 = .08$, $F = 2.59$, $p < .05$) in the DITLAC, but not in the TAC ($\Delta R^2 = .05$, $F = 1.39$, $p > .10$). However, the comparison of ΔR^2 values revealed that they were not significantly different (ΔR^2_{diff}= .03, LCI = -.05, HCI = .10).

Aggression PEDRs. When the conceptually irrelevant variables were added to the regression analysis predicting aggression PEDRs, they did not account for significant unique variance over that of conceptually relevant variables, cognitive ability and familiarity in either the DITLAC ($\Delta R^2 = .06$, $F = 1.58$, $p > .10$) or the TAC ($\Delta R^2 = .03$, $F = .77$, $p > .10$). Furthermore, the two ΔR^2 values were not significantly different (ΔR^2_{diff}= .03, LCI = -.03, HCI = .10).

Consideration PEDRs. The conceptually irrelevant variables did not significantly account for unique variance in PEDRs of consideration in the DITLAC ($\Delta R^2 = .03$, $F = .65$, $p > .10$) nor

the TAC (ΔR^2 = .05, F = 1.16, p > .10). When compared, the two ΔR^2 did not significantly differ (ΔR^2_{diff}= .01, LCI = -.05, HCI = .05).

Planning and organizing PEDRs. Lastly, conceptually irrelevant variables did not account additional variance in PEDRs of planning and organizing for either type of AC (DITLAC: ΔR^2 = .05, F = 1.01, p > .10; TAC: ΔR^2 = .04, F = .79, p > .10) and there was not a significant difference in the amount of variance that the set of conceptually irrelevant predictors accounted for in PEDRs of planning and organizing (ΔR^2_{diff}= .02, LCI = -.04, HCI = .09).

In sum, conceptually irrelevant external factors only accounted for unique variance in PEDRs for influencing others in the DITLAC. However, this additional variance was not significantly more than the variance conceptually irrelevant external factors accounted for in influencing others PEDRs in the TAC. For all other PEDRs, conceptually irrelevant external factors did not account for unique variance over that of conceptually relevant external factors, cognitive ability, or familiarity.

Internal Convergent Validity

A summary of the intercorrelations between exercise ratings is given in Table 18. Hypothesis 3 stated that the internal convergent validity of PEDRs (i.e., same-dimension different exercise correlations) would be stronger for the DITLAC than for the TAC. In order to test this hypothesis, same-dimension different-exercise (SDDE) correlations were averaged for the two conditions and then run through *Fisher's r to z transformation*. The average SDDE correlation from DITLAC ($\bar{r}$ = .25) was not significantly different (z = .23, p > .10) than the average SDDE correlations from the TAC ($\bar{r}$ = .22). Next, these comparisons were conducted for each exercise. These results are displayed in Tables 18. Unfortunately, in only two comparisons out of the fifteen the internal convergent validity for ratings obtained in the DITLAC was

stronger than for ratings obtained in the TAC. Specifically, the relationship between influencing others PEDRs from the inbox exercise and the interview exercise in the DITLAC ($r = .30$, $p < .01$) was significantly stronger than the same relationship in the TAC ($r = .04$, $p > .10$; $z = 1.96$, $p < .01$) and the relationship between influencing others from the role play exercise and the interview exercises was significant stronger in the DITLAC ($r = .55$, $p < .01$) than in the TAC ($r = .20$, $p < .05$; $z = 2.29$, $p < .05$).

Internal Discriminant Validity

Finally, hypothesis 4 sought to demonstrate that the internal discriminant validity evidence in the TAC was significantly weaker than in the DITLAC. In order to test this hypothesis, the different dimension same exercise (DDSE) correlations were calculated for both types of ACs and then compared to one another. Thus, the DITLAC would be shown to have stronger internal discriminant validity than the TAC if the DDSE correlations in the TAC were higher than they were in the DITLAC. As such, *Fisher's r to z transformation* was used to compare the DDSE correlations directly to one another. As can be seen in Tables 18, none of the comparisons supported hypothesis 6. That is, the average DDSE correlation in the DITLAC ($\bar{r} = .13$) was not significantly different than the average DDSE correlation from the TAC ($\bar{r} = .25$). This was also true when the comparisons were made within exercises as well.

Table 18. Correlations between PEDRs by exercise (TAC on above top diagonal, DITLAC below diagonal)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1. CM-IB		.47**	.30**	-.09	-.11	.07	-.03	-.18	-.17	.01	.21*	.00	-.01	-.07	-.11
2. CM-RP	.41**		.43**	-.14	.12	.39**	-.19*	-.24**	-.06	.07	.33**	.09	-.03	.05	.07
3. CM-IN	.41**	.52**		.03	.18*	.35**	-.15	-.02	.04	.03	.21*	.07	.17	-.03	.10
4. IO-IB	.15	-.03	.07		.28**	.02	-.01	.17	.20*	.43**	-.04	.11	.09	.15	.08
5. IO-RP	.06	.33**	.27**	.31**		.30**	.12	.19*	.28**	.11	.15	.25**	.06	.18*	.25**
6. IO-IN	.21*	.27**	.26**	.31**	.57**		-.11	-.07	-.02	.11	.24**	.08	.15	-.02	.31**
7. AGG-IB	-.21*	-.26**	-.22**	.02	.09	-.02		.23**	.16	-.18	-.15	.18	.01	.02	-.12
8. AGG-RP	-.02	-.15	.05	.19*	.16*	.01	.22**		.63**	-.07	-.03	-.04	.26**	.07	.20*
9. AGG-IN	.00	-.07	.10	.10	.07	.10	.18*	.60**		-.02	.10	.10	.13	.15	.24**
10. CON-IB	.06	.17	.17	.34**	.23*	.31**	-.30**	.09	.18		.06	.11	.16	.20	.08
11. CON-RP	.11	.31**	.25**	.03	.44**	.26**	.12	-.09	-.18*	.08		.19*	-.02	.17*	.16
12. CON-IN	-.04	.09	.11	.19*	.43**	.23**	.13	.21*	.09	-.04	.38**		-.07	.02	.03
13. PO-IB	.36**	.02	.18	.09	.17	.03	-.08	.12	.05	.01	.06	-.04		.06	.26*
14. PO-RP	.05	.12	.18*	.02	.13	.15	.08	.14	.05	.04	.17*	-.03	.01		.14
15. PO-IN	.18*	.16	.28**	.28**	.27**	.29**	.06	.17*	.18*	.35**	.35**	.13	.31**	.14	
Means	3.29	3.32	3.10	2.71	3.21	4.12	1.11	1.35	1.43	2.64	2.39	2.62	2.52	1.66	2.35
SD	1.24	.83	1.00	1.18	1.05	1.29	.30	.45	.85	.72	.69	1.00	1.26	1.21	1.14
Means	1.94	2.90	2.73	2.60	3.30	4.54	1.06	1.40	1.34	2.64	2.28	2.32	2.46	1.79	2.47
SD	1.04	.88	1.02	1.14	.96	1.03	.13	.45	.55	.84	.64	.92	1.25	1.20	1.09

Notes. ** $p < .01$, * $p < .05$
Familiarity in months
PEDRs = Post Exercise Dimension Ratings
COM = Communication
IO = Influencing Others
AGG = Aggressiveness
CON = Consideration
PO = Planning and Organizing
IB = Inbox
RP = Role Play
IN = Interview

CHAPTER 5: DISCUSSION

The purpose of the current study was to demonstrate that a day-in-the-life AC (DITLAC) would demonstrate better construct validity evidence than a traditional AC (TAC). The basic premise behind this assertion is that when formatted in a traditional manner, assessment centers are not eliciting typical performance from job candidates. TACs do not represent the structure of how work is conducted everyday. In fact, I argued that the structure of TACs actually created a situation that elicited maximum performance for job candidates. In contrast, DITLACs are designed to represent typical performance situations, which, in turn, should produce better construct validity ratings. Construct validity was tested in two ways, using both an external construct validity (i.e., nomological network) approach and an internal construct validity (i.e., MTMM Matrix) approach.

External Convergent Validity

For the first hypothesis, support depended on the PEDR of interest. There was no support for hypothesis 1 or 2 for the PEDR of aggression. In fact, there was no evidence of convergent or discriminant validity in either the DITLAC or TAC for aggression. One explanation concerns the range restriction surrounding aggressive behaviors. Aggressiveness is a highly undesirable as a well as a highly transparent behavior. That is, it is very clear when someone is being actively aggressive and it has a low base rate of occurring in the workplace (Mitchell & Ambrose, 2007). In fact, both within and across condition, aggressiveness PEDRs had the lowest *SD* out of all PEDRs. In support of this explanation, there was no evidence of external convergent validity in either the DITLAC or the TAC.

