AN INVESTIGATION OF THE STABILITY OF ASSESSMENT CENTER

PERFORMANCE CONSISTENCY AND RATINGS OF

JOB PERFORMANCE

Except where reference is made to the works of others, the work described in this dissertation is my own or was done in collaboration with my advisory committee. This dissertation does not include proprietary or classified information.

_____
Julie M. Hetzler

Certificate of Approval:

| | |
|---|---|
| _____ | _____ |
| Bryan D. Edwards<br>Assistant Professor<br>Psychology | Daniel J. Svyantek, Chair<br>Professor<br>Psychology |
| _____ | _____ |
| L. Allison Jones-Farmer<br>Associate Professor<br>Management | Jacqueline K. Mitchelson<br>Assistant Professor<br>Psychology |

_____
George T. Flowers
Dean
Graduate School

# AN INVESTIGATION OF THE STABILITY OF ASSESSMENT CENTER PERFORMANCE CONSISTENCY AND RATINGS OF JOB PERFORMANCE

Julie M. Hetzler

A Dissertation

Submitted to

the Graduate Faculty of

Auburn University

in Partial Fulfillment of the

Requirements for the

Degree of

Doctor of Philosophy

Auburn, Alabama
August 10, 2009

AN INVESTIGATION OF THE STABILITY OF ASSESSMENT CENTER

PERFORMANCE CONSISTENCY AND RATINGS OF

JOB PERFORMANCE

Julie M. Hetzler



_____
Signature of Author

_____
Date of Graduation

DISSERTATION ABSTRACT

AN INVESTIGATION OF THE STABILITY OF ASSESSMENT CENTER PERFORMANCE CONSISTENCY AND RATINGS OF JOB PERFORMANCE

Julie M. Hetzler

Doctor of Philosophy, August, 10, 2009
(M.S., Auburn University, 2007)
(B.A., Emory University, 2004)

108 Typed Pages

Directed by Daniel J. Svyantek

The consistent lack of evidence for the construct validity of assessment center (AC) performance dimensions has led some researchers to propose alternative approaches to investigating candidate performance in ACs (Lance, 2008a; Lance et al., 2000). The lack of convergent and disciminant (i.e., construct) validity for AC performance dimensions has been referred to as "the construct validity problem" for ACs (Howard, 1997, p.21). The lack of construct validity evidence has been viewed as a problem by some researchers because the lack of convergent and discriminant validity for dimension ratings has traditionally been attributed to measurement error (Gibbons, 2007; Lance, 2008a; Lance et al., 2000). Recent research regarding the consistency of AC performance suggests that performance consistency may actually be a measurable individual difference where participants may reliably perform

differently in the individual exercises within ACs, also referred to as cross-situationally specific performance (Gibbons, 2007; Lance et al., 2000). Gibbons (2007) presented a performance consistency index to measure performance consistency in the context of an AC, which was adapted for the present study to measure dimension-level AC performance consistency. The results of the present study show that only overall consistency in an AC exhibits temporal stability in a sample of police officers who participated in a promotional AC at two separate points in time. Dimension-level AC performance consistency was not found to be stable over time. Additionally, it was found that AC performance consistency was not related to supervisor or peer ratings of job performance with the exception of a significant positive relationship found between overall AC consistency at time 2 and peer ratings of teamwork/cooperation. Lastly, supervisor ratings of job performance consistency were not found to be related to AC consistency at either time 1 or time 2.

ACKNOWLEDGEMENTS

I would like to thank my advisor and chair of this dissertation, Dan Svyantek, for his guidance and support during my years in the Auburn Psychology Department. Dan also deserves thanks from all of the I/O psychology students for taking a chance on Auburn and resurrecting our program. I would also like to thank Allison Jones-Farmer for providing me with an excellent foundation of statistical knowledge and inspiring me to be confident in my knowledge and ability. Thanks to Bryan Edwards and Jackie Mitchelson for providing me with constructive feedback and advice both on this dissertation and my future career as an I/O psychologist.

I would also like to thank my parents, Bill and Carol Hetzler, for supporting me throughout my education and enabling me to reach my goals. I have been able to achieve this dream because of my parents and their generosity and sacrifice. My friends deserve thanks as well for being good listeners and reminding me of life outside of graduate school. Finally, I would like to thank the Center for Business at Auburn Montgomery and Katherine Jackson for allowing me to have access to the archival data used for this study.

Style Guide: *Publication Manual of the American Psychological Association, Fifth Edition*

Software: *Statistical Package for the Social Sciences 17.0*

*AMOS 17.0*

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

# INTRODUCTION

The assessment center method has been implemented in a variety of organizations for more than fifty years and continues to be one of the most popular techniques for the selection/promotion, development, and diagnosis of training needs for employees (Guidelines and Ethical Considerations for Assessment Center Operations, 2000; Howard, 1997; Thornton & Byham, 1982; Thornton & Rupp, 2006). The continued use and popularity of assessment centers (ACs) is further indication of their utility for organizational selection, development, and diagnosis decisions. Despite the widespread popularity and implementation of ACs, there has been an ongoing debate in the literature regarding the validity of the performance dimensions measured in ACs, specifically, the construct validity (i.e., convergent and discriminant validity) of the dimensions measured by ACs (Lance, 2008a; Lance, Foster, Nemeth, Gentry, & Drollinger, 2007; Lance, Lambert, Gewin, Lievens, & Conway, 2004; Sackett & Dreher, 1982; Sackett & Tuzinski, 2001; Turnage & Muchinsky, 1982). Research on ACs consistently demonstrates that there is greater consistency on participant dimension scores within a single exercise in an AC than on a single dimension measured across multiple exercises.

Three distinct lines of research have evolved to attempt to account for the unexpected variability in both AC exercise and dimension scores that have contributed to the so-called "paradox" of AC construct validity (Gibbons, 2007; Lance, 2008a; Lance et

al., 2007). The first line of research focused on design flaws in ACs (Gaugler & Thornton, 1989; Lievens, 1998; Woehr & Arthur, 2003). Although design fixes such as reducing the number of dimensions rated and providing more extensive assessor training were found to improve AC construct validity, the findings were still inconsistent overall and did not result in substantially higher levels of construct validity evidence (Lance, Foster, Nemeth, Gentry, & Drollinger, 2007; Lievens, 1998; Woehr & Arthur, 2003). The second line of research has focused on characteristics of the individual exercises within an AC and specifically what traits should be activated within each exercise. This line of research combines both trait activation theory (Tett & Burnett, 2003) and research regarding the nature or types of exercises (Bem & Funder, 1978; Highhouse & Harris, 1993). Although both of these earlier lines of research resulted in improvements in AC construct validity, a more recent line of research has begun to attribute the low levels of AC construct validity to true candidate performance inconsistency across situations or exercises (Gibbons, 2007; Lance, 2008a; Lance et al. 2007; Lance, et al., 2004; Lievens, 2002). The present study seeks to contribute to this recent line of research which suggests that AC participants exhibit cross-situationally inconsistent behavior which may be both measurable and meaningful, as opposed to the previous conception that such inconsistencies were a result of measurement error.

Historically, industrial/organizational psychologists were not the first psychologists to research the consistency of human behavior across situations (Bem & Allen, 1974; Fleeson, 2001; Mischel, 1968). Personality psychologists have debated the inconsistency of personality in the person-situation debate with most personality psychologists now agreeing on the interactionist perspective where both the person and

the situation account for individual behavior. To this end, recent research in the field of personality psychology has begun to investigate the theory that behavioral consistency may be a measurable individual difference (Fleeson, 2001). Building on this research, Gibbons (2007) proposed and tested an index of performance consistency based on a candidate's performance within an AC. Gibbons (2007) found that candidates reliably differed in performance consistency and performance consistency was predictive of supervisor ratings of job performance. While Gibbons' research provided a theoretical and empirical foundation for viewing performance consistency as a measurable individual difference, additional research is needed to investigate performance consistency further. The present study builds on the initial work of Gibbons and provides three primary contributions in addition to replicating Gibbons' findings.

The present study measured the performance consistency of police officers in a promotional assessment center for the rank of police sergeant. Assessment center data were gathered for a large metropolitan police department in the Southeastern United States over the span of 7 years and 4 separate administrations. Using the performance consistency index established by Gibbons (2007), performance consistency was calculated for all candidates that participated in the sergeant's promotional AC at least twice. In order to establish the stability of performance consistency over time, the performance consistency indices for candidates were compared between the first and second time in which they participated in the AC. Additionally, supervisor and peer ratings of job performance were collected and the relationship between performance consistency from AC1 and AC2 and ratings of job performance was investigated. Lastly, supervisor and peer ratings of actual job performance consistency were also collected and

were compared to candidates' AC performance consistency to determine if AC performance consistency is significantly related to job performance consistency. In summary, the primary objectives of the present study are presented below:

1) Investigate the stability of individual performance consistency as measured in two separate promotional assessment centers.
2) Investigate the relationship between supervisor and peer ratings of job performance for candidates in two separate promotional assessment centers.
3) Investigate the relationship between AC performance consistency and supervisor ratings of job performance consistency.

The following section will summarize the relevant literature concerning the history of assessment centers, the definition/criteria for what constitutes an assessment center, and the empirical evidence supporting the validity of assessment centers. Additionally, a brief review of personality theory is included in order to provide an additional body of theoretical and empirical evidence to support the nature of behavioral consistency and inconsistency within different contexts as a measurable individual difference. The implications of performance consistency as an individual difference will be reviewed in terms of what organizations can expect from their organizational members' behavior and an alternative view of how measures of construct validity may be interpreted in assessment center contexts.

# LITERATURE REVIEW

## *Defining the Assessment Center Method*

The International Congress on Assessment Center Methods developed the first *Guidelines and Ethical Considerations for Assessment Center Operations* in 1975 (International Task Force, 2000). The *Guidelines* have since been updated to reflect the evolving technology and adaptability of the AC method. The most recent publication of the AC *Guidelines* defines an *assessment center* as "a standardized evaluation of behavior based on multiple inputs" (International Task Force, 2000, p.319). The AC *Guidelines* further specify ten criteria that must be met to use the term "assessment center", which ensures that the research and integrity associated with the AC method are not associated with less rigorous methods (Thornton & Rupp, 2006).

To summarize the essential criteria for an AC, the *Guidelines* require that an AC consist of behavioral judgments made from specifically developed simulations using multiple behavioral assessment techniques and those simulations must be developed based on a thorough job analysis. Also, assessees must be rated by more than one trained assessor on specific dimensions or competencies and the assessor ratings must be combined according to professional guidelines. Typical assessment devices in ACs include in-baskets, interviews, and leaderless group discussions (Gatewood & Feild, 2001; Thornton & Byham, 1982). A complete list of the *Guidelines* ten essential AC criteria can be found in Appendix A.

*History of the Assessment Center Method*

Prior to the appearance of assessment centers in the United States, German military psychologists were involved in the development of multiple-assessment techniques for the selection of German army, air force, and naval officers in the late 1930s (Thornton & Byham, 1982). The German military psychologists were the first to use multiple assessors and emphasize holistic measurements based on behavioral samples, which are three key elements present in modern ACs. Building on the early work of the Germans, the British War Office Selection Boards (WOSBs) began implementing assessment centers, as well, to select British military officers. Perhaps the most important contribution of the British version of the assessment center was their investigation of the psychometric properties of the method and their subsequent emphasis on performing validation studies (Thornton & Byham, 1982). In fact, the WOSBs were so successful that the British Civil Service Selection Board (CSSB) began to implement the same method to make promotion decisions, which was the first implementation of the AC method in a non-military setting.

The first site to implement the AC method in the United States was the Office of Strategic Services (OSS) which oversaw intelligence services during WWII (Moses & Byham, 1977). The OSS performed ACs for a relatively brief period of time, from 1943 to 1945, but Fiske and his colleagues are frequently credited with being the first Americans to implement the modern assessment center method (Bray, Campbell, & Grant, 1974; Bray & Grant, 1966; Fiske, Hanfmann, MacKinnon, Miller, & Murray, 1948; Moses & Byham, 1977; Thornton & Byham, 1982). The OSS was charged with selecting individuals that would become intelligence agents, “saboteurs”, and propaganda

experts during the war (Thornton & Byham, 1982, p.39). These were critical positions and the OSS recognized the need to implement a method beyond paper-and-pencil testing to observe candidates' behavior in a variety of situations. To that end, assessment center staff lived with the candidates to observe their behavior at all times (Fiske et al., 1948). Significant for the purposes of the present study, the early AC practitioners considered performance inconsistency in the various situations to be a meaningful characteristic of the individual which would be documented in the final summative report for that individual (Fiske et al., 1948; OSS Assessment Staff, 1948; Thornton & Byham, 1982). Fiske would later go on to study consistency as a stable trait and provided a foundation for current research regarding the conceptualization of consistency as an individual difference (Fiske, 1961; Fleeson, 2001). The concept that behavioral inconsistency between exercises is due to measurement error is a more recent development (Sackett & Dreher, 1982; Sackett & Tuzinski, 2001).

The first industrial application of the assessment center occurred in 1956 when Douglas Bray and the American Telephone and Telegraph Company (AT&T) started the Management Progress Study (MPS; Bray &Grant, 1966; Thornton & Byham, 1982). The MPS followed the careers of over 400 managers at AT&T for more than 25 years. The MPS was both a study of management potential as well as a study of the complete lives of the managers. The results of the MPS provided both qualitative and quantitative information regarding the predictive validity of ACs and the personal characteristics associated with successful managers (Bray & Campbell, 1968). The MPS continues to be frequently cited as evidence of the criterion-related validity of ACs.

Several themes emerge from the brief history of ACs presented here: a) early assessment center practitioners were interested in more holistic measures of individuals including situations where behaviors were expected to vary and b) early AC ratings consisted of *both* qualitative and quantitative measures. These early characteristics of ACs may provide some explanation for the on-going debate regarding the validity of AC performance dimensions. Historically, ACs have included a qualitative element which may be partially responsible for the difficulty in explaining the divide between how AC practitioners and researchers conceptualize the dimensions measured in ACs and what these dimensions are intended to measure (Howard, 1997; Lance, 2008a).

*The Paradox of Assessment Center Validity*

As mentioned previously, ACs are conducted for multiple purposes including selection/promotion, development, and diagnosis of training needs (Thornton & Rupp, 2006). Surveys of AC users have found that the majority (95.8%) of ACs are developed for selection and promotion (Gaugler, Rosenthal, Thornton, & Bentson, 1987; Spychalski, Quinones, Gaugler, & Pohley, 1997; Woehr & Arthur, 2003). Spychalski et al. (1997) performed the most recent survey of assessment center practitioners and discovered that more than 95% of the respondents utilized content validation as opposed to criterion or construct validation for their ACs. This finding is not surprising as the evidence of AC construct validity has been debated since the early 1980s, thus making practitioners more apprehensive of a validation strategy which may not produce favorable results (Neidig & Neidig, 1984; Sackett & Dreher, 1982; Sackett & Dreher, 1984).

Although the evidence of the construct validity of ACs has been questioned, much evidence exists in the form of criterion and content-related validation studies (Arthur,

Day, McNelly, & Edens, 2003; Gaugler et al., 1987; Howard, 1997; Thornton & Mueller-Hanson, 2004). The conflicting evidence of the criterion validity and construct validity of ACs has been labeled by some researchers as the "paradox" of AC validity (Lance et al., 2007). The following sections will provide a brief review of the concept of validity, a summary of the types of validity evidence that has been found for ACs, and an introduction to the current debate regarding the construct validity of the dimensions measured by ACs.

*Review of validity.* The assessment of validity evidence is one component of the psychometric quality of a test or measure. Gathering evidence of the validity of a measure ensures that the intended measurement target is actually being measured (Binning & Barrett, 1989; Crocker & Algina, 1986; Cronbach, 1990; Hunter & Hunter, 1984; Landy, 1986; Lawshe, 1985; Sackett, Schmitt, Ellingson, & Kabin, 2001; Schmidt & Hunter, 1998; Tenopyr, 1977). In assessment centers, the intended measurement targets are dimensions, such as problem analysis or leadership, which are measured in a variety of simulations (i.e., role play, leaderless group discussion, and in-basket). Binning and Barrett (1989) note that validity does not refer so much to the selection test itself, but to the validity of the inferences that can be made from a test. In the case of ACs, this would refer to the validity of the selection or promotion decisions an organization makes based on the results of the assessment center. The validity of these selection decisions is determined based upon evidence that is collected. Both the *Principles for the Validation and Use of Personnel Selection Procedures* (2003) and the *Standards for Educational and Psychological Testing* (1999) stipulate that evidence of validity can be collected in multiple forms, but the most common types of validation studies include criterion-related

validation, content validation, or construct-related validation. These three types of evidence of validity should not be confused as three different *types* of validity (Landy, 1986). The most current conceptualization of validity is the unitary theory which considers validity to be singular, but with multiple forms of evidence (e.g. criterion-relatedness, content relevance, or construct validity; Binning & Barrett, 1989). The following sections will include first the criterion-related validity evidence of ACs, then the content and construct validity evidence. This order of validity evidence parallels the historical development of AC validity, as well, which initially focused on criterion (i.e., predictive) validity, and concludes with the current debate regarding the construct validity of AC performance dimensions.

*Validity of Assessment Centers*

*Criterion-related validity.* The earliest evidence of assessment center validity came in the form of criterion-related validity. As mentioned previously, the MPS at AT&T provided the first evidence of the predictive validity of ACs (Cascio, 1998; Gatewood & Field, 2001; Thornton & Byham, 1982). Bray and his colleagues tracked a group of men as they progressed through their careers and linked their early AC ratings with criteria such as salary and management level attained (Bray & Grant, 1966; Thornton & Byham, 1982). They found that ACs were able to significantly predict which individuals would move into middle management positions 8 years after the AC ratings were obtained (Howard, 1997). The findings of the MPS are also significant as the AC ratings were not made available to the participants or their supervisors, thus, the criteria were not contaminated by prior knowledge of the participants' performance in the AC (Bray & Grant, 1966).

Additional evidence of the predictive validity of ACs comes from a frequently cited meta-analysis by Gaugler et al. (1987) where they obtained 107 validity coefficients from 50 different studies. Gaugler et al. (1987) found that the individual validity coefficients from the studies obtained for their meta-analysis ranged from -.25 to .78. After combining the results from the individual studies and correcting for sampling error, restriction of range, and criterion unreliability, they found the mean validity coefficient to be .37 for overall AC ratings. Also, Gaugler et al. (1987) found that validity coefficients differed depending upon the purposes of the AC. The validity of ACs for promotion was .30, but the validity coefficient for ACs used for initial selection was .41. They were also able to explain some of the variability in validity coefficients by differences between the implementation and design of ACs. Stronger validity coefficients were found in those ACs that had a wider variety of exercises, used psychologists as assessors, included peer evaluations, and had more women as assessees (Gaugler et al., 1987).

More recent research has been somewhat critical of the Gaugler et al. (1987) meta-analysis as they used overall ratings (OARs) from ACs as opposed to ratings obtained at the dimension level (Arthur et al., 2003). In an effort to investigate the validity of ACs at the dimension level, Arthur et al. (2003) performed a meta-analysis where results from 34 different studies were combined and 168 different dimensions across studies were consolidated into a set of 7 conceptually distinct dimensions. The validities for the separate dimensions ranged from .25 to .39. They also found that, collectively, 4 of the 7 dimensions of AC ratings accounted for 20% of the variance in job performance, which is more than the 14% of variance in performance accounted for by the Gaugler et al. (1987) study. Specifically, Arthur et al. (2003) found that the

dimensions of problem solving, influencing others, organizing/planning, and communication accounted for the criterion-related validity of ACs, while the dimensions of drive and consideration/awareness of others did not make a significant contribution. The most significant dimension identified by Arthur et al. (2003) was problem solving; problem solving alone was found to account for 15% of the variability in job performance.

The findings of Arthur et al. (2003) are consistent with previous studies which suggest that fewer dimensions be used in ACs in order to ensure higher validity (Jones & Whitmore, 1995; Lievens, 1998; Russell, 1985). In a review of the AC literature, Woehr and Arthur (2003) found that the mean number of AC dimensions measured was 10.60, but given the findings of Arthur et al. (2003), it is likely that an AC with 10 dimensions would contain some dimensions that are "dead weight" and are not accounting for additional variance in job performance. Also, there is additional research to suggest that a large number of AC dimensions may place too large a cognitive load on assessors to the extent that they cannot meaningfully distinguish between the dimensions (Gaugler & Rudolph, 1992; Gaugler & Thornton, 1989). These findings will be further discussed in a later section pertaining to the design of ACs.

Meriac, Hoffman, Woehr, and Fleisher (2008) recently extended the findings of Arthur et al. (2003). Meriac et al. (2008) performed a meta-analysis of the incremental criterion-related validities of AC dimension ratings and found that 6 of the 7 dimensions proposed by Arthur et al. (2003) accounted for a significant amount of the variance in job performance. Although Arthur et al. (2003) and Meriac et al. (2008) found different results regarding which individual dimensions were valid predictors of job performance,

taken together, their findings provide a significant contribution to the body of research regarding AC validity. Specifically, these researchers have provided a taxonomy of AC dimensions and investigated which dimensions accounted for the most variance in job performance. Additionally, their studies provide direction for future researchers to focus on the validity of individual AC dimensions as opposed to referring generically to overall AC validity.

*Temporal stability of AC criterion-related validity.* There has been some research to suggest that the predictive validity of ACs may change over time (Jansen & Stoop, 2001; Bray, Campbell, & Grant, 1974). The criteria that ACs predict are often long-term criteria such as management potential and career advancement. Results from the MPS at AT&T showed that the predictive validity of the OARs for management level attained decreased from .46 in the years just after the AC to .33 sixteen years later (Thornton & Byham, 1982). Tziner, Ronen, and Hacohen (1993) also found that the predictive validity of OARs for upper-level management potential decreased over time while other studies have found that the predictive validity of AC ratings increased over time (Hinrichs, 1978; McEvoy & Beatty, 1989).

In order to further investigate the temporal stability of the criterion-related validity of ACs, Jansen and Stoop (2001) tracked a group of 679 recruits for an organization in the Netherlands who participated in an AC as part of their selection process. Over a 7-year period, they used OARs from the AC to predict salary growth and found that the validity of the OARs was .39 at the end of 7 years. Jansen and Stoop (2001) also found that the validity of OARs diminished between years 3 and 5, but then increased in years 6 and 7. A more recent study by Jansen and Vinkenburg (2006) looked

at OAR predictive validity over a 13-year period and found similar variability in their results. Thus, multiple studies have established that the criterion-related validity of ACs is dynamic.