The convergent validity results surrounding communication are very mixed. Both types of AC demonstrated some degree of convergent validity evidence. In the DITLAC, agreeableness significantly predicted PEDRs of communication, whereas in the TAC agreeableness did not. However, typical ratings of communication significantly predicted PEDRs of communication in the TAC, but not in the DITLAC. This finding could be explained by the relationship between the familiar peers and the focal participant. Many of the partners knew the focal participant from class and, likely, their observation of professional communication is limited to classroom settings. Given the factors in this manuscript that purport to create strong situations, it is possible that partners understood the peer's communication behavior strictly in the highly-structured and formalized educational context. As this type of structure strongly represents that of the TAC, it makes sense that typical ratings of communication demonstrated a strong relationship with PEDRs in the TAC and not the DITLAC. It follows then that the variance explained in PEDRs by personality factors in the DITLAC and typical ratings of communication in the TAC would cause the R^2 between conditions to be relatively equal.

The support for consideration PEDRs was also mixed. DITLAC demonstrated strong construct validity evidence for the PEDR of consideration. Even though neuroticism fell through, both agreeableness and typical ratings of consideration showed a strong relationship with consideration ratings glean from the DITLAC. Furthermore, the set of external factors accounted for until variance in communication PEDRs over cognitive ability and familiarity. Though this latter instance was not true of TAC, typical ratings of consideration did significantly predict PEDRs of consideration. Typical ratings of performance were used in this study as an external indicator of what the participant would on the job. As such, I heavily relied on the assumption that focal participant's peers would have a decent understanding of the focal participant's typical

behavior. However, it could be that partners only have a relationship with the focal participant under certain situations and their ratings would be skewed depending on the situations to which they have been exposed. Thus, the validity of partners' ratings may be contingent upon the setting in which they have observed the focal participant.

I ran supplementary analyses to test this assumption. Specifically, in the TAC, the relationship between typical ratings of consideration and PEDRs of consideration grew stronger the more the exposure the peer had to the focal participant in maximum performance situations (indicated by interactions with authority figure at work). This pattern was exactly the opposite for the DITLAC. The more familiarity the peer had with the focal participant in maximum performance situations, the weaker the relationship between PEDRs of consideration and typical consideration ratings. Furthermore, the relationship between PEDRs of consideration and typical performance ratings became increasingly stronger the more the partner witnessed the focal participant behaving in weak situations (socializing with friends) in the DITLAC, while this relationship was not contingent upon context familiarity in the TAC. These supplement analysis demonstrate the validity of performance ratings is sometimes dependent on context. This is an area of future research, which is elaborated upon later).

All three of the conceptually relevant external factors related to planning and organizing were bordering on significance in predicting PEDRs. Since these effects are all relatively small in magnitude, sample size is a clear reason why these effects did no reach significance. However, when compared to the relatively poor effect sizes from the TAC, I argue that this difference is still meaningful and can be interpreted as support for hypothesis 1. This is supplemented by the fact that the difference in the variance these factors explained in PEDRs between types of ACs was also almost significant. Like the effect sizes, the R^2 values being dealt with in this study are

incredibly small because I controlled for both cognitive ability and familiarity. Coupled with the knowledge that the external factors did account for significance variance in PEDRs of planning and organizing over that of the control variables demonstrates that there is value to the DITLAC structure.

Finally, the clearest results were influencing others PEDRs. Even though extraversion and neuroticism did not significantly predict PEDRs of influencing others, three of the factors did so with relatively large effect sizes. In fact, it could be because these three factors accounted for so much variance in the PEDRs, there was not much left over for extraversion and neuroticism to explain. Additionally, results supported hypothesis 1 in that these external factors explained significantly more variance in PEDRs in the DITLAC than in the TAC.

Overall, though there were a few inconsistencies, the majority of the results were in line with expectations. DITLAC demonstrated better convergent validity evidence than the TAC. Even when the differences in effect sizes were not significantly different, results showed trends that were in line the hypothesized effects.

External Discriminant Validity Evidence.

Hypothesis 2 posited that conceptually irrelevant external factors would show stronger relationships with PEDRs in TAC than in the DITLAC. However, none of the tests demonstrated this. This is not to say that no discriminant validity evidence was found in either condition. Indeed, the fact that conceptually irrelevant factors did not account for unique variance in four out of the five PEDRs for the DITLAC suggests that conceptually irrelevant factors are not related to behaviors that should not be related to. The one instance where conceptually irrelevant factors did predict PEDRs was for influencing others. Breaking down the results, this was largely due to the typical ratings of conceptually irrelevant variables. Looking at the individual effect

sizes, typical ratings of aggressiveness and consideration accounted for significant variance in PEDRs of influencing others. The behaviors that represent influencing others (e.g., persuasiveness, stranding up for beliefs, remaining calm and collected in stressful situations) have also been used to describe aggressive (e.g., directly refusing requests; Baron, Neuman, & Geddes, 1999) and considerate (e.g., awareness of social environment, confrontation; Arthur et al., 2003) behaviors. As such, it is not entirely unexpected that these behavioral ratings predict one another. Even with this contrasting finding, conceptually relevant external factors accounted for more variance in PEDRs of influencing others than the conceptually irrelevant factors. In fact, this was true of all five PEDRs within DITLAC.

Internal Convergent Validity Evidence.

Though ratings from the DITLAC did not demonstrate significantly stronger convergent validity than ratings from the TAC, this is not to suggest that there was no convergent validity evidence in DITLAC. In 11 out of the 15 correlations comparisons from the DITLAC, there was a relatively strong magnitude between same dimension different exercise correlations. The only two dimensions where convergent validity evidence was lacking was for consideration and planning and organizing. The lack of convergent validity in these ratings highlights the issue of situationally specific behavior and the trait activation potential of each of exercise. Lievens et al (2006) suggest that exercises vary in their ability to tap certain traits. For instance, conscientiousness is best tapped by the inbox exercise and, further, conscientiousness is highly linked to the dimension of planning and organizing. Thus, there may have been more and different opportunities to express planning and organizing behaviors in the inbox exercise compared to that of the role play or interview exercises. The lack of convergent validity evidence for considerate behavior could have been an artifact of how interpersonal behaviors were assess

in each of the exercises. Opportunities to assess considerate behaviors requires a degree of interpersonal communication, which different in the in-basket, role play, and interview exercises. For the inbox and role play exercises, opportunities to engage in considerate behaviors were directly towards the recipient (e.g., customer, coworker) of the considerate behaviors, while considerate behaviors were directed towards a third-party in the interview exercise. Further, the medium in which the considerate behaviors were express different between the inbox and role-play exercises. Specifically, considerate behaviors were exclusively delivered via written communication in the inbox exercise and exclusively oral communication in the role-play exercise. Thus, these differences could have had an impact on the degree to which considerate behaviors were elicited and how in what form they manifested.

Internal Discriminant Validity Evidence.

There was not much evidence for internal discriminant validity in either the TAC or the DITLAC. I believe there are two main reasons for this. First, the way in which the analysis was conducted does not account for the potential conceptual similarities between PEDRs nor the degree to which they are both tapped in that same exercise. For instance, as mentioned earlier, there is a degree of overlap between the behaviors that represent influencing others, aggressiveness, and consideration. When these behaviors are not properly delineated in a rater's mind, it is likely that there will be a strong relationship between these different dimensions. This problem was exacerbated in the current study as a different set of raters were assigned to each PEDR. The proposed benefits having a unique set of raters for each PEDR is that the raters would only have to hold one concept in their head, lessening the cognitive burden. However, this approach had an unintended downside. It is possible that when a rater needs to rate more than one dimension, a single behavior exhibited by the participant will be cognitively assigned to a

particular dimension, creating less of an overlap between different dimensions. In the present study, it was possible that a single participant behavior was rated on multiple dimension. Consequently, rating between different dimensions within the same scenario was a greater possibility. The results here are in line with this explanation.

The second reason is related to the first in that it involves the lack of mutual exclusivity between PEDRs. Dimension chosen to be assessed in the AC are all related to effective behaviors on the job. As such, it is unreasonable to expect that there should be no relationship between different dimensions in the same exercise. In fact, the evidence found in the current study suggest that the average relationship between different dimensions from the same exercise is relatively small. However, because the hypothesis specifically asserted that DITLAC would have better discriminant validity than the TAC, hypothesis 4 was not supported.

Limitations

One of the limitations of the study was in the length of the assessment period. It is very common for ACs to last the entire day and even take part over several days. The length of the current study from start to finish was a total of four hours. Within the shorten time frame it was only feasible create three exercises, which is rather uncommon. According to Eurich et al. (2009), only 13% of all assessment conducted in the United States used three exercises or less. More commonly, ACs have four to five exercises (64%). If this study had used the common number of exercises, the assessment period would have been extended. In turn, if the study would have taken 6 or 7 hours a stronger argument could have been made of behaviors from the DITLAC design representing typical performance. In the present study, the four hour assessment period likely lessened the impact of aspects such as fatigue and motivation on the assessee's performance. However, if a longer study had been in place, the effects discovered in this study

could have been amplified. One of the features of maximum performance situations is the brevity of the assessment period (Sackett, 2007). As the assessment period increases in length, behavior begins to better represent typical performance. Therefore, if a longer study designed would have been incorporated, I believe stronger evidence would have been found.