*Content-oriented validity.* While the majority of published validation studies on ACs investigate criterion-related and construct-related validity, Sackett (1987) notes that content validation is usually the most viable validation strategy for assessment centers in organizational settings. Content-related validity refers to content-related evidence that a "measurement procedure contains a fair sample of the universe of situations it is supposed to represent" (Cascio, 1998, p.101). Since the establishment of the 1978 *Uniform Guidelines on Employee Selection Procedures,* content validation has become increasingly popular. In fact, Spychalski et al. (1997) found that an overwhelming majority (95.8%) of AC practitioners chose content validation as their validation strategy.

Although content validation may be more practically feasible for organizations, content-related validation is not appropriate for all ACs (Sackett, 1987). Content-related validity is most appropriate for ACs that intend to sample work behaviors so that candidates can be successful on the job right away (Dreher & Sackett, 1981; Sackett, 1987; Wernimont & Campbell, 1968). If the purpose of the AC is to identify potential and/or predict future job performance, then a criterion-related validation strategy would be more appropriate (Thornton & Rupp, 2006). Content validation is also sometimes chosen because other validation strategies are not feasible (Sackett, 1987). Criterion-related validation requires a large sample size and may last for a long period of time, which can be costly for organizations. Also, criterion-related validation or construct validation may produce the unfavorable result that the AC is not valid, which is

something that many practitioners are unwilling to risk, considering the substantial costs of developing and administering ACs (Sackett, 1987; Thornton & Rupp, 2006).

Unlike criterion-related validity and construct validity, there are no quantitative, definitive methods for establishing content validity. Goldstein, Zedeck, and Schneider (1993) point out that content validation involves inherently qualitative data and more research is needed to identify the judgmental processes that subject matter experts (SMEs) undergo when helping to link KSAs to work behaviors, work behaviors to AC dimensions, and AC dimensions to job performance. The greater the inferential "leaps" in this process, the less fidelity and validity an AC is likely to have (Dreher & Sackett, 1981; Goldstein et al., 1993; Sackett, 1987; Tenopyr, 1977).

The significant amount of research cited above regarding the criterion and content-related validity evidence of ACs confirms why practitioners support and defend the validity of ACs. It is also noteworthy that most practitioners use a content validation strategy. Since content validation is legally defensible and ACs have a high degree of face validity, practitioners have not been as concerned about the so-called paradox of AC validity (Howard, 1997, 2008). The following section summarizes the current debate regarding the construct validity of ACs and is followed by recent research proposing alternative interpretations of candidate performance consistency in ACs.

*Construct validity.* Perhaps the most contentious issue concerning ACs at the present time is the lack of evidence of construct-related validity for the dimensions measured in ACs (Bowler & Woehr, 2006; Lance et al., 2000; Lievens, 2002; Lievens & Klimoski, 2001; Sackett & Dreher, 1982; Sackett & Tuzinski, 2001). As stated previously, the accumulation of evidence of the criterion-related validity of ACs and the

comparative lack of evidence of construct validity of AC dimensions has been referred to as the construct validity paradox of assessment centers (Lance, Foster, Gentry, & Thoreson, 2004). ACs are designed to measure specific behavioral dimensions, which are assessed in a variety of simulations or exercises. The dimensions assessed in ACs are considered to be constructs. Thornton and Byham defined a behavioral dimension as "a cluster of behaviors that are specific, observable, and verifiable, and that can be reliably and logically classified together" (Thornton & Byham, 1982, p.117). It is important to note that Thornton and Byham's definition clearly states that dimensions are *behaviorally* based and not trait based.

One common method for determining the construct validity of AC dimensions involves the assessment of both the convergent and discriminant validity of the dimension ratings made both within and across exercises often using the multitrait-multimethod matrix proposed by Campbell and Fiske (1959). Since ACs consist of multiple exercises containing multiple dimensions, the correlations between the same dimensions across exercises and different dimensions within exercises are assessed to determine the degree of convergent and discriminant validity (Arthur, Woehr, & Maldegen, 2000). Since the same dimensions are typically measured in multiple exercises, some researchers have argued that these dimensions should be highly correlated if they are supposed to represent a single construct (Klimoski & Brickner, 1987; Lance et al., 2000). For example, one might expect that if a candidate scores highly on oral communication for a role-play exercise, then that candidate should also score highly on oral communication for an interview. In turn, one would expect that the correlations of the different dimensions measured within exercises would be lower than

the correlations of the same dimension measured across exercises. This pattern of relationships would establish the assessors' ability to distinguish between dimensions within exercises and consistently measure candidates' performance across exercises. Conversely, the majority of construct validation studies show that the correlations of dimensions within exercises are consistently stronger than dimensions across exercises (Bowler & Woehr, 2006; Howard, 1997, Lievens & Conway, 2001; Lievens, 2001a, 2001b, 2002; Sackett & Dreher, 1982; Schneider & Schmitt, 1992; Woehr & Arthur, 2003).

There are several different sources that may be contributing to the lack of convergent and discriminant validity of AC ratings (Thornton, 1992; Lievens, 1998). Woehr and Arthur (2003) provide a thorough review of the literature on AC construct validity. In their review, they grouped the literature into two primary categories: methodological/design characteristics and construct misspecification. Relevant research in each of these areas will be briefly discussed in the following section.

There is a large body of literature that focuses on the methodological factors that may be manipulated to enhance the construct validity of AC ratings. Lievens (1998) grouped these design issues into categories according to differences in dimensions, assessors, observations, and ratings. A recent meta-analysis by Woehr and Arthur (2003) found construct validity was enhanced when there are fewer dimensions, dimensions are well-defined (Bycio, Alvares, & Hahn, 1987; Gaugler & Thornton, 1989; Joyce, Thayer, & Pond, 1994; Meriac et al., 2008), psychologists are used as assessors (Gaugler et al., 1987; Sagie & Magnezy, 1997), frame-of-reference training is provided for assessors (Schleicher, Day, Mayes, & Riggio, 2002; Woehr & Huffcut, 1994), and there is a small

ratio of assessees to assessors (Bycio et al., 1987). Research regarding "design fixes" for ACs has resulted in slightly improved estimates of convergent and discriminant validity of AC dimensions, but the more researchers have fine-tuned their measurements, the more apparent it has become that assessor ratings of candidate performance in ACs represent cross-situationally specific candidate behavior (Lance, 2008b; Lance et al., 2007).

Recently, trait activation theory (TAT) has been offered as one explanation for why candidate performance may vary across exercises (Haaland & Christiansen, 2002; Lievens, Chasteen, Day, & Christiansen, 2006; Tett & Burnett, 2003). TAT involves a person-situation interactionist perspective which provides a theoretical explanation to predict how individuals will perform in certain situations based on what traits are activated by the situation (Bush, 2003). Lievens et al. (2006) found that convergence was better for those exercises that elicited behavior relating to similar traits. Also, they found that TAT works best for the extraversion and conscientiousness traits and suggested that this might be because these traits are observed more easily by assessors.

In addition to research investigating design improvements for ACs and TAT, another line of research exists which suggests that ACs may not be measuring the dimensions that practitioners claim they are measuring (Lance, 2008a). This line of research has been termed "construct misspecification" (Woehr & Arthur, 2003). Researchers fear that the construct misspecification of ACs may result in negative consequences that practitioners have not considered (Russell & Domm, 1995; Woehr & Arthur, 2003). Specifically, that assessment centers might be measuring dimensions or constructs other than those that they identified and selected based on the job analysis.

This construct misspecification would not be as problematic for selection purposes because whatever construct is being measured may still be predictive of job performance; however, it may create problems for developmental ACs as assessees might be given inaccurate feedback (Lance et al., 2000). An alternative explanation for the misspecification of the construct involves the misinterpretation of the dimensions. Lance et al. (2000) posit that the various exercises in ACs represent different situations and the resulting ratings represent cross-situational specificity not assessor inaccuracy, as mentioned previously. Thus, the ratings might represent true performance differences across exercises (Haaland & Christiansen, 2002; Lievens, 2001, 2002). Given that candidate performance in ACs may be cross-situationally specific, Lance et al. (2000) argue that the application of the MTMM design may be inappropriate for measuring the construct validity of AC dimensions as the MTMM design was originally developed to measure traits which were thought to be cross-situationally stable (Campbell & Fiske, 1959; Howard, 2008). As mentioned previously, Thornton and Byham's definition emphasizes that AC dimensions measure "observable behaviors", not traits (Howard, 2008, p.99).

To further investigate the internal structure of ACs, Lance, Lambert, et al. (2004) found that ACs consist of 2 primary factors: a general performance factor (GP) that represents cross-situationally consistent behavior and an exercise factor that represents situationally specific or inconsistent behavior. Lance et al. (2007) found that personality, specifically the dimensions of conscientiousness, emotional stability, and openness to experience, was significantly related to the GP factor. Additionally, Lance et al. suggest that the exercise-based performance factor is most likely due to task characteristics or

unique aspects of the individual simulations within an AC. The present study seeks to further account for what Lance and colleagues have termed the "exercise" factor and propose that cross-situational specificity may vary meaningfully between candidates and may be a measurable, stable, individual difference.

*Personality Theory and Behavioral Consistency*

The debate regarding the nature of individual consistency is not unique to the field of industrial/organizational psychology or even the AC method. Personality psychologists have debated for years regarding the existence of traits and whether it was the person or the situation that accounted for behavior (Epstein & O'Brien, 1985; Mischel, 1968, 2004). The person-situation debate was eventually resolved in the personality psychology literature with most personality researchers currently agreeing on the interactionist perspective with both the person and the situation accounting for individual behavior (Shoda, 1999, 2003). Although most psychologists currently agree on the interactionist perspective, some trait-based research continues (McCrae et al., 2000), while other researchers have elected to focus on the situation and ways to categorize situations (Bem & Funder, 1978; Kendrick & Funder, 1988; Mischel, 1973). Mischel (1973) proposed that there are two primary types of situations: strong and weak situations. According to Mischel, strong situations result in the same response from most individuals due to normative expectations of behavior. In contrast, weak situations contain fewer normative behavioral cues and allow for the expression of individual differences (Mischel, 1973). Thornton and Mueller-Hanson (2004) suggest that the exercises included in ACs be weak enough to allow for varied candidate responses.

Following the work of personality researchers, some AC research has investigated similarities between the various situations included in ACs. Highhouse and Harris (1993) applied Bem and Funder's (1978) template matching technique to exercises in ACs to determine if exercise similarity could account for AC candidate performance inconsistency. They found some evidence that candidates performed more consistently in exercises that were perceived to be more similar by assessors. Brannick, Michaels, and Baker (1989) also investigated whether situational similarity could account for performance inconsistency in an AC. They developed parallel forms of an in-basket exercise to see if candidate performance would be consistent in situations that were developed to be as similar as possible. In contrast to the findings of Highhouse and Harris (1993), they found weak evidence of convergent and divergent validity between candidate scores on the two in-basket exercises. Thus, research investigating situational explanations of candidate performance inconsistency in ACs has not resolved the construct-validity paradox.

*Consistency as an individual difference.* Bem and Allen (1974) were among the first researchers to investigate consistency as an individual difference. They theorized that individuals should only be consistent on those traits with which people identify themselves. Thus, Bem and Allen (1974) focused on consistency only as a means of identifying who should be consistent on particular traits, or as they stated, "predicting some of the people some of the time" (p.517). More recently, research in the personality literature provides evidence that consistency may be an individual trait itself. Shoda, Mischel, and Wright (1994) tracked the behavior of boys at a summer camp and subsequently developed profiles of behavior for each boy based on the situation. Their

results allowed them to predict which boys would respond to criticism from certain individuals.

Following the work of Shoda et al. (1994), Fleeson (2001) tracked the personality and affect reported by individuals over a period of several weeks and found that the standard deviations for each individual's behavior distribution were fairly stable. Fleeson's research is particularly significant as he is one of the few personality researchers to emphasize the importance of the entire range of behavioral ratings as opposed to simply taking the average of a set of ratings. He saw behavioral consistency as an important measure itself, which in the context of the present study may well be related to significant outcomes in the workplace.

Gibbons (2007) provided the most recent evidence that performance consistency is a measurable individual difference. Gibbons presented a series of studies that demonstrated that individuals reliably differ in levels of consistency. Additionally, Gibbons linked consistency to team performance in both an athletic team sample and an operational AC sample. Perhaps most importantly, Gibbons proposed and tested an index to measure consistency in an AC context. Taken together, the work of Shoda, Mischel, and Wright (1994), Fleeson (2001), and Gibbons (2007) support the proposition that performance consistency is a stable individual difference and that performance consistency is related to outcomes such as team performance and supervisory performance ratings.

*Hypothesis 1: Candidate AC performance consistency will be stable over time such that candidate performance consistency at AC1 will be significantly positively related to candidate performance consistency in AC2.*

*Linking Assessment Center Performance Consistency to Job Performance*

In order for AC practitioners to implement consistency metrics in their ACs, consistency must be first be found to relate to job performance. AC performance dimensions are said to be valid due to their relation with job performance as evidenced by both content and criterion validity; however, if consistency in an AC is to provide incremental validity and be considered a unique dimension of AC performance, AC consistency must be found to be related to job performance above and beyond what is already accounted for by the mean dimension scores and overall AC scores (Gibbons & Rupp, 2009).

In order to provide preliminary information regarding the AC consistency-job performance link, both supervisor and peer ratings of AC participant job performance were obtained for the purposes of the present study. A large body of literature exists on multisource job performance ratings. The following section reviews this literature as it pertains to the present study.

*Multisource ratings of job performance.* . The performance appraisal literature shows that peer and supervisor ratings of job performance differ significantly in the context of multisource and 360 feedback appraisals (Conway & Huffcutt, 1997; Mount et al., 1998; Murphy & Cleveland, 1991). In a meta-analysis comparing multisource performance ratings, Conway and Huffcutt found that peer and supervisor ratings of job performance had the highest degree of convergence when comparing between self, peer, supervisor, and subordinate pairs of ratings ($r$ = .34; 1997). Although peer and supervisor ratings of job performance have been found to have the highest level of agreement, significant differences persist between the two different groups of raters. Thus, for the

purposes of the present study, peer and supervisor ratings of job performance will be treated as separate sets of dependent variables.

*Present Study*

The present study proposes a departure from previous research on assessment centers which, as summarized previously, has focused largely on design fixes and the construct validity of the dimensions measured in ACs. In turn, the present study seeks to build on the work of Gibbons (2007) and Fleeson (2001) to demonstrate that inconsistent candidate scores in AC dimensions reflect true performance differences. Although Lance and colleagues have previously made this suggestion, Gibbons (2007) recently developed an index to allow for further investigation of performance consistency in the context of an AC.

The present study provides a unique contribution to the existing literature by testing the stability of performance consistency in an AC for candidates that participated in the same AC at two separate points in time. Additionally, supervisor and peer performance ratings were obtained to investigate to what degree performance consistency can account for variance in others ratings of job performance. The present study will analyze data from 4 operational assessment centers for police officer promotions in a large, metropolitan police department in the southeastern United States. Although no formal hypotheses are made regarding AC performance consistency on the dimension level, the relationships between AC dimension performance consistency and job performance on the dimension level will be investigated, as well.

*Hypothesis 2a: Candidate performance consistency as measured in AC1 and AC2 will account for significant variance in supervisor ratings of customer service.*

*Hypothesis 2b: Candidate performance consistency as measured in AC1 and AC2 will account for significant variance in supervisor ratings of quantity of work.*

*Hypothesis 2c: Candidate performance consistency as measured in AC1 and AC2 will account for significant variance in supervisor ratings of quality of work.*

*Hypothesis 2d: Candidate performance consistency as measured in AC1 and AC2 will account for significant variance in supervisor ratings of job knowledge.*

*Hypothesis 2e: Candidate performance consistency as measured in AC1 and AC2 will account for significant variance in supervisor ratings of teamwork/cooperation.*

*Hypothesis 3a: Candidate performance consistency as measured in AC1 and AC2 will account for significant variance in peer ratings of customer service.*

*Hypothesis 3b: Candidate performance consistency as measured in AC1 and AC2 will account for significant variance in peer ratings of quantity of work.*

*Hypothesis 3c: Candidate performance consistency as measured in AC1 and AC2 will account for significant variance in peer ratings of quality of work.*

*Hypothesis 3d: Candidate performance consistency as measured in AC1 and AC2 will account for significant variance in peer ratings of job knowledge.*

*Hypothesis 3e: Candidate performance consistency as measured in AC1 and AC2 will account for significant variance in peer ratings of teamwork/cooperation.*

In addition to linking AC performance consistency to supervisor and peer ratings of job performance, further validation of AC consistency is needed to determine if there is relationship between consistency in an assessment center and consistency while on the job. In order to investigate this relationship, supervisor ratings of job performance consistency were obtained.

*Hypothesis 4: Supervisor ratings of candidate job performance consistency will be significantly related to candidate performance consistency in AC1 and AC2.*

*Summary of Hypotheses*

*Hypothesis 1: Candidate AC performance consistency will be stable over time such that candidate performance consistency at AC1 will be significantly positively related to candidate performance consistency in AC2.*

*Hypothesis 2a: Candidate performance consistency as measured in AC1 and AC2 will account for significant variance in supervisor ratings of customer service.*

*Hypothesis 2b: Candidate performance consistency as measured in AC1 and AC2 will account for significant variance in supervisor ratings of quantity of work.*

*Hypothesis 2c: Candidate performance consistency as measured in AC1 and AC2 will account for significant variance in supervisor ratings of quality of work.*

*Hypothesis 2d: Candidate performance consistency as measured in AC1 and AC2 will account for significant variance in supervisor ratings of job knowledge.*

*Hypothesis 2e: Candidate performance consistency as measured in AC1 and AC2 will account for significant variance in supervisor ratings of teamwork/cooperation.*

*Hypothesis 3a: Candidate performance consistency as measured in AC1 and AC2 will account for significant variance in peer ratings of customer service.*

*Hypothesis 3b: Candidate performance consistency as measured in AC1 and AC2 will account for significant variance in peer ratings of quantity of work.*

*Hypothesis 3c: Candidate performance consistency as measured in AC1 and AC2 will account for significant variance in peer ratings of quality of work.*

*Hypothesis 3d: Candidate performance consistency as measured in AC1 and AC2 will account for significant variance in peer ratings of job knowledge.*

*Hypothesis 3e: Candidate performance consistency as measured in AC1 and AC2 will account for significant variance in peer ratings of teamwork/cooperation.*

*Hypothesis 4: Supervisor ratings of candidate job performance consistency will be significantly related to candidate performance consistency in AC1 and AC2.*

# METHOD

## *Participants*

Participants were police officers in a metropolitan police department in the Southeastern United States. The officers participated in a promotional assessment center for the rank of sergeant. AC scores were obtained for 4 separate AC administrations over the span of 7 years within the same police department. Data were screened to include only those participants that repeated the AC ($n$ = 162). The screened sample of repeat AC participants consisted of 59.3% African American officers ($n$ = 96), 38.9% Caucasian officers ($n$ = 63), and 1.9% of the officers were other races ($n$ = 3). The screened sample was 90.1% male ($n$ = 146) and 9.9% female ($n$ = 16). The racial and gender composition of the screened sample was similar to the police department as a whole, which is predominantly African American and male.

## *Procedure*

The AC scores for the police officers were provided by a consulting group operating within a large university in the southeastern United States. As mentioned previously, scores were provided for participants in 4 separate AC administrations over the span of 7 years within the same police department. Although the data for the present study were candidates for promotion to the rank of sergeant, the consulting group also administered ACs for the ranks of lieutenant and captain in the same police department. The candidates participating in the lieutenant and captain ACs included the peers and

supervisors of the candidates participating in the sergeant ACs. In order to obtain ratings of job performance, candidates for the lieutenant and captain ACs were given the opportunity to voluntarily provide ratings of job performance for their peers and subordinates that had previously participated in an AC. The peer and supervisor job performance ratings were obtained from current police sergeants and lieutenants once they had completed participating in their respective AC (either for promotion to lieutenant or captain). Participants were asked to rate only those officers with whom they had worked for a sufficient period of time to view the officer's performance. The ratings were anonymous and could not be linked back to the rater. The job performance rating scale consisted of performance dimensions currently assessed in the performance appraisal tool used by the police department. The job performance dimensions measured included customer service, quantity of work, quality of work, job knowledge, and teamwork/cooperation. Additionally, the dimension of 'consistency of work' was added to allow for comparisons between AC performance consistency and job performance consistency. The job performance scale can be found in Appendix B.

*Assessment Center Exercise Development and Administration*

The following section will describe the development and administration of the AC exercises by the consulting group. The sergeant promotional ACs in 3 of the 4 administrations (administrations 2, 3, and 4) were based on the same job analysis. The sergeant AC from the first administration was based on a slightly different job analysis, although the content of the target position (rank of sergeant) was very similar between the first administration and the later administrations. The promotional AC for the rank of sergeant consisted of an initial multiple-choice job knowledge exam followed by

structured interview questions. After completion of the job knowledge exam and structured interview questions, the candidate scores were banded and only the highest two scoring bands went on to participate in a set of 2 role plays. For the purposes of the present study, only the candidate scores for the structured interview portion of the AC will be analyzed. Therefore, only the development and administration of the structured interview component of the AC will be included.

*Development of the structured interview questions.* A comprehensive job analysis was conducted with incumbent police sergeants (also referred to as subject matter experts or SMEs) and ratings of importance and necessity at entry were used to determine which knowledges, skills, and abilities (KSAs) were eligible for testing in each AC. Qualifying KSAs were then grouped based on similarity into performance dimensions by SMEs. The KSAs were grouped into 5 performance dimensions: problem analysis, management ability, supervisory ability, technical and departmental knowledge, and oral communication. Each of these performance dimensions was measured in multiple structured interview questions. The operational definitions of the performance dimensions can be found in Appendix C as well as a matrix of which dimensions were measured in each structured interview question by administration.

The structured interview questions were developed using small groups of SMEs ($n$ = 6 per group) to generate examples of critical incidents from each individual's personal experience as a sergeant. Two groups of SMEs were convened to develop the questions and response guidelines for each question. The SMEs were provided with a list of the KSAs that were eligible for testing in order to generate critical incidents where those KSAs would be needed. The response guidelines included 3 primary rating

categories scored on a 7-point likert-type scale: Clearly Unacceptable (1), Clearly Acceptable (3), and Clearly Superior (7) with scores between the primary categories to allow for finer distinction of candidate performance. Each category was anchored by examples of behaviors or responses generated by the SMEs.