Another limitation of this study concerns the lack of focus on situationally-specific variance. As mentioned earlier, some behaviors are more likelihood to occur in situation in which that behavior is more readily taped. For instance, the relationship between conscientiousness and planning and organizing was much stronger in the inbox exercise (r = .16) than in the interview exercise (r = .02) in the DITLAC, but relatively the same for the TAC (inbox, r = .10; interview, r = .10). The degree to which personality traits predicted PEDRs highly depended on the situation. However, the degree to which the exercise tapped each of the five dimensions was not assess a priori and, consequently, hypotheses were not formed around situation specificity. One way to compensate for this lack of foresight would have been to use suggestion from other research that has investigated this issue previously. However, the tasks in this study were explicitly designed to tap all five dimensions. Thus, future research will assess the trait-activation potential for each task and hypothesize with the expected effects. Given the general idea of trait-activation theory (Haaland & Christiansen, 2002; Lievens et al., 2006; Tett & Burnett, 2003), it would be logical to assume that there would be more evidence for the DITLAC. Since DITLAC is hypothesized to better assess typical performance, which is representative of an individual's stranding on personality traits, the greater these traits are 'activated' in a scenario the stronger the relationship between personality characteristics and PEDRs should be.

Lastly, the manipulation of this study could have been stronger. Though caution and care was taken to ensure that the tasks the individuals performed in each condition were equivalent, the focus on ensuring the strength of the manipulation was in the formation of the DITLAC. That is, more thought could have been given to the features of the TAC such that it would more accurately represent a maximum performance situation in a selection environment. For instance, the breaks that the participants took were clear and transparent, but they remained in the same room during the break. Other maximum performance features could have been incorporated as well. For instance, an experimenter acting as a rater could have been present in the room in order to reinforce the knowledge that they were being assessed. However, this should have increased the likelihood of finding the hypothesized effects. That is, the results of this study might actually represent attenuated effects.

Research Implications.

This is one of the few studies to investigate assessment centers in a laboratory environment. ACs are so seldom studied in the laboratory that it is not mentioned in many of the literature reviews. Although psychological realism may be lost by transferring this study into laboratory environment, the increased control over experimental condition may outweigh this concern. Specifically, the majority of field studies investigating construct validity often only manipulated what they had control over: the raters. They manipulated rater training, the measurements that raters used, the number of dimensions that needed to be rated by each rater, and the like. Hence, the conclusions made concerning the construct validity of ACs has almost exclusively been made based on results from a single aspect of an AC. This study demonstrates that not only are AC a viable option for laboratory studies, but the manipulation of aspects of the

AC can lead to differences in construct validity. As such, future research should look into to changing aspects of the AC itself in order to improve assessment of typical behavior.

Another implication for research concerns the criterion-related validity evidence of ACs. Over the past several decades, ACs have consistently provided strong criterion-related validity evidence (e.g., Arthur, et al., 2003; Gaugler, et al., 1987; Sackett, 1987). However, the results of this study bring some of those results into question. If DITLAC does indeed have better construct validity evidence than TAC, the question then becomes, why have TACs produced such strong criterion-related validity evidence? The answer to this question may lye within the metrics used to capture the criterion. Often, job performance ratings are single sourced rating and those ratings are often from the candidate's supervisor (Viswesvaran, Ones, & Schmidt, 1996). Supervisors, however, are more privy to witnessing an incumbent in maximum performance situations (Ployhart, et al., 2001). Thus, the stronger criterion-related evidence of TACs could be a consequence of the situational similarity. That is, behavioral ratings from TACs and job performance ratings are gleaned from maximum performance situations, which could explain the strong criterion-related validity evidence. Thus, future research you investigate the criterion related validity of ACs using multisource performance ratings to discover if the criterion-related validity evidence is as strong as the research as demonstrated over the past several decades.

Practical Implications.

The most transparent practical implication of the result from this study concerns the possibility of redesigning an AC to mirror typical day on the job. Redesigning standing ACs from the traditional structure to a day-in-the-life structure should not involve a vast amount of effort. Since the redesign is purely structural, current ACs can use the same tasks that they currently employ. The change that would occur would primarily be logistical in nature.

Accompany this design could lead to better decisions make about job candidates, which could lead to better fit, less absenteeism, and less turnover. Furthermore, utilizing a DITLAC structure, it is possible to improve the effectiveness of developmental feedback.

One of the longstanding issues surrounding the construct validity dilemma of ACs was that if it was not possible to accurately specific what dimension is being assessed, it is not possible to give effective feedback. That is, if you are not sure what you are measuring than how can you state that an employee needs improvement on certain dimension. Given the construct validity evidence of DITLAC, it would be possible to provide more effective developmental feedback to employees. As ACs become more affordable to run, this can be done for the more traditional AC jobs (e.g., high-level executives) to the high-volume jobs. This is especially true given the automation and the cost of the AC used in the current study.

The cost of developing and running an AC has always been a financial roadblock for utilizing this method for high-volume jobs. However, the costs of ACs have dramatically decreased with the dropping cost of technology and the outsourcing of ratings. This study adds to this evidence by being almost completely. Through the use of voice detection software, thought-out scripting, and pre-recorded videos, this study only required the labor of one research assistant to run the study. As such, ACs are becoming a financially viable option for high volumes jobs.

Directions for Future Research

There are two directions I believe researchers should focus on in the future. The first concerns the utilization of the day-in-the-life structure in future investigations. To my knowledge, this is the first multi-exercise DITLAC used in a laboratory study. Thus, there were many lessons learned which gave way to future research agendas. The structural features chosen to represent a typically day on the job could have been made more transparent. For instance,

many jobs require employee to take a break, however, that break may be up to the employees' direction. In the present study, the participant had no real incentive to take a break (e.g., the absence of a break room, coworkers to talk to, etc.). As such, future research should look into amplifying the characteristics of a DITLAC in future research. Additionally, I believe that there may be other structural features that future research should investigate. One feature that was not incorporated in this study was the utilization of time. A typical workday lasts for around eight hours, while the present experiment lasted only four. Thus, future could utilize new or different aspects of the workday into a DITLAC to see if behaviors are any more or less typical.

The second direction I believe future research should investigate concerns situational specificity. The hypotheses of this study were general such effects were compared across exercises and dimensions. However, a more fine-grained approach may lead to more meaningful (and stronger) results. For instance, while most of the conceptually-linked personality variables significantly predicted influencing others across all exercises, when this relationship broken down the exercise level, this relationship disappears in the inbox and interview exercises for some personality variables. Thus, certain personality traits only predict the manifestation of conceptually-linked behavior under certain situations. Though this idea has been investigated in TACs, it has yet to be explored in DITLAC.

Conclusion

The purpose of this study was to investigate the construct validity evidence in a day-in-the-life structured AC and compare it to a traditionally designed AC. Though the results of this study seldom demonstrated that the DITLAC design provided statistically better construct validity evidence compared to the TAC design, it did show that the DITLAC design provided statistically significant construct validity evidence when the TAC design did not. Thus, it serves

to reason that the DITLAC does provide the same, if not better, construct validity evidence of the TAC.

To my knowledge, this study is the first investigation that directly compares the construct validity of two different AC designs to one another. Additionally, this may be the first study that actively manipulates the structure of an AC beyond changes to the raters (e.g., use of behavioral checklists, training) or information provided to the assessees (e.g., skill transparency). As such, this study should serve as a foundation to directly manipulating the structure of the exercises in order to shed light on the construct validity problems of ACs. I believe there is a bright future in the utilization of the day-in-the-life design as it should theoretically assess typical performance better and, consequently should led to better selection/promotion decisions.

APPENDIX A: E-MAIL CONTENT

E-Mail 1

Dear SUMMIT Hospital,

My daughter and I came into the emergency room two days ago after she had fallen off her bicycle. When we first came in, someone told us to go to the ER administrators to check-in so the triage nurse could examine the severity of my daughter's injury. Though she was bleeding profusely and couldn't stop crying that stupid nurse didn't think she was bad enough to go back to see a doctor. We had to wait over an hour until they took us back. They just gave her some medicine for the pain, but I know my daughter and that didn't help at all. Every time I asked one of the ER assistants to go back or for more medicine, they just denied my requests. How could they do nothing while my child was in such pain?! INEXCUSABLE!!!

I asked to talk to the supervisor on duty, only to find out that she was just as rude as everyone else at your hospital. If I didn't argue, my daughter's arm could have gotten infected and it would have been SUMMIT Hospital's fault. You and your entire staff are completely incompetent. You are the worse hospital ever and I'm going to make sure to tell all my friends to avoid your hospital at all costs.

Infuriated,

Meghan Frizgerald

E-Mail 2

Hello there,

I'm writing you concerning my recent trip to your Emergency Room. My wife and I came into your hospital just a few days ago because she was having stomach issues. I had never seen her in so much pain in my entire life and I was really worried about her. We checked in and waited over 2 hours to see someone. The only people that I saw in the emergency room lobby were an old guy with a cough and some lady – I didn't see anything wrong with her. I didn't even get to speak to the attendant because I couldn't even find one.

I know an emergency room is an extremely chaotic and stressful environment – and I understand that you have a lot going on with your job. I was just really disappointed that I couldn't find anyone to help with my wife's pain. It was really hard seeing her in so much pain. I was expecting a lot better given your hospital's reputation.

Regards,

Steven McPherson

E-Mail 3

Hello SUMMIT Hospital Staff,

I just wanted to send all of you a message with how please I was with our experience. When my son and I first arrived to the emergency room, I was really concerned because my son had been up the whole night before with a fever and I couldn't get ahold of our pediatrician. I was afraid something was really wrong with him, so I took him to your ER.

When I first arrive, the triage nurse asked me a couple questions about his medical history and told me not to worry about anything. She gave my son some medicine and talked to him so sweetly. She even gave him a sticker for being a "brave little man" and a balloon. This was beyond my expectations because I have had some terrible emergency room experiences before. But it gets better. Once we were called back, it only took us about a half a hour; the doctor we saw was equally as sweet. He assured me that it was only a fever and it would break soon. He also gave me the number of an emergency pediatrician if I couldn't get a hold of mine again.

We were only in the hospital for about an hour and I don't think our experience could have been any more pleasant. My son and I were so appreciative; we just wanted to let you all know.

Sincerely,

Marcy and Kyle Magnolia

E-Mail 4

Dear SUMMIT Hospital Staff,

What is wrong with you? I was in some serious pain and none of your "staff" seemed to care. Honestly, your inability to get me back to see a doctor takes real skill. I think that a bunch of monkeys do your job better than you. I felt like half my body was numb and none of your assistants lifted a finger.

Even when I did get back to see a doctor, he was essentially useless. I now have an appointment to get a CAT scan, but have no diagnosis. What is the use of your hospital if you can't even diagnose a medical condition?

Thanks for nothing!

Jeremiah Plankton

E-Mail 5

Dear Summit Hospital,

I am writing this email to inform you of my recent experience with your emergency room staff. Last weekend I brought my 16 year old daughter to your ER in hopes that you would help her as best you could. My daughter, Katie, was having terrible abdominal pain. We had to wait over

two hours to be seen by a nurse and once we were taken back – the nurse insisted on asking very personal questions. So personal, that I do not wish to repeat them here.

When the doctor came in, it seemed like he was in a rush and we weren't a priority. He simply told Katie that she should eat healthier and exercise more and I swear he said he had more urgent matters to attend to. Not a single test was run to see if anything urgent could be causing her abdominal pain.

I don't mean to be rude, I just really care about my daughter. This was her first trip to the emergency room and I know she was really scared. I had heard such good things about SUMMIT Hospital and I was just really disappointed you were unable to confirm those opinions.

-Vicki Westerly

E-Mail 6

Summit Hospital Staff,

I recently came into your ER with a broken wrist. The wait time was a short 20 minutes, and the staff was beyond helpful. I was seen by two nurses, and the doctor stayed with me for almost an hour to run other tests and make certain that my wrist was the only thing injured in my accident. Afterwards, he prescribed pain medication and put my wrist in a cast.

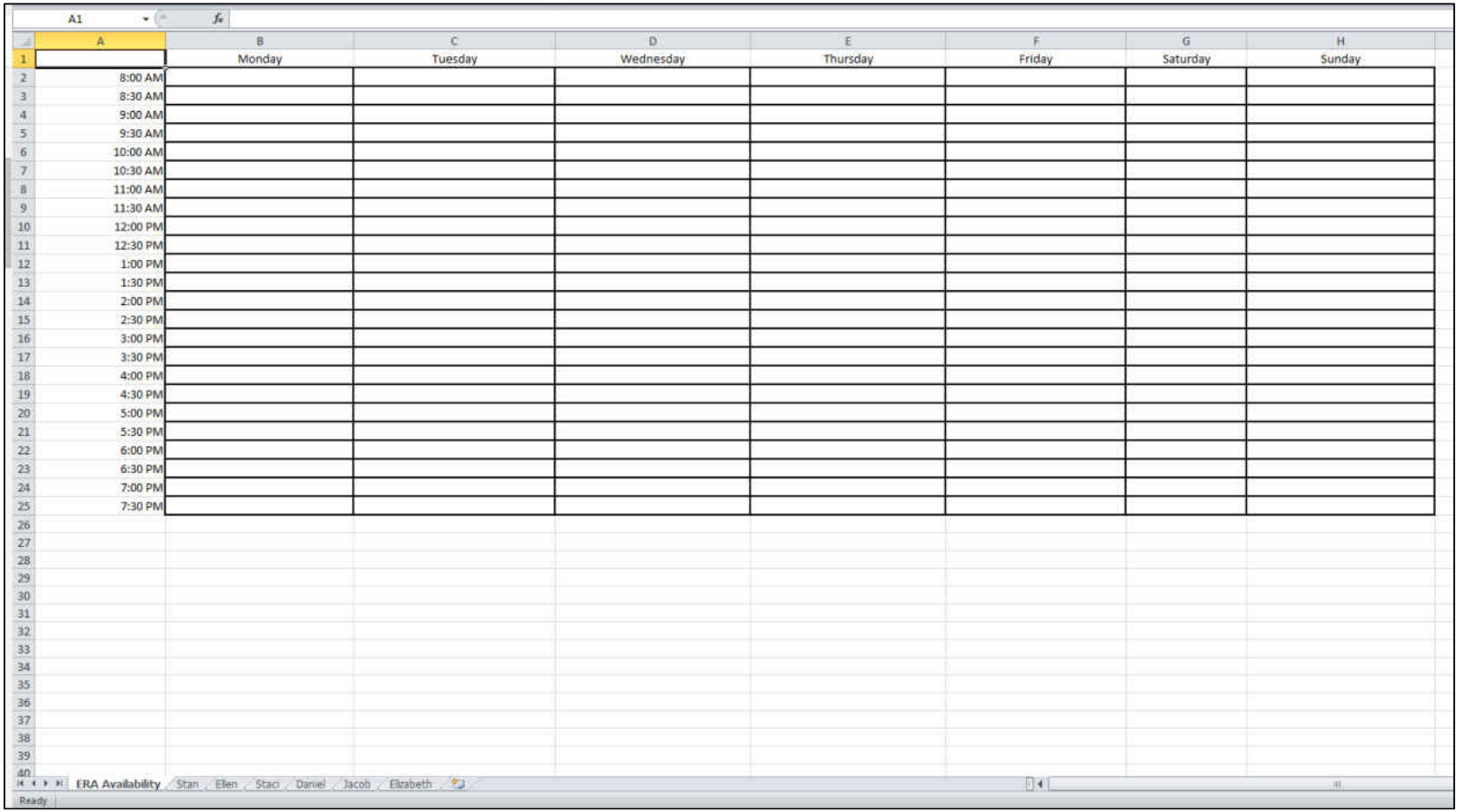
I am a very clumsy man and have been to many emergency rooms, but yours was by far the only one I would return to without hesitation should something like this happen again (and it will). I just wanted to thank you for making an unpleasant experience more bearable.