*Validation of structured interview questions.* A content validation process was followed for the structured interview questions. The SMEs that participated in the development of the questions and response guidelines rated the questions to ensure that the questions met the following criteria: (a) job-related, (b) distinguished between high and low performing police sergeants, (c) quality of the question (i.e., was the question too easy, too difficult, ambiguous, biased), (d) the questions measured the performance dimensions and underlying KSAs that they were designed to measure. Additionally, all questions and response guidelines were reviewed by higher ranking police officers to ensure accuracy and consistency with department policy.

*Assessor training.* Assessors at the rank of police sergeant or higher were recruited from departments across the United States. Assessors participated in 1 day of assessor training (approximately 7 hours) prior to the administration of the AC. During training, assessors were informed of common rating errors and provided with examples of such ratings errors (e.g., halo, central tendency, etc). Additionally, assessors were instructed on how to accurately record candidate behavior and were encouraged to only record observable behaviors. Assessors were then provided with examples of structured interview questions and were allowed to rate responses provided by actors as opposed to actual police officers. The consultants ensured that all assessors consistently provided accurate ratings before dismissing the assessors from training.

*Administration of structured interview questions.* Candidates were given the opportunity to attend an information session regarding the AC several weeks prior to the administration. On the day of the administration, candidates were taken to a room with a panel of two assessors representing both racial and gender diversity (one male, one female, one Caucasian, one African American). Candidates were given 8 minutes to prepare their response and an additional 8 minutes to present their response. Candidates followed this same procedure until they had responded to all structured interview questions. The assessors rated each candidate's response independently and then came to a consensus rating where their individual ratings were within 1 point for each dimension rating. Candidates were rated by different panels of assessors for each structured interview question. The average of the individual assessor's ratings was calculated to produce the resulting dimension scores for each structured interview question.

*Analysis*

*Calculating consistency.* Presented below is the consistency index developed and validated by Gibbons (2007) to measure performance consistency in an AC:

$$C = 1 - \frac{\overline{cik}}{R}$$

The $\frac{\overline{cik}}{R}$ term represents the average of score differences from every possible pairing of exercises measuring each performance dimension divided by the range of the rating scale. For the present study, the rating scale ranged from 1 to 7, therefore $R = 6$. The value of this term is then subtracted from 1 so that higher values of $C$ indicate higher levels of consistency. Consistency will first be calculated for each performance dimension and then will be aggregated across all dimensions to reflect total performance consistency in

the AC. See Figure 1 in Appendix D for a more complete explanation of how consistency will be calculated and operationalized for the present study and a comparison of how consistency was originally proposed and measured by Gibbons (2007).

Consistency indices will be calculated for all participants for both of the ACs in which they participated. Once the consistency indices are calculated, the hypotheses can then be tested. The hypotheses proposed for the present study are presented below followed by the proposed analysis to test each hypothesis or set of hypotheses.

*Hypothesis 1: Candidate AC performance consistency will be stable over time such that candidate performance consistency at AC1 will not differ significantly from candidate performance consistency in AC2.*

A model will be tested with AC performance consistency in both ACs modeled as separate factors ($C_1$ and $C_2$). The dimension-level consistencies will be included as indicators of the latent consistency construct for both $C_1$ and $C_2$. The model to test H1 is included as Figure 2 in Appendix E.

*Hypothesis 2a: Candidate performance consistency as measured in AC1 and AC2 will account for significant variance in supervisor ratings of customer service.*

*Hypothesis 2b: Candidate performance consistency as measured in AC1 and AC2 will account for significant variance in supervisor ratings of quantity of work.*

*Hypothesis 2c: Candidate performance consistency as measured in AC1 and AC2 will account for significant variance in supervisor ratings of quality of work.*

*Hypothesis 2d: Candidate performance consistency as measured in AC1 and AC2 will account for significant variance in supervisor ratings of job knowledge.*

*Hypothesis 2e: Candidate performance consistency as measured in AC1 and AC2 will account for significant variance in supervisor ratings of teamwork/cooperation.*

*Hypothesis 3a: Candidate performance consistency as measured in AC1 and AC2 will account for significant variance in peer ratings of customer service.*

*Hypothesis 3b: Candidate performance consistency as measured in AC1 and AC2 will account for significant variance in peer ratings of quantity of work.*

*Hypothesis 3c: Candidate performance consistency as measured in AC1 and AC2 will account for significant variance in peer ratings of quality of work.*

*Hypothesis 3d: Candidate performance consistency as measured in AC1 and AC2 will account for significant variance in peer ratings of job knowledge.*

*Hypothesis 3e: Candidate performance consistency as measured in AC1 and AC2 will account for significant variance in peer ratings of teamwork/cooperation.*

In order to test hypotheses 2a through 2e and 3a through 3e, job performance will be incorporated into the model with $C_1$ and $C_2$. The preliminary model to test hypotheses 2a through 2e and 3a through 3e is included as Figure 3 in Appendix E. For simplicity, job performance is represented at the global level, although the hypotheses are stated at the dimension level.

*Hypothesis 4: Supervisor ratings of candidate job performance consistency will be significantly related to candidate performance consistency in AC1 and AC2.*

Hypothesis 4 will be tested using Pearson's correlation coefficient to determine if the relationship between job performance consistency and AC performance consistency is statistically significant.

# RESULTS

*Descriptive Statistics and Preliminary Analyses*

Consistency indices were calculated for the AC performance dimensions including problem analysis (C-PA), management ability(C-MA), technical and departmental knowledge (C-TDK), and oral communication (C-OC). The supervisory ability dimension was dropped as that dimension was only measured once during the most recent administration which prevented calculation of consistency indices. Thus comparison of consistency for supervisory ability consistency was not possible. All dimension level consistency indices were transformed into Z scores and values exceeding +/-3 were removed from subsequent analyses. Consistency indices were transformed back into non-standardized units for subsequent analyses. After removal of outliers ($n$ = 9), the final sample for analysis consisted of 153 officers. Subsequent analyses were run with and without outliers and the removed outliers were not found to alter the significance of the obtained results. The sample included 91 (59.5%) African American officers, 59 (38.6%) Caucasian officers, and 3 officers who did not report a race. The sample also consisted of 137 (89.5%) males and 16 (10.5%) females. Table 1 contains means, standard deviations, and intercorrelations for the AC dimension consistency indices and overall consistencies for AC1 and AC2. Overall consistency at time 1 and time 2 was obtained by summing the dimension-level consistency indices for first and second AC administrations in which the officers participated.

Although the proposed analysis section stated that structural equation modeling (SEM) would be used to test the models presented in Figure 2 and Figure 3, the correlations between the AC performance dimension consistency indices were not statistically significant and thus it was not appropriate to test the model presented in Figure 2. As the model proposed in Figure 3 built upon the model in Figure 2, the model proposed in Figure 3 could not be tested using SEM either. Thus, the proposed hypotheses will be tested using alternative analyses.

*Hypotheses Testing*

Hypothesis 1 was tested with Pearson's correlation coefficient. The correlation between AC consistency at time 1 and AC consistency at time 2 was statistically significant and thus hypothesis 1 was supported (r =.44, $p$ <.001). To further investigate hypothesis 1, the correlations of the individual dimension-level consistency indices between time 1 and time 2 were also obtained and can be found in Table 1 in Appendix F. Although the aggregated consistency indices between time 1 and time 2 were significantly related, none of the dimension-level consistency indices were correlated between time 1 and time 2. Also worth noting, within both time 1 and time 2 the consistency indices for the problem analysis and technical departmental knowledge dimensions were significantly related ($r$ = .18, $p$ = .03; $r$ = .27, $p$ = .001). The implications of these results will be included in the Discussion section.

Hypotheses 2a-2e proposed that AC consistency at time 1 and time 2 would relate to supervisor ratings of job performance. Hypotheses 3a-3e included the proposed relationships between AC consistency and peer ratings of job performance. Table 2a

contains the means, standard deviations, and intercorrelations for the AC dimension and overall consistency indices at time 1 and time 2 and supervisor ratings of job performance. Table 2b contains the means, standard deviations, and intercorrelations for the AC dimension and overall consistency indices at time 1 and time 2 and peer ratings of job performance. Hypotheses 2a-2e were not supported as evidenced by the non-significant correlation coefficients found in Table 2a. Although hypotheses 2a-2e were stated on the level of overall consistency, the correlations between the dimension-level consistencies and supervisor ratings of job performance were also obtained. The pattern of relationships between the dimension-level consistency indices and the supervisor ratings of job performance were in the opposite direction that what would be expected. Supervisor ratings of job performance were most significantly related to the problem analysis consistency indices both at time 1 and at time 2, however, the sign of the correlation coefficient changes from positive to negative from time 1 to time 2. The implications of this effect will be included in the discussion.

Hypotheses 3a, 3b, 3c, and 3e were not supported. Hypothesis 3d was the only hypothesis linking overall AC consistency to peer ratings of job performance that was partially supported. Specifically, hypothesis 3d linked AC consistency to peer ratings of teamwork/cooperation. Hypothesis 3d was only partially supported because the correlation was only statistically significant at time 2 ($r = .27$, $p = .02$). The dimension-level consistency indices were also correlated with the peer ratings of job performance and no patterns were observed. Only problem analysis consistency and management ability consistency both at time 2 were significantly related to customer service and job knowledge respectively ($r = .29$, $p = .01$; $r = .30$, $p = .01$). Hypothesis 4 was not

supported, as well. The correlation coefficients were non-significant between the AC consistency indices at time 1, time 2, and supervisor ratings of job performance consistency ($r = .06$, $p = .52$; $r = .14$, $p = .15$).

*Additional Analyses*

Due to the lack of support for the relationship between AC performance consistency and ratings of job performance, additional analyses were conducted to provide further information regarding the nature of the relationship between AC performance consistency and other measures of variability, the reliability of the AC itself, and the relationship between the mean-level AC dimension scores and overall scores and ratings of job performance. Each of these analyses will be discussed individually in the following subsections.

*Relationship of consistency index to other measures of variability.* As the calculation of the consistency index was modified to the AC dimension level for the present study, further analyses were conducted to understand how the dimension-level consistency indices related to other more common measures of variability, specifically the standard deviation and the range of scores in each AC dimension. Table 3a contains means, standard deviations, and intercorrelations for the AC dimension-level consistencies and the standard deviations for each AC dimension. Table 3b contains means, standard deviations, and intercorrelations for the AC dimension-level consistencies and the range for each AC dimension. As can be seen in Table 3a, the dimension-level AC consistency indices were perfectly negatively correlated with the standard deviation for each AC dimension. Additionally, the dimension-level AC consistency indices were strongly, negatively correlated with the range for each AC

dimension. Tables 3c, 3d, 3e, and 3f show the means, standard deviations, and intercorrelations for the standard deviation and range and peer and supervisor ratings of job performance. The standard deviation and range for the AC dimensions exhibited the same pattern of relationships with ratings of job performance as the consistency indices.

*Relationship between mean-level AC dimension scores and overall AC scores and ratings of job performance.* The means, standard deviations, and intercorrelations between the overall AC scores and mean-level AC dimension scores and supervisor ratings of job performance are presented in Table 4a. The means, standard deviations, and intercorrelations between the overall AC scores and mean-level AC dimension scores and peer ratings of job performance are presented in Table 4b. The correlations among mean dimension scores are stronger within time 1 and time 2 than across times, although the dimensions are significantly correlated over time. Additionally, the overall AC scores are significantly related over time ($r = .25$, $p = .008$). Although the majority of the mean scores from the AC dimensions at both time 1 and time 2 were not significantly related to supervisor or peer ratings of job performance, the overall AC scores at time 2 were significantly related to the aggregated supervisor ratings of job performance ($r = .21$, $p = .026$).