Keep up the good work,

Tim Myers

APPENDIX B: INBOX EXCEL TOOLS

Appendix B.1 – Schedule Entry Form



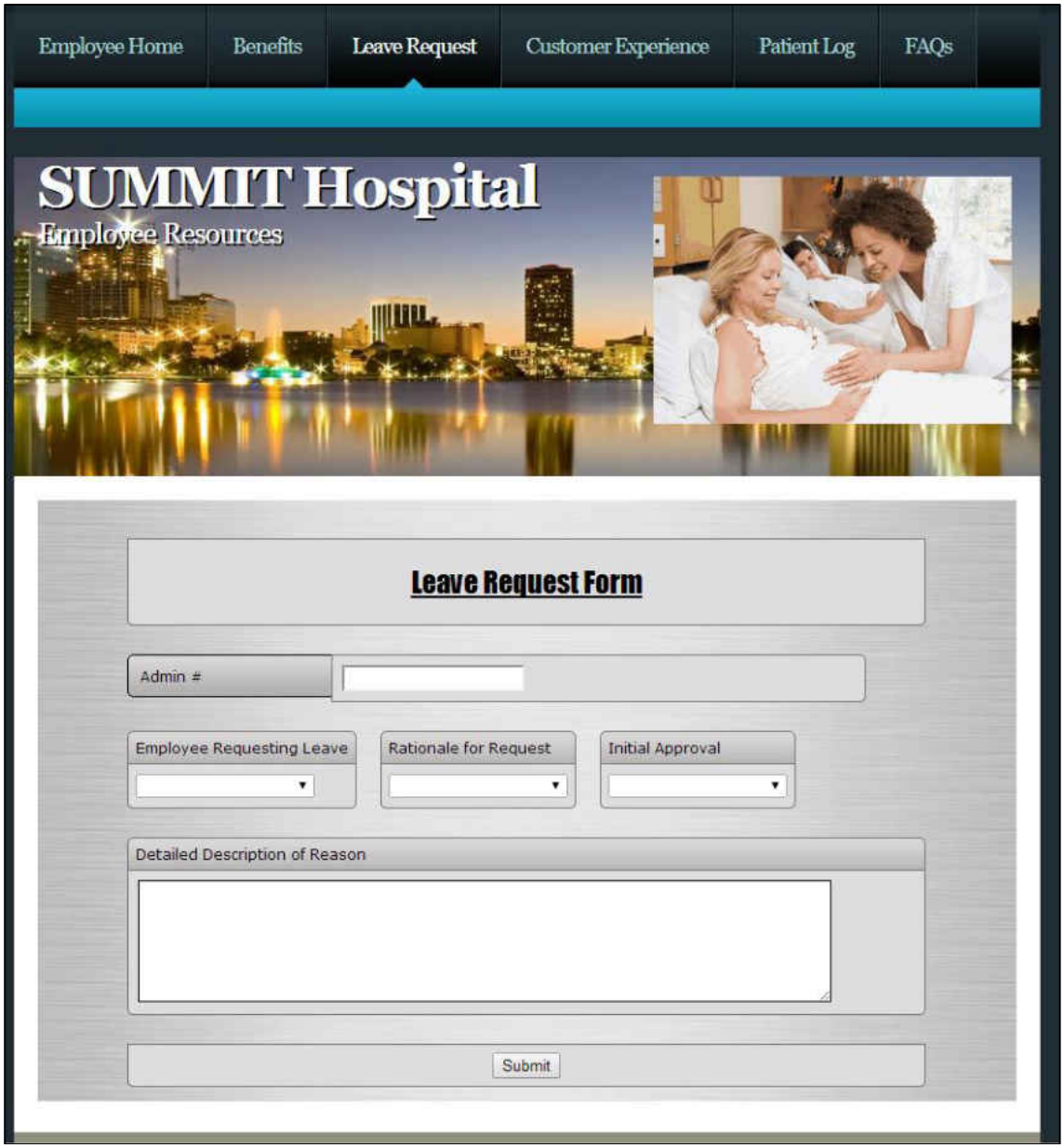
	A	B	C	D	E	F	G	H
1		Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
2	8:00 AM							
3	8:30 AM							
4	9:00 AM							
5	9:30 AM							
6	10:00 AM							
7	10:30 AM							
8	11:00 AM							
9	11:30 AM							
10	12:00 PM							
11	12:30 PM							
12	1:00 PM							
13	1:30 PM							
14	2:00 PM							
15	2:30 PM							
16	3:00 PM							
17	3:30 PM							
18	4:00 PM							
19	4:30 PM							
20	5:00 PM							
21	5:30 PM							
22	6:00 PM							
23	6:30 PM							
24	7:00 PM							
25	7:30 PM							

Appendix B.2 – Final Schedule Form

	A	B	C	D	E	F	G	H
1		Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
2	8:00 AM							
3								
4	8:30 AM							
5								
6	9:00 AM							
7								
8	9:30 AM							
9								
10	10:00 AM							
11								
12	10:30 AM							
13								
14	11:00 AM							
15								
16	11:30 AM							
17								
18	12:00 PM							
19								
20	12:30 PM							
21								
22	1:00 PM							
23								
24	1:30 PM							
25								
26	2:00 PM							
27								
28	2:30 PM							
29								
30	3:00 PM							
31								
32	3:30 PM							
33								
34	4:00 PM							
35								
36	4:30 PM							
37								
38	5:00 PM							
39								
40	5:30 PM							
41								
42	6:00 PM							
43								
44	6:30 PM							
45								
46	7:00 PM							
47								
48	7:30 PM							
49								
50								
51	**Total Hours Schec**	**Minutes**		**Hours**				
52	Stan	0		0				
53	Ellen	0		0				
54	Staci	0		0				
55	Daniel	0		0				
56	Jacob	0		0				
57	Elizabeth	0		0				
58								
59								

ERA Final Schedule

APPENDIX C: WEBSITE EXAMPLE


Employee Home
Benefits
Leave Request
Customer Experience
Patient Log
FAQs
SUMMIT Hospital
Employee Resources
Leave Request Form
Admin #
Employee Requesting Leave
Rationale for Request
Initial Approval
Detailed Description of Reason
Submit

APPENDIX D: EXAMPLE OF SUMMIT HOSPITAL PATIENT PACKET

Administrator #: 522

SUMMIT HOSPITAL PATIENT PACKET

Check-In: 4:30 pm Admitted: 5:12 pm Check-Out: 6:30 pm

Please completely fill out all information. If you have any questions, please ask the emergency room attendant.

GENERAL INFORMATION

Name: Kevin (First) S (M) Flynn (Last)

Address: 8230 Pamlico Street
Street Address Line 1

Street Address Line 2

Orlando (City) Florida (State)

32817 (Zip Code) United States (Country)

Home Phone: 407-251-6957 Cell Phone: 407-593-8328

DOB (MM/DD/YY): 05/14/1984 Gender: Male

Height: 5'8" Weight: 173 lbs

E-Mail: B.Flynn@bellsouth.net Marital Status: Single

SUMMIT HOSPITAL

Administrator #: 522

SUMMIT HOSPITAL PATIENT PACKET

MEDICAL HISTORY

Have you ever had (Please check all that apply)

☐ Anemia	☑ Asthma	☐ Arthritis	☐ Cancer
☐ Gout	☐ Diabetes	☐ Emotional Disorder	☐ Epilepsy Seizures
☑ Fainting Spells	☐ Gallstones	☐ Heart Disease	☐ Heart Attack
☐ Rheumatic Fever	☑ High Blood Pressure	☐ Digestive Problems	☐ Ulcerative Colitis
☐ Ulcer Disease	☐ Hepatitis	☐ Kidney Disease	☐ Liver Disease
☐ Sleep Apnea	☐ Use a C-PAP machine	☐ Thyroid Problems	☐ Tuberculosis
☐ Venereal Disease	☐ Neurological Disorders	☐ Bleeding Disorders	☐ Lung Disease
☐ Emphysema			

☐ Other: _____

Current Medications

Please list your current medications:

Prozac, Tylenol, Naproxen, & Asmanex

Administrator #: _____

SUMMIT HOSPITAL PATIENT PACKET

Medical Operations

Please list all major medical operations you've have in the past 10 years

Date	Operation Description
2003	had appendix removed
2005	ACL surgery

Administrator #: 522

SUMMIT HOSPITAL PATIENT PACKET

Healthy and Unhealthy Habits

To get a better idea of your current health, please answer the following questions:

How often do you exercise each week?

☑ Never ☐ 1-2 days ☐ 3-4 days ☐ 5+ days

To what degree to you follow a structured diet?

☑ Loose Diet Plan ☐ Strict Diet Plan ☐ No Diet Plan

How many glasses of alcohol do you consume a day?

☐ I don't drink ☑ 1-2 glasses ☐ 3-4 glasses ☐ 5+ glasses

How many cups of caffeine do you consume a day?

☐ I don't drink ☐ 1-2 cups ☑ 3-4 cups ☐ 5+ cup

How many packs of cigarettes do you smoke a week?

☐ I don't smoke ☑ 1-2 packs ☐ 3-4 packs ☐ 5+ packs

Administrator #: _____

SUMMIT HOSPITAL PATIENT PACKET

Comments on Medical History

Please include any other comments regarding your medical history:

Have had asthma most of my life. Could be contributing to my respiratory problem.

Triage Nurse Initial Assessment Form

PHYSIOLOGICAL INFORMATION

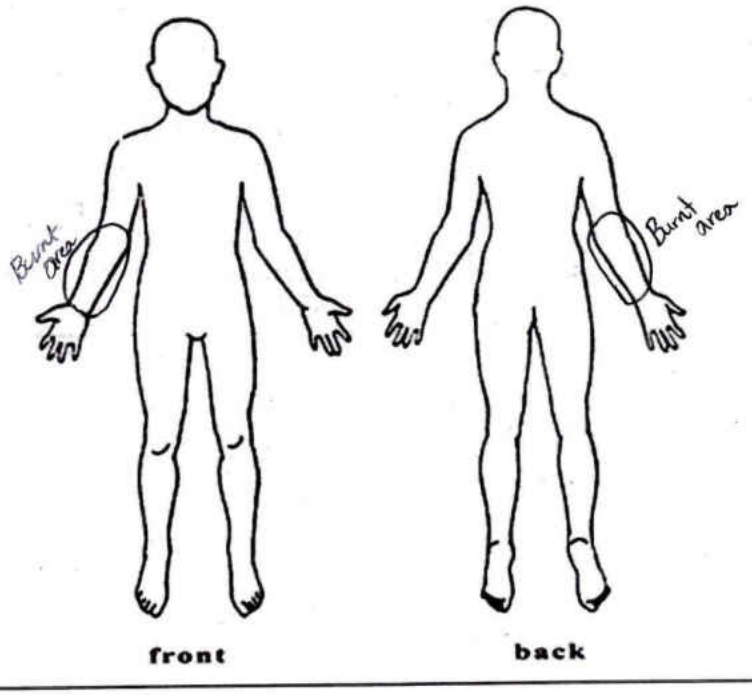
Please report body temperature and blood pressure below

Temperature 98.9

Heart rate 94 ~~89 BPM~~ Systolic BP 140 Stage 1 Hypertension Diastolic BP 89 Prehypertension

INDICATION OF PHYSICAL ALIMENTS

Please describe the patient's physical aliments



DESCRIPTION OF SYMPTOMS

Please describe the patient's symptoms

Patient has burns on right arm, starting at ~~finger~~ fingertips to almost the elbow. Heart may have stopped during electrocution - Heart monitor should be checked

Severity

Please indicate the level of severity below

Low Severity — High Severity

1 2 3 4 5 6 7 8 (9) 10

APPENDIX E: WONDERLIC EXAMPLE ITEMS

Sample Questions

Wonderlic Contemporary Cognitive Ability Test

Formerly the Wonderlic Personnel Test – Revised (WPT-R)

1. Which of the following is the earliest date?

   A) Jan. 16, 1898 B) Feb. 21, 1889 C) Feb. 2, 1898 D) Jan. 7, 1898 E) Jan. 30, 1889

2. **LOW** is to **HIGH** as **EASY** is to __?__.

   J) **SUCCESSFUL** K) **PURE** L) **TALL** M) **INTERESTING** N) **DIFFICULT**

3. A featured product from an Internet retailer generated 27, 99, 80, 115 and 213 orders over a 5-hour period. Which graph below best represents this trend?

   A B C D E

4. What is the next number in the series? 29 41 53 65 77 __?__

   J) 75 K) 88 L) 89 M) 98 N) 99

5. *One word below appears in color. What is the OPPOSITE of that word?*
   She gave a complex answer to the question and we all agreed with her.

   A) long B) better C) simple D) wrong E) kind

6. Jose's monthly parking fee for April was $150; for May it was $10 more than April; and for June $40 more than May. His average monthly parking fee was __?__ for these 3 months.

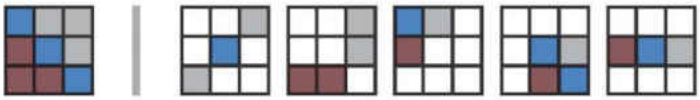
   J) $66 K) $160 L) $166 M) $170 N) $200

7. *If the first two statements are true, is the final statement true?*
   Sandra is responsible for ordering all office supplies.
   Notebooks are office supplies.

   Sandra is responsible for ordering notebooks.

   A) yes B) no C) uncertain

8. Which THREE choices are needed to create the figure on the left? Only pieces of the same color may overlap.



   J K L M N

9. Which THREE of the following words have similar meanings?

   A) observable B) manifest C) hypothetical D) indefinite E) theoretical

10. Last year, 12 out of 600 employees at a service organization were rewarded for their excellence in customer service, which was __?__ of the employees.

    J) 1% K) 2% L) 3% M) 4% N) 6%

Correct Answers: 1. E, 2. N, 3. D, 4. L, 5. C, 6. M, 7. A, 8. KLM, 9. CDE, 10. K

Pinpoint potential.
Wonderlic.
www.wonderlic.com | 800.323.3742

APPENDIX F: IRB APPROVAL LETTER

University of Central Florida Institutional Review Board
Office of Research & Commercialization
12201 Research Parkway, Suite 501
Orlando, Florida 32826-3246
Telephone: 407-823-2901 or 407-882-2276
www.research.ucf.edu/compliance/irb.html

Approval of Human Research

From: **UCF Institutional Review Board #1**
FWA00000351, IRB00001138

To: **Christopher Wiese**

Date: **August 28, 2013**

Dear Researcher:

On 8/28/2013, the IRB approved the following human participant research until 8/27/2014 inclusive:

Type of Review:	UCF Initial Review Submission Form
Project Title:	Day-In-The-Life Assessment Center
Investigator:	Christopher Wiese
IRB Number:	SBE-13-09575
Funding Agency:	
Grant Title:	
Research ID:	N/A

The scientific merit of the research was considered during the IRB review. The Continuing Review Application must be submitted 30days prior to the expiration date for studies that were previously expedited, and 60 days prior to the expiration date for research that was previously reviewed at a convened meeting. Do not make changes to the study (i.e., protocol, methodology, consent form, personnel, site, etc.) before obtaining IRB approval. A Modification Form **cannot** be used to extend the approval period of a study. All forms may be completed and submitted online at https://iris.research.ucf.edu .

If continuing review approval is not granted before the expiration date of 8/27/2014, approval of this research expires on that date. When you have completed your research, please submit a Study Closure request in iRIS so that IRB records will be accurate.

Use of the approved, stamped consent document(s) is required. The new form supersedes all previous versions, which are now invalid for further use. Only approved investigators (or other approved key study personnel) may solicit consent for research participation. Participants or their representatives must receive a copy of the consent form(s).

In the conduct of this research, you are responsible to follow the requirements of the Investigator Manual.

On behalf of Sophia Dziegielewski, Ph.D., L.C.S.W., UCF IRB Chair, this letter is signed by:

Signature applied by Joanne Muratori on 08/28/2013 03:02:53 PM EDT

Joanne Muratori

IRB Coordinator

REFERENCES

Allport, G. W. (1937). *Personality: A Psychological Interpretation*. Oxford, England: Holt.

Arthur, W. J., Day, E. A., McNelly, T. L., & Edens, P. S. (2003). A Meta-Analysis of the Criterion-Related Validity of Assessment Center Dimensions. *Personnel Psychology*, *56*(1), 125–153. doi:10.1111/j.1744-6570.2003.tb00146.x

Arthur, W., Woehr, D. J., & Maldegen, R. (2000). Convergent and Discriminant Validity of Assessment Center Dimensions: A Conceptual and Empirical Reexamination of the Assessment Center Construct-Related Validity Paradox. *Journal of Management*, *26*(4), 813–835. doi:10.1177/014920630002600410

Bailey, B. P., & Iqbal, S. T. (2005). *Leveraging Changes in Mental Workload During Task Execution to Mitigate Effects of Interruption* (No. UIUCDCS-R-2005-2623) (pp. 741–750). University of Illinois at Urbana-Champaign.

Bakan, D. (1966). *The Duality of Human Existence: Isolation and Communion in Wester Man*. Boston, MA: Beacon.

Beaty, J. C., Jr, Cleveland, J. N., & Murphy, K. R. (2001). The Relation Between Personality and Contextual Performance in" Strong" Versus“ Weak” Situations. *Human Performance*, *14*(2), 125–148.

Bettencourt, B. A., Talley, A., Benjamin, A. J., & Valentine, J. (2006). Personality and aggressive behavior under provoking and neutral conditions: A meta-analytic review. *Psychological Bulletin*, *132*(5), 751–777. doi:10.1037/0033-2909.132.5.751

Beus, J. M., & Whitman, D. S. (2012). The Relationship Between Typical and Maximum Performance: A Meta-Analytic Examination. *Human Performance*, *25*(5), 355–376. doi:10.1080/08959285.2012.721831

Bray, D. W., & Grant, D. L. (1966). The Assessment Center in the Measurement of Potential for Business Management. *Psychological Monographs: General and Applied*, *80*(17), 1–27. doi:10.1037/h0093895

Chan, D., & Schmitt, N. (1997). Video-based versus paper-and-pencil method of assessment in situational judgment tests: subgroup differences in test performance and face validity perceptions. *Journal of Applied Psychology*, *82*(1), 143–159. doi:10.1037/0021-9010.82.1.143

Chatman, J. A. (1989). Improving Interactional Organizational Research: a Model of Person-Organization Fit. *Academy of Management Review*, *14*(3), 333–349. doi:10.5465/AMR.1989.4279063

Cooper, W. H., & Withey, M. J. (2009). The Strong Situation Hypothesis. *Personality and Social Psychology Review*, *13*(1), 62–72. doi:10.1177/1088868308329378

Cronbach, L. J. (1960). *Essentials of Psychological Testing* (Vol. 2). New York, NY: Harper.