*Comparing peer and supervisor ratings of job performance.* Table 5 contains the intercorrelations for the supervisor and peer ratings of job performance both on the dimension and aggregate level. The aggregated supervisor and peer ratings of job performance were significantly correlated ($r = .47$, $p < .001$). The internal consistency was higher for the supervisor ratings of job performance than for the peer ratings,

although the internal consistency of the peer ratings was still at an acceptable level (α = .93, α = .85; Cortina, 1993).

To further investigate the reliability of the supervisor and peer ratings of job performance, $r_{wg}$ was calculated to assess the degree of interrater agreement for both supervisor and peer ratings of job performance. The $r_{wg}$ values are presented in Table 6 below. The resulting values for $r_{wg}$ below are all statistically significant and indicate acceptable levels of agreement for both supervisor and peer raters ($p$ <.05, Dunlap, Burke, & Smith-Crowe, 2003).

# DISCUSSION

The following section will summarize the key findings of the present study and provide a discussion of each hypothesis. The implications of the additional analyses provided in the Results section will also be discussed. Directions for future research and limitations of the present study are provided.

## *Implications*

*Relationship of AC performance consistency over time.* Hypothesis 1 stated that AC performance consistency at time 1 would be significantly related to AC performance consistency at time 2. Hypothesis 1 was found to be supported. Although hypothesis 1 was supported, additional analyses were performed to investigate the relationships between AC dimension-level consistency indices over time. As stated in the Results section, none of the AC dimension-level consistency indices were significantly correlated between time 1 and time 2. The finding that the aggregated dimension-level consistency indices were significantly related, yet the dimension-level indices were not related between time 1 and time 2 is somewhat conflicting. Some evidence has been found suggesting that a general measure of consistency exists, although the present study provides the first empirical investigation of AC consistency at the dimension level (Bray, Campbell, & Grant, 1974; Gibbons & Rupp, in press). Bray and colleagues found evidence that a general consistency factor included in the Management Progress Study was found to be predictive of future promotions. Additionally, Gibbons (2007) index of

consistency was developed to measure overall AC consistency. The index of AC consistency developed by Gibbons was modified for the purposes of the present study to measure AC consistency at the dimension level. For further detail on how Gibbons (2007) originally calculated consistency see Figure 1 in Appendix G. Perhaps an alternative measure of consistency is needed to accurately capture consistency at the narrower dimension level.

In order to further explore the nature of AC consistency at the dimension level, frequency distributions were obtained for each AC dimension consistency index at both time 1 and time 2. All distributions were normal, suggesting that AC consistency varies reliably for each AC dimension, although dimension-level consistency at time 1 was not found to be related to consistency on the same dimension at time 2.

Analysis of the intercorrelations of the AC dimension consistencies within each AC administration time revealed that the problem analysis and technical departmental knowledge consistency indices were significantly related within both AC administration times ($r$ = .18; $r$ = .27). A possible explanation for this finding is that technical departmental knowledge is necessary for effective problem analysis, therefore consistency (or inconsistency) on one dimension may lead to consistency on the other. The stable relationship between consistency on problem analysis and technical departmental knowledge is an interesting finding that is worthy of further investigation and replication.

The finding that overall AC consistency is correlated over time is relevant for several reasons. First, the moderate stability of AC performance consistency supports the conceptualization of consistency as an individual difference that is measurable, but has

been overlooked by AC practitioners. Second, the finding that performance consistency is significantly related over time may be most applicable to developmental ACs. Future developmental ACs may elect to provide feedback to candidates regarding their overall performance consistency in the AC. Further research could also investigate whether this feedback may result in improved levels of consistency in later ACs, which would further validate the recognition of consistency as an individual difference. Consistency feedback may be particularly important for occupations or jobs where inconsistency in performance is extremely costly (e.g., air traffic controllers, police officers, combat troops, etc.). Gibbons and Rupp (in press) further suggest that AC consistency may represent differing individual skill patterns which could be enhanced through developmental feedback. Although future research may provide evidence of more specific measures of consistency, the findings of the present study suggest that AC consistency can only be reliably measured at a general level. Thus, any feedback given regarding AC performance consistency should be kept at a general level until reliable measures of consistency at the dimension level are developed and patterns of consistency are more identifiable.

*Linking AC performance consistency to ratings of job performance.* Hypotheses 2a through 2e and 3a through 3e proposed that AC performance consistency would relate to supervisor and peer ratings of job performance, which included the job performance dimensions of customer service, quantity of work, quality of work, job knowledge, and teamwork/cooperation. Only hypothesis 3d was supported where overall AC performance consistency at time 2 was found to be significantly correlated with peer ratings of teamwork/cooperation. The majority of the hypotheses linking overall AC performance

consistency to ratings of job performance were not supported. This may be due to the quality of the job performance ratings and methodological constraints encountered when collecting the job performance ratings.

The lack of support for hypotheses 2a-2e and 3a-3e is cause for some concern as these hypotheses were proposed to further validate the meaningfulness of AC consistency as a possible predictor of job performance. An alternative conceptualization of the relationship between AC consistency and job performance may be that AC consistency may serve as a moderator between AC performance and ratings of job performance. Edwards and Woehr (2007) found a similar effect where personality consistency moderated the relationship between self and other ratings of personality. An earlier body of research by Bem and Allen (1974) and Bem and Funder (1978) focused on operationalizing personality consistency as a means of "predicting more of the people more of the time" where the relationship between personality and behavior was stronger for those individuals with more consistent behavior. In relation to the present study, it may be that AC consistency only provides additional utility for decisions made from AC performance for those individuals who are consistent throughout the AC. These areas may be possible directions for future research. Limited prior research exists linking AC consistency to job performance and additional research is needed before any conclusions can be drawn regarding the relationship of AC consistency to job performance (Gibbons & Rupp, 2007; Gibbons & Rupp, in press).

Hypothesis 4 proposed that overall AC consistency at time 1 and time 2 would relate to supervisor ratings of job performance consistency; however, hypothesis 4 was not supported. As stated previously, the methodological issues encountered during

collection of the ratings of job performance could have interfered with measuring the actual relationship between AC performance consistency and job performance consistency. Gibbons and Rupp (in press) recently issued a call for additional research linking AC consistency to measures of job performance consistency as a means of further validating AC consistency. Although the findings of the present study did not support a relationship between AC consistency and job performance consistency, additional research may find support for this link.

*Exploratory analyses.* Additional analyses were performed to further explore why many of the proposed hypotheses were not supported. First, correlations between the AC dimension consistency indices and the standard deviations for the AC dimensions were obtained. The resulting correlation coefficients revealed that Gibbons' (2007) consistency index, when modified to the dimension-level of ACs, is perfectly negatively correlated with the standard deviation for each AC dimension. In addition to comparing the consistency index to the standard deviations, the AC dimension consistency indices were correlated with the range for each AC dimension. As the standard deviation and the range are related, it was not surprising that the AC dimension consistency indices were strongly, negatively correlated with the range for each dimension, as well. Thus, rather than going through the somewhat complex process of computing AC dimension consistency indices using the modified Gibbons consistency index, obtaining the standard deviation for each AC dimension may be preferable based on the findings of the present study. See Figure 4 for an example of how the AC dimension consistency indices and the AC dimension standard deviations may be practically implemented into operational ACs. Figure 4 serves only as a hypothetical example of how the consistency index and the

standard deviation for the AC dimensions *may* be operationalized. Additional research is needed to determine what methods of integrating consistency measurement into AC scores are most effective and if AC performance consistency is predictive of job performance, which was not found in the present study.

Although overall AC consistency was found to be stable across AC administrations, that stability did not hold up at the AC dimension level. In order to further understand these findings, the correlations between the mean scores for each AC dimension at time 1 and time 2 were obtained. Although the correlations among AC dimensions within each time were higher than the correlations of the mean scores on each dimension over time, the AC dimension mean scores were still significantly correlated across time 1 and time 2. Also, the AC dimension scores were more strongly correlated between time 1 and time 2 than the overall AC scores.

Overall AC scores were found to be significantly related between time 1 and time 2; however, the finding that overall AC scores at time 1 and time 2 have a relatively low correlation ($r$ = .25) brings into question the reliability of the AC method itself. Although the sample obtained for the present study includes individuals who averaged a two to three year gap between the first and second time that they participated in the AC, it is somewhat surprising that a higher correlation was not found between the overall AC score at time 1 and time 2. Brannick (2008) recently issued a call for research to assess the psychometric properties of assessment centers. In particular, Brannick emphasizes the poor reliability of AC scores both by dimension and exercise. The poor reliability of ACs is relevant for research investigating consistency in ACs because score variability may in fact be due to measurement error and not represent an additional dimension of AC

performance as the present study proposed. If in fact ACs are not reliable measures of the performance dimensions that they claim to measure, then inferences made regarding AC performance consistency are likely not meaningful. Avenues for future research will be discussed in a later section concerning AC performance consistency and the AC method itself. The following section will present several factors that may limit the generalizability of the results of the present study.

*Limitations*

*Ratings of job performance.* As with all studies, the present study had several limitations. The ratings of job performance may not have been an appropriate criterion to link to the AC dimension consistencies. First, the collection of peer and supervisor ratings of job performance was constrained to occur only during the administration of the most recent AC. Thus, only a small and potentially non-representative subgroup of peers and supervisors were allowed to provide ratings for the AC participants. Also, the peers and supervisors were allowed to choose who to rate and the ratings were anonymous. Thus, there was no way to verify that the raters had actually worked with the individuals whom they rated or hold raters accountable for their ratings. Additionally, no formal training was provided to decrease the effects of rating errors. Although the training for the AC assessors is a rigorous and standardized process, the collection of job performance ratings for the purposes of the present study was not held to the same level of rigor as the training process for the AC assessors.

Beyond the lack of rater training, the scale used to rate AC participants' job performance consisted of five single-item measures for each job performance dimension. The dimensions included on the job performance measure were the same dimensions

used in the current performance evaluation tool used by the police department. The decision was made to use these dimensions because the raters would be familiar with these dimensions which may somewhat compensate for the lack of rater training. Although the raters may have been more familiar with the job performance dimensions measured, there was less of a theoretical foundation linking these dimensions to the AC dimension consistencies. Future research should select more comprehensive measures of job performance to provide the most accurate job performance ratings possible. Additionally, future research should seek to match the AC dimension consistencies with job performance dimensions or perhaps simply use a general measure of job performance.

Another factor affecting the accuracy of the ratings of job performance was the lack of advance notice given to the peer and supervisor raters. The raters were informed of the opportunity to rate their peers and subordinates the same day that the ratings were obtained. Thus, the raters based their ratings entirely off of recall. Although the collection of the job performance ratings was constrained by many methodological flaws, the present study sought to provide an initial exploratory investigation of how AC consistency may be related (or unrelated) to multisource ratings of job performance.

*Sample limitations.* In addition to the methodological flaws involved in the collection of job performance ratings, the participant sample obtained for the present study included only those individuals who participated in an AC at least twice. Thus, the range of participant scores was limited as those individuals who scored highly after their first AC were promoted. In addition to those individuals who were promoted after a single AC, some individuals who performed very poorly may have left the department

due to decreased likelihood of promotion. These limitations are inherent to data from operational ACs, however, and cannot be controlled due to the expense associated with ACs.