Cronbach, L. J., & Meehl, P. E. (1955). Construct Validity In Psychological Tests. *Psychological Bulletin*, *52*(4), 281–302.

Donahue, L. M., Truxillo, D. M., Cornwell, J. M., & Gerrity, M. J. (1997). Assessment Center Construct Validity and Behavioral Checklists: Some Additional Findings. *Journal of Social Behavior & Personality*, *12*(5), 85–108.

Donovan, J. J., & Radosevich, D. J. (1999). A Meta-Analytic Review of the Distribution of Practice Effect: Now You See It, Now You Don't. *Journal of Applied Psychology*, *84*(5), 795–805.

Driskell, J. E., Willis, R. P., & Copper, C. (1992). Effect of Overlearning on Retention. *Journal of Applied Psychology*, *77*(5), 615–622.

DuBois, C. L., Sackett, P. R., Zedeck, S., & Fogli, L. (1993). Further Exploration of Typical and Maximum Performance Criteria: Definitional Issues, Prediction, and White-Black Differences. *Journal of Applied Psychology*, *78*(2), 205–211. doi:10.1037/0021-9010.78.2.205

Dugan, B. (1988). Effects of Assessor Training on Information Use. *Journal of Applied Psychology*, *73*(4), 743–748.

Endler, N. S., & Parker, J. D. (1992). Interactionism Revisited: Reflections on the Continuing Crisis in the Personality Area. *European Journal of Personality*, *6*(3), 177–198.

Eurich, T. L., Krause, D. E., Cigularov, K., & Thornton, G. C. (2009). Assessment Centers: Current Practices in the United States. *Journal of Business and Psychology*, *24*(4), 387–407. doi:10.1007/s10869-009-9123-3

Eyrolle, H., & Cellier, J.-M. (2000). The effects of interruptions in work activity: Field and laboratory results. *Applied Ergonomics*, *31*(5), 537–543. doi:10.1016/S0003-6870(00)00019-3

Finstad, K., Bink, M., McDaniel, M., & Einstein, G. O. (2006). Breaks and Task Switches in Prospective Memory. *Applied Cognitive Psychology*, *20*(5), 705–712. doi:10.1002/acp.1223

Furnham, A., Crump, J., & Whelan, J. (1997). Validating the NEO Personality Inventory Using Assessor's Ratings. *Personality and Individual Differences*, *22*(5), 669–675. doi:10.1016/S0191-8869(96)00261-9

Galton, F. (1965). Galton on Mental Capacity. In R. J. Herrnstein & E. G. Boring (Eds.), *A Sourcebook in the History of Psychology* (pp. 421–423). Cambridge, MA: Harvard University Press.

Gaugler, B. B., Rosenthal, D. B., Thornton, G. C., & Bentson, C. (1987). Meta-analysis of

assessment center validity. *Journal of Applied Psychology*, *72*(3), 493–511. doi:10.1037/0021-9010.72.3.493

Gaugler, B., & Thornton, G. (1989). Number of assessment center dimensions as a determinant of assessor accuracy. *Journal of Applied Psychology*, *74*(4), 611–618.

Goldstein, H. W., Yusko, K. P., Braverman, E. P., Smith, D. B., & Chung, B. (1998). The Role of Cognitive Ability in the Subgroup Differences and Incremental Validity of Assessment Center Exercises. *Personnel Psychology*, *51*(2), 357–374. doi:10.1111/j.1744-6570.1998.tb00729.x

Haaland, S., & Christiansen, N. D. (2002). Implications of Trait-Activation Theory for Evaluating the Construct Validity of Assessment Center Ratings. *Personnel Psychology*, *55*(1), 137–163. doi:10.1111/j.1744-6570.2002.tb00106.x

Hennessy, J., Mabey, B., & Warr, P. (1998). Assessment Centre Observation Procedures: an Experimental Comparison of Traditional, Checklist and Coding Methods. *International Journal of Selection and Assessment*, *6*(4), 222–231. doi:10.1111/1468-2389.00093

Highhouse, S., & Harris, M. M. (1993). The Measurement of Assessment Center Situations: Bem's Template Matching Technique for Examining Exercise Similarity. *Journal of Applied Social Psychology*, *23*(2), 140–155. doi:10.1111/j.1559-1816.1993.tb01057.x

Hinrichs, J. R. (1978). An Eight-Year Follow-Up of a Management Assessment Center. *Journal of Applied Psychology*, *63*(5), 596–601. doi:10.1037/0021-9010.63.5.596

Howard, A. (1974). An Assessment of Assessment Centers. *The Academy of Management Journal*, *17*(1), 115–134.

International Task Force on Assessment Center Guidelines. (2009). Guidelines and Ethical Considerations for Assessment Center Operations. *International Journal of Selection and*

Assessment, *17*(3), 243–253. doi:10.1111/j.1468-2389.2009.00467.x

Johns, G. (2006). The Essential Impact of Context on Organizational Behavior. *Academy of Management Review*, *31*(2), 386–408. doi:10.5465/AMR.2006.20208687

Klehe, U.-C., & Latham, G. (2006). What Would You Do—Really or Ideally? Constructs Underlying the Behavior Description Interview and the Situational Interview in Predicting Typical Versus Maximum Performance. *Human Performance*, *19*(4), 357–382. doi:10.1207/s15327043hup1904_3

Kleinmann, M. (1993). Are rating dimensions in assessment centers transparent for participants? Consequences for criterion and construct validity. *Journal of Applied Psychology*, *78*(6), 988–992.

Klimoski, R., & Brickner, M. (1987). Why Do Assessment Centers Work? the Puzzle of Assessment Center Validity. *Personnel Psychology*, *40*(2), 243–260. Retrieved from http://resolver.flvc.org/ucf?sid=google&auinit=R&aulast=Klimoski&atitle=Why+do+assessment+centers+work%3F+The+puzzle+of+assessment+center+validity&id=doi:10.1111/j.1744-6570.1987.tb00603.x&title=Personnel+psychology&volume=40&issue=2&date=1987&spage=243&issn=0031-5826

König, C., Melchers, K., Kleinmann, M., Richter, G., & Klehe, U. (2007). Candidates'Ability to Identify Criteria in Nontransparent Selection Procedures: Evidence from an assessment center and a structured interview. *International Journal of Selection and Assessment*, *15*(3), 283–292.

Kvavilashvili, L. (1998). Remembering Intentions: Testing a New Method of Investigation. *Applied Cognitive Psychology*, *12*(6), 533–554. doi:10.1002/(SICI)1099-

0720(1998120)12:6<533::AID-ACP538>3.0.CO;2-1

Lachman, R., Lachman, L. J., & Butterfield, E. C. (1979). *Cognitive Psychology and Information Processing: an Introduction*. Hillsdale, NJ: Erlbaum.

Lance, C. (2008). Why Assessment Centers Do Not Work the Way They Are Supposed to. *Industrial and Organizational Psychology*, *1*(1), 84–97. doi:10.1111/j.1754-9434.2007.00017.x

Lance, C. E., Noble, C. L., & Scullen, S. E. (2002). A critique of the correlated trait-correlated method and correlated uniqueness models for multitrait-multimethod data. *Psychological Methods*, *7*(2), 228–244. doi:10.1037//1082-989X.7.2.228

Lewin, K. (1943). Defining the "Field at a Given Time.." *Psychological Review*, *50*(3), 292–310. doi:10.1037/h0062738

Lievens, F. (1998). Factors Which Improve the Construct Validity of Assessment Centers: a Review. *International Journal of Selection and Assessment*, *6*(3), 141–152. doi:10.1111/1468-2389.00085

Lievens, F. (2001). Assessor Training Strategies and Their Effects on Accuracy, Interrater Reliability, and Discriminant Validity. *Journal of Applied Psychology*, *86*(2), 255–264. doi:10.1037//0021-9010.86.2.255

Lievens, F., & Conway, J. M. (2001). Dimension and Exercise Variance in Assessment Center Scores: a Large-Scale Evaluation of Multitrait-Multimethod Studies. *Journal of Applied Psychology*, *86*(6), 1202–1222. doi:10.1037//0021-9010.86.6.1202

Lievens, F., Chasteen, C. S., Day, E. A., & Christiansen, N. D. (2006). Large-Scale Investigation of the Role of Trait Activation Theory for Understanding Assessment Center Convergent and Discriminant Validity. *Journal of Applied Psychology*, *91*(2), 247–258.

doi:10.1037/0021-9010.91.2.247

Macan, T. H., Avedon, M. J., Paese, M., & Smith, D. E. (1994). The Effects of Applicants' Reactions to Cognitive Ability Tests and an Assessment Center. *Personnel Psychology*, *47*(4), 715–738.