Another limitation of the present study may have been the pairing of scores over four different administrations. This method was chosen for the present study to increase the size of the sample available to study the temporal stability of AC consistency. Although the AC process is highly standardized at the consulting firm that conducted the ACs, some differences exist between administrations that cannot be controlled. To investigate this effect, a control variable was created to code for the administrations in which the individual participated. This variable was not found to attenuate any significant relationships that were found.

A final limitation of the present study was the inclusion of only scores from the structured interview component of the AC. Prior research has investigated the psychometric properties of the individual components of ACs, specifically the reliability of alternate forms of in-basket exercises (Brannick, 2008 and Brannick, Michaels, & Baker, 1989). Although AC consistency was measured using only the participant scores from structured interviews questions, the results of the present study are intended to provide preliminary guidance for future research concerning AC consistency based on scores from full ACs. More research is needed to replicate the findings of the present study using full ACs before stronger conclusions can be drawn regarding the stability of AC consistency and how it relates to job performance.

*Directions for Future Research*

Several suggestions for future research have already been made in the previous sections. To summarize, more research is needed to investigate the reliability of ACs. Although the present study provided some initial insight into the test-retest reliability of ACs, more research is needed to replicate the degree of reliability found in the present study. Additional research is also needed to extend the findings of the present study regarding the relationship between AC performance consistency and job performance.

*Measuring consistency.* Perhaps the most significant finding of the present study is the significant correlation of overall AC consistency over time. The initial evidence provided by the present study that overall AC consistency is relatively stable over time may lead to the development of new and better ways to measure performance consistency, both in contexts such as ACs and in the domain of job performance. Ultimately, the goal of measuring consistency in a promotional AC should be to identify a new performance dimension that will account for unique variance in job performance that is not already explained by the dimensions currently included in ACs. Also, if consistency is to be measured for developmental ACs, then research should show that consistency is something that individuals can change. If performance consistency is not something that can be developed, then participants may be discouraged or frustrated with feedback received from developmental ACs.

Several suggestions have already been made regarding alternative methods for assessing consistency in ACs and on the job. One method proposed by Gibbons and Rupp (in press) is to have the assessors rate a dimension in ACs that pertains to participants' performance consistency. Additionally, Deadrick and Gardner (1997) found that

performance distributions can be used to obtain reliable estimates of performance variability, however, their study involved objective measures of job performance as opposed to the more subjective performance dimensions used in ACs. More recent research by Edwards and Woehr (2007) supports the use of frequency-based estimation in the measurement of personality factors. Given the inconsistent findings of the present study, perhaps frequency-based ratings would be a viable alternative rating format for performance consistency within ACs. Including a frequency-based rating format in ACs may be more difficult to implement as the same assessors would likely need to observe candidates across all exercises to result in a frequency-based rating of dimension consistency (Gibbons & Rupp, in press). Also, perhaps a frequency-based rating format would be more likely to relate to measures of job performance as opposed to a measure of consistency based on average score differences.

*Predictors of consistency.* Once reliable ACs and metrics for consistency are developed, additional research is needed to understand why some individuals are more consistent than others in ACs and on the job. One possible predictor of AC performance consistency is performance anxiety. Multiple physiological measures exist that are relatively non-invasive such as skin conductance measures and saliva swabs which measure cortisol levels. In addition to physiological measures of anxiety, other lines of research have begun to investigate the moderator roles of neuroticism and/or communication apprehension in AC performance (Blume, 2006; Collins et al., 2003). Beyond the effects of personality variables and measures of anxiety, more qualitative methods might also be informative. Specifically, allowing AC participants to provide

qualitative feedback and suggest possible explanations for their own performance consistency in the AC may bring to light variables that have not yet been considered.

*Conclusion*

The results of the present study offer mixed support for research regarding AC performance consistency. While overall AC consistency was found to be stable over time, consistency was not found to relate to ratings of job performance or job performance consistency. Future research should seek to avoid the methodological flaws that may have affected the results of the present study. Additional suggestions were provided to guide future research and encourage productive analysis of performance consistency. Research regarding the validity of ACs has accumulated for several decades and the sustained popularity and growth of ACs ensures that the demand for additional research on ACs will continue.

REFERENCES

American Educational Research Association, American Psychological Association, and National Council for Measurement in Education (1999). *Standards for educational and psychological testing.* Washington, DC: American Educational Research Association.

Arthur, W., 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,* 125-154.

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,* 813-835.

Bem, D.J., & Allen, A. (1974). On predicting some of the people some of the time: The search for cross-situational consistencies in behavior. *Psychological Review, 81,* 506-520.

Bem, D.J., & Funder, D.C. (1978). Predicting more of the people more of the time: Assessing the personality of situations. *Psychological Review, 85,* 485-501.

Binning, J.F., & Barrett, G.V. (1989). Validity of personnel decisions: A conceptual analysis of the inferential and evidential bases. *Journal of Applied Psychology, 74,* 478-494.

Blume, B.D. (2006). Construct confusion and assessment centers: A person-situation interactionist perspective (Doctoral dissertation, Indiana University, 2006). *Dissertation Abstracts International, 67,* 2225.

Bowler, M.C., & Woehr, D.J. (2006). A meta-analytic evaluation of the impact of dimension and exercise factors on assessment center ratings. *Journal of Applied Psychology, 91,* 1114-1124.

Brannick, M.T. (2008). Back to basics of test construction and scoring. *Industrial and Organizational Psychology: Perspectives on Science and Practice, 1,* 131-133.

Brannick, M.T., Michaels, C.E., & Baker, D.P. (1989). Construct validity of in-basket scores. *Journal of Applied Psychology, 74,* 957-963.

Bray, D.W., Campbell, R.J., & Grant, D.L. (1974). *Formative years in business: A long-term A.T.&T. study of managerial lives.* New York: Wiley.

Bray, D.W., & Grant, D.L. (1966). The assessment center in the measurement of potential for business management. *Psychological Monographs: General & Applied, 80,* 1-27.

Bush, M.A. (2003). Assessment center construct validity: Establishing expectations based on the dimension activation theory (Doctoral dissertation, University of Tennessee, Knoxville). *Dissertation Abstracts International: Section B: The Sciences & Engineering, 51* (1-B), 469.

Bycio, P., Alvares, K.M., & Hahn, J. (1987). Situation specificity in assessment center ratings: A confirmatory factor analysis. *Journal of Applied Psychology, 72,* 463-474.

Campbell, D.T., & Fiske, D.W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin, 56,* 81-105.

Cascio, W.F. (1998). *Applied psychology in human resource management* (5th ed.). NJ: Prentice Hall.

Collins, J., Schmidt, F., Sanchez-Ku, M., Thomas, L., McDaniel, M., & Le. H. (2003). Can basic individual differences shed light on the construct meaning of assessment center evaluations? *International Journal of Selection and Assessment, 11,* 17-29.

Conway, J.M., & Huffcutt, A.I. (1997). Psychometric properties of multisource performance ratings: A meta-analysis of subordinate, superior, peer, and self-ratings. *Human Performance, 10,* 331-360.

Cortina, J.M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology, 78,* 98-104.

Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory.* Orlando, FL: Holt, Rinehart, and Winston.

Cronbach, L.J. (1990). *Essential of psychological testing* (5th ed.). New York: Harper & Row.

Deadrick, D.L., & Gardner, D.G. (1997). Distributional ratings of performance levels and variability. *Group & Organization Management, 22,* 317-342.

Dreher, G.F., & Sackett, P.R. (1981). Some problems with applying content validity evidence to assessment center procedures. *Academy of Management Review, 6,* 551-560.

Dunlap, W.P., Burke, M.J., & Smith-Crowe, K. (2003). Accurate tests of statistical significance for *rwg* and average deviation interrater agreement indexes. *Journal of Applied Psychology, 88,* 356-362.

Edwards, B. D., Woehr, D.J. (2007). An examination and evaluation of frequency-based personality measurement. *Personality and Individual Differences, 43,* 803-814.

Epstein, S., & O'Brien, E.J. (1985). The person-situation debate in historical and current perspective. *Psychological Bulletin, 98,* 513-537.

Fiske, D.W. (1961). The inherent variability of behavior. In D.W. Fiske & S.R. Maddi (Eds.), *Functions of varied experience* (pp.326-354). Homewood, IL: Dorsey Press.

Fiske, D.W., Hanfmann, E., MacKinnon, D.W., Miller, J.G., & Murray, H.A. (1948). Selection of personnel for clandestine operations: Assessment of men. Walnut Creek, CA: Aegean Park Press.

Fleeson, W. (2001). Toward a structure- and process-integrated view of personality: Traits as density distributions of states. *Journal of Personality & Social Psychology, 80,* 1011-1027.

Gatewood, R.D., & Feild, H.S. (2001). *Human resource selection* (5th ed.). OH: Harcourt.

Gaugler, B.B., Rosenthal, D.B., Thornton, G.C., & Bentson, C. (1987). Meta-analysis of assessment center validity. *Journal of Applied Psychology, 72,* 493-511.

Gaugler, B.B., & Rudolph, A.S. (1992). The influence of assessee performance variation on assessors' judgments. *Personnel Psychology, 45,* 77-98.

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

Gibbons, A.A.M. (2007). Inconsistency in assessment center performance: Measurement error or something more? (Doctoral dissertation, University of Illinois, Urbana - Champaign). *Dissertation Abstracts International, 68*, 4874.

Gibbons, A.M., & Rupp, D.E. (2007, April). *Inconsistency in assessment center performance: A meaningful individual difference?* Paper presented at the 22nd meeting of the Society for Industrial and Organizational Psychology, New York, NY.

Gibbons, A.M., & Rupp, D.E. (in press). Dimension consistency as an individual difference: A new (old) perspective on the assessment center construct validity debate. *Journal of Management.*

Goldstein, I.L., Zedeck, S., & Schneider, B. (1993). An exploration of the job-analysis-content validity process. In N. Schmitt and W. Borman (Eds.), *Personnel Selection in Organizations.* San Francisco: Jossey-Bass.

Haaland, S., & Christiansen, N.D. (2002). Implications of trait-activation theory for evaluating the construct validity of assessment center ratings. *Personnel Psychology, 55,* 137-163.

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,* 140-155.

Hinrichs, J.R. (1978). An eight-year follow-up of a management assessment center. *Journal of Applied Psychology, 63,* 596-601.

Howard, A. (1997). A reassessment of assessment centers: Challenges for the 21st century. *Journal of Social Behavior and Personality, 12,* 13-52.

Howard, A. (2008). Making assessment centers work the way they are supposed to. *Industrial and Organizational Psychology: Perspectives on Science and Practice, 1,* 98-104.

Hunter, J.E., & Hunter, R.F. (1984). Validity and utility of alternative predictors of job performance. *Psychological Bulletin, 96,* 72-98.

International Task Force on Assessment Center Guidelines. (2000). Guidelines and ethical considerations for assessment center operations. *Public Personnel Management, 28,* 315-331.

Jansen, P.G.W., & Stoop, B.A.M. (2001). The dynamics of assessment center validity: Results of a 7-year study. *Journal of Applied Psychology, 86,* 741-753.