Meyer, R. D., Dalal, R. S., & Bonaccio, S. (2009). A Meta-Analytic Investigation Into the Moderating Effects of Situational Strength on the Conscientiousness-Performance Relationship. *Journal of Organizational Behavior*, *30*(8), 1077–1102. doi:10.1002/job.602

Meyer, R. D., Dalal, R. S., & Hermida, R. (2010). A Review and Synthesis of Situational Strength in the Organizational Sciences. *Journal of Management*, *36*(1), 121–140. doi:10.1177/0149206309349309

Meyer, R. D., Dalal, R. S., Jose, I. J., Hermida, R., Chen, T. R., Vega, R. P., et al. (2011). Measuring Job-Related Situational Strength and Assessing Its Interactive Effects With Personality on Voluntary Work Behavior. *Journal of Management*. doi:10.1177/0149206311425613

Mischel, W. (1973). Toward a Cognitive Social Learning Reconceptualization of Personality. *Psychological Review*, *80*(4), 252–283. doi:10.1037/h0035002

Mischel, W. (1977). The Interaction of Person and Situation. In D. Magnusson & N. S. Endler (Eds.), *Personality at the Crossroads: Current Issues in Interactional Psychology* (pp. 333–352). Hillsdale, NJ: Lawrence Erlbaum.

Mischel, W., & Shoda, Y. (1998). Reconciling Processing Dynamics and Personality Dispositions. *Annual Review of Psychology*, *49*(1), 229–258. doi:10.1146/annurev.psych.49.1.229

Mitchell, M. S., & Ambrose, M. L. (2007). Abusive supervision and workplace deviance and the

moderating effects of negative reciprocity beliefs. *Journal of Applied Psychology, 92*(4), 1159-1168. doi: 10.1037/0021-9010.92.4.1159

Motowidlo, S. J., & Beier, M. E. (2010). Differentiating specific job knowledge from implicit trait policies in procedural knowledge measured by a situational judgment test. *Journal of Applied Psychology*, *95*(2), 321–333. doi:10.1037/a0017975

Motowidlo, S. J., Hooper, A. C., & Jackson, H. L. (2006). Implicit Policies About Relations Between Personality Traits and Behavioral Effectiveness in Situational Judgment Items. *Journal of Applied Psychology*, *91*(4), 749–761. doi:10.1037/0021-9010.91.4.749

Neidig, R. D., & Neidig, P. J. (1984). Multiple Assessment Center Exercises and Job Relatedness. *Journal of Applied Psychology*, *69*(1), 182–186. doi:10.1037/0021-9010.69.1.182

Oostrom, J. K., Born, M. P., 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:10.1080/08959285.2012.703732

Ployhart, R. E., Lim, B.-C., & Chan, K. Y. (2001). Exploring Relations Between Typical and Maximum Performance Ratings and the Five Factor Model of Personality. *Personnel Psychology*, *54*(4), 809–843. doi:10.1111/j.1744-6570.2001.tb00233.x/full

Ratwani, R. M., Andrews, A. E., Sousk, J. D., & Trafton, J. G. (2008). The Effect of Interruption Modality on Primary Task Resumption. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *52*(4), 393–397. doi:10.1177/154193120805200441

Reilly, R. R., Henry, S., & Smither, J. W. (1990). An Examination of the Effects of Using Behavior Checklists on the Construct Validity of Assessment Center Dimensions. *Personnel*

Psychology, *43*(1), 71–84.

Robie, C., Brown, D. J., & Bly, P. R. (2008). Relationship Between Major Personality Traits and Managerial Performance: Moderating Effects of Derailing Traits. *International Journal of Management*, *25*(1), 131–139.

Rogers, T. B. (1981). A Model of the Self as an Aspect of the Human Information Processing System. In N. Cantor & J. F. Kihlstrom (Eds.), *Personality, Cognition, And Social Interaction* (pp. 193–214). Hillsdale, NJ: Lawrence Erlbaum.

Sackett, P. R. (1987). Assessment centers and content validity: Some neglected issues. *Personnel Psychology*, *40*(1), 13–25.

Sackett, P. R. (2007). Revisiting the Origins of the Typical-Maximum Performance Distinction. *Human Performance*, *20*(3), 179–185. doi:10.1080/08959280701332968

Sackett, P. R., & Dreher, G. F. (1982). Constructs and assessment center dimensions: Some troubling empirical findings. *Journal of Applied Psychology*, *67*(4), 401–410.

Sackett, P. R., & Hakel, M. D. (1979). Temporal stability and individual differences in using assessment information to form overall ratings. *Organizational Behavior and Human Performance*, *23*(1), 120–137. doi:10.1016/0030-5073(79)90051-5

Sackett, P. R., & Harris, M. M. (1988). A further examination of the constructs underlying assessment center ratings. *Journal of Business and Psychology*, *3*(2), 214–229.

Sackett, P. R., Zedeck, S., & Fogli, L. (1988). Relations between measures of typical and maximum job performance. *Journal of Applied Psychology*, *73*(3), 482.

Sagie, A., & Magnezy, R. (1997). Assessor Type, Number of Distinguishable Dimension Categories, and Assessment Centre Construct Validity. *Journal of Occupational Psychology*, *70*(1), 103–108.

Schleicher, D. J., Day, D. V., Mayes, B. T., & Riggio, R. E. (2002). A new frame for frame-of-reference training: Enhancing the construct validity of assessment centers. *Journal of Applied Psychology*, *87*(4), 735–746. doi:10.1037//0021-9010.87.4.735

Schneider, J. R., & Schmitt, N. (1992). An Exercise Design Approach to Understanding Assessment Center Dimension and Exercise Constructs. *Journal of Applied Psychology*, *77*(1), 32–41.

Smithikrai, C. (2008). Moderating Effect of Situational Strength on the Relationship Between Personality Traits and Counterproductive Work Behaviour. *Asian Journal of Social Psychology*, *11*(4), 253–263. doi:10.1111/j.1467-839X.2008.00265.x

Song, L. J., Tsui, A. S., Law, K. S., & undefined author. (2007). Unpacking Employee Responses to Organizational Exchange Mechanisms: The Role of Social and Economic Exchange Perceptions. *Journal of Management*, *35*(1), 56–93. doi:10.1177/0149206308321544

Spector, P. E., Schneider, J. R., Vance, C. A., & Hezlett, S. A. (2000). The Relation of Cognitive Ability and Personality Traits to Assessment Center Performance. *Journal of Applied Social Psychology*, *30*(7), 1474–1491. doi:10.1111/j.1559-1816.2000.tb02531.x

Spychalski, A. C., Quiñones, M. A., Gaugler, B. B., & Pohley, K. (1997). A Survey of Assessment Center Practices in Organizations in the United States. *Personnel Psychology*, *50*(1), 71–90. doi:10.1111/j.1744-6570.1997.tb00901.x

Tett, R. P., & Burnett, D. D. (2003). A personality trait-based interactionist model of job performance. *Journal of Applied Psychology*, *88*(3), 500–517. doi:10.1037/0021-9010.88.3.500

Tett, R. P., & Guterman, H. A. (2000). Situation Trait Relevance, Trait Expression, and Cross-

Situational Consistency: Testing a Principle of Trait Activation. *Journal of Research in Personality*, *34*(4), 397–423. doi:10.1006/jrpe.2000.2292

Tett, R. P., Jackson, D. N., & Rothstein, M. (1991). Personality measures as predictors of job performance: A meta-analytic review. *Personnel Psychology*, *44*(4), 703–742.

Ury, W. (1991). *Getting Past No*. New York, NY: Random House LLC.

Viswesvaran, C., Ones, D. S., & Schmidt, F. L. (1996). Comparative analysis of the reliability of job performance ratings. *Journal of Applied Psychology*, *81*(5), 557-574. doi: 10.1037/0021-9010.81.6.660

Weiss, H., & Adler, S. (1984). Personality and Organizational Behavior. In B. M. Staw & L. L. Cummings (Eds.), *Research in Organizational Behavior* (Vol. 26, pp. 1–50). Greenwich, CT: JAI.

Wiggins, J. S., & Trapnell, P. D. (1996). A Dyadic-Interactional Perspective on the Five-Factor Model. In J. S. Wiggins (Ed.), *The Five-Factor Model of Personality: Theoretical Perspectives*. New York, NY: Guilford Pres