Jansen, P.G.W., & Vinkenburg, C.J. (2006). Predicting management career success from assessment center data: A longitudinal study. *Journal of Vocational Behavior, 68,* 253-266.

Jones, R.G., & Whitmore, M.D. (1995). Evaluating developmental assessment centers as interventions. *Personnel Psychology, 48,* 377-388.

Joyce, L.W., Thayer, P.W., & Pond, S.B. (1994). Managerial functions: An alternative to traditional assessment center dimensions? *Personnel Psychology, 47,* 109-121.

Kendrick, D.T., & Funder, D.C. (1988). Profiting from controversy: Lessons from the person-situation debate. *American Psychologist, 43,* 23-34.

Klimoski, R.J., & Brickner, M. (1987). Why do assessment centers work? The puzzle of assessment center validity. *Personnel Psychology, 40,* 243-260.

Lance, C.E. (2008a). Why assessment centers do not work the way they are supposed to. *Industrial and Organizational Psychology: Perspectives on Science and Practice, 1,* 84-97.

Lance, C.E. (2008b). Where have we been, how did we get there, and where shall we go? *Industrial and Organizational Psychology: Perspectives on Science and Practice, 1,* 140-146.

Lance, C.E., Foster, M.R., Gentry, W.A., & Thoreson, J.D. (2004a). Assessor cognitive processes in an operational assessment center. *Journal of Applied Psychology, 89,*22-35.

Lance, C.E., Foster, M.R., Nemeth, Y.M., Gentry, W.A., & Drollinger, S. (2007). Extending the nomological network of assessment center construct validity: Prediction of cross-situationally consistent and specific aspects of assessment center performance. *Human Performance, 20,* 345-362.

Lance, C.E., Newbolt, W.H., Gatewood, R.D., Foster, M.R., French, N., & Smith, D.E. (2000). Assessment center exercise factors represent cross-situational specificity, not method bias. *Human Performance, 13,* 323-353.

Lance, C.E., Lambert, T.A., Gewin, A.G., Lievens, F., & Conway, J.M. (2004). Revised estimates of dimension and exercise variance components in assessment center post-exercise dimension ratings. *Journal of Applied Psychology, 89,* 377-385.

Landy, F.J. (1986). Stamp collecting versus science: Validation as hypothesis testing. *American Psychologist, 41,* 1183-1192.

Lawshe, C.H. (1975). A quantitative approach to content validity. *Personnel Psychology, 28,* 563-575.

Lievens, F. (1998). Factors which improve the construct validity of assessment centers: A review. *International Journal of Selection and Assessment, 6,* 141-152.

Lievens, F. (2001). Assessors and use of assessment centre dimensions: A fresh look at a troubling issue. *Journal of Organizational Behavior, 22,* 203-221.

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,* 247-258.

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,* 1202-1222.

Lievens, F., & Klimoski, R.J. (2001). Understanding the assessment center process: Where are we now? *International Review of Industrial and Organizational Psychology, 16,* 246-286.

McEvoy, G.M., & Beatty, R.W. (1989). Assessment centers and subordinate appraisals of managers: A seven-year examination of predictive validity. *Personnel Psychology, 42,* 37-52.

Meriac, J.P., Hoffman, B.J., Woehr, D.J., & Fleisher, M.S. (2008). Further evidence for the validity of assessment center dimensions: A meta-analysis of the incremental criterion-related validity of dimension ratings. *Journal of Applied Psychology, 93,* 1042-1052.

Mischel, W. (1968). *Personality and assessment.* New York: John Wiley and Sons.

Mischel, W. (1973). Toward a cognitive social learning reconceptualization of personality. *Psychological Review, 80,* 252-283.

Mischel, W. (2004). Toward an integrative science of the person. *Annual Review of Psychology, 55,* 1-22.

Mount, M.K., Judge, T.A., Scullen, S.E., Sytsma, M.R., & Hezlett, S.A. (1998). Trait, rater, and level effects in 360-degree performance ratings. *Personnel Psychology, 51*, 557–576.

Murphy, K.R. & Cleveland, J.N. (1991) *Performance appraisal: An organizational perspective.* Boston: Allyn and Bacon.

Neidig, R.D., & Neidig, P.J. (1984). Multiple assessment center exercises and job relatedness. *Journal of Applied Psychology, 69,* 182-186.

Office of Strategic Services Assessment Staff. *Assessment of Men: Selection of personnel for the Office of Strategic Services.* New York: Rinehart, 1948.

Rupp, D.E., Thornton, G.C., III, & Gibbons, A.M. (2008). The construct validity of the assessment center method and usefulness of dimensions as focal constructs. *Industrial and Organizational Psychology: Perspectives on Science and Practice, 1,* 116-120.

Russell, C. J. (1985). Individual decision processes in an assessment center. *Journal of Applied Psychology, 70,* 737-746.

Russell, C.J., & Domm, D.R. (1995). Two field tests of an explanation of assessment centre validity. *Journal of Occupational and Organizational Psychology, 68,* 25-47.

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

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

Sackett, P.R., & Dreher, G.F. (1984). Situation specificity of behavior and assessment center validation strategies: A rejoinder to Neidig and Neidig. *Journal of Applied Psychology, 69,* 187-190.

Sackett, P.R., Schmitt, N., Ellingson, J.E., & Kabin, M.B. (2001). High-stakes testing in employment, credentialing, and higher education: Prospects in a post-affirmative-action world. *American Psychologist, 56,* 302-318.

Sackett, P.R., & Tuzinski, K. (2001). The role of dimensions and exercises in assessment center judgments. In M. London (Ed.), *How people evaluate others in organizations* (pp.111-129). Malwah, NJ: Earlbaum.

Sagie, A., & Magnezy, R. (1997). Assessor type, number of distinguishable categories, and assessment centre construct validity. *Journal of Occupational and Organizational Psychology, 70,* 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,* 735-746.

Schmidt, F.L., & Hunter, J.E. (1998). The validity and utility of selection methods in personnel psychology: practical and theoretical implications of 85 years of research findings. *Psychological Bulletin, 124,* 262-274.

Schneider, J.R., & Schmitt, N. (1992). An exercise design approach to understanding assessment center dimension and exercise constructs. *Journal of Applied Psychology, 77,* 32-41.

Shoda, Y. (1999). A unified framework for the study of behavioral consistency: Bridging person x situation interaction and the consistency paradox. *European Journal of Personality, 13,* 361-387.

Shoda, Y. (2003). Studying persons in order to understand situations: studying situations in order to understand persons. In C. Sansone, C. Morf, & A. Panter (Eds.), *Handbook of methods in social psychology* (pp.128-153). Thousand Oaks, CA: Sage.

Shoda, Y., Mischel, W., & Wright, J.C. (1993). The role of situational demands and cognitive competencies in behavior organization and personality coherence. *Journal of Personality and Social Psychology, 65,* 1023-1035.

Shoda, Y., Mischel, W., & Wright, J.C. (1994). Intraindividual stability in the organization and patterning of behavior: Incorporating psychological situations into the idiographic analysis of personality. *Journal of Personality and Social Psychology, 67,* 674-687.

Society for Industrial and Organizational Psychology, Inc. (2003). *Principles for the validation and use of personnel selection procedures*. (4th ed.). Bowling Green, OH: Author.

Spychalski, A.C., Quinones, M.A., Gaugler, B.B., & Pohley, K. (1997). A survey of assessment center practices in organizations in the united states. *Personnel Psychology, 50,* 71-90.

Tenopyr, M.L. (1977). Content-construct confusion. *Personnel Psychology, 30,* 47-54.

Tett, R.P., & Burnett, D.D. (2003). A personality trait-based interactionist model of job performance. *Journal of Applied Psychology, 88,* 500-517.

Thornton, G.C., III (1992). *Assessment centers in human resource management.* Reading, MA: Addison-Wesley.

Thornton, G.C., III, & Byham, W.C. (1982). *Assessment centers and managerial performance.* New York: Academic Press.

Thornton, G.C., III, & Mueller-Hanson, R.A. (2004). *Developing organizational simulations.* Mahwah, NJ: Lawrence-Earlbaum.

Thornton, G.C., III, & Rupp, D.E. (2006). *Assessment centers in human resource management.* Mahwah, NJ: Lawrence Erlbaum.

Turnage, J.J., & Muchinsky, P.M. (1982). Transsituational variability in human performance within assessment centers. *Organizational Behavior & Human Performance, 30,* 174-200.

Tziner, A., Ronen, S., & Hacohen, D. (1993). A four-year validation study of an assessment center in a financial corporation. *Journal of Organizational Behavior, 14,* 225-237.

Wernimont, P.F., & Campbell, J.P. (1968). Signs, samples, and criteria. *Journal of Applied Psychology, 52,* 372-376.

Woehr, D.J., & Arthur, W. (2003). The construct-related validity of assessment center ratings: A review and meta-analysis of the role of methodological factors. *Journal of Management, 29,* 231-258.

Woehr, D.J., & Huffcut, A.I. (1994). Rater training for performance appraisal: A quantitative review. *Journal of Occupational and Organizational Psychology, 67,* 189-205.

# APPENDIX A

**Assessment Center Criteria**

1. A thorough job analysis must be conducted in order to determine what behaviors will be evaluated in the assessment center.
2. Behaviors observed during the assessment center must be classified into meaningful and relevant categories (i.e., dimensions, skills, abilities, competencies, etc.).
3. Assessment techniques must elicit behavior on dimensions/categories identified through job analysis.
4. Multiple assessment techniques must be used.
5. At least one job-related simulation must be included in order for the candidate to exhibit job-related behavior.
6. Multiple assessors must be used.
7. Assessors must receive thorough training and exhibit adequate assessment proficiency before evaluating candidates' behavior.
8. Assessors must systematically record candidates' behavior.
9. Assessors must independently report observations of candidate behavior before integrating information with other assessors.
10. The combination of candidate scores or behavioral information must be combined either statistically or through assessor's consensus discussion.

# APPENDIX B

**Name of Officer: _____**

**This officer is a (circle one): Peer Subordinate**

Please rate your peer/subordinate officer's performance while on the job in the following areas using the scale below:

| 5 | 4 | 3 | 2 | 1 |
|---|---|---|---|---|
| Far Exceeds Standards | Exceeds Standards | Consistently Meets Standards | Marginal Standards | Below Standards |

| Performance Area | Rating |
|---|---|
| Customer Service | |
| Quantity of Work | |
| Quality of Work | |
| Job Knowledge | |
| Teamwork/Cooperation | |
| Consistency of Work | |

# APPENDIX C

# Police Department
# Sergeant Promotional Assessment Center
# Operational Performance Dimension Definitions

**Problem Analysis**
Effectiveness in identifying problem areas, securing relevant information, relating and comparing information from different sources, determining the source of a problem, and implementing task-resolving decisions. This includes developing short- or long-range plans to determine objectives, identify problems, establish priorities, set standards, provide guidelines, and identify resource needs.

**Supervisory Ability**
The extent to which subordinates are provided with directions and guidance toward the accomplishment of specified performance goals. This includes the ability to set and enforce performance standards, recognize problem behavior, evaluate subordinate work performance, provide guidelines, and monitor subordinate performance in order to provide assistance, extend recognition, discipline, and motivate or counsel. Supervisory Ability differs from Management Ability in that Supervisory Ability is concerned with the work performance and professional development of individuals in one's area of responsibility; whereas Management Ability focuses on allocating personnel and equipment to meet Division or Unit work responsibilities or assignments.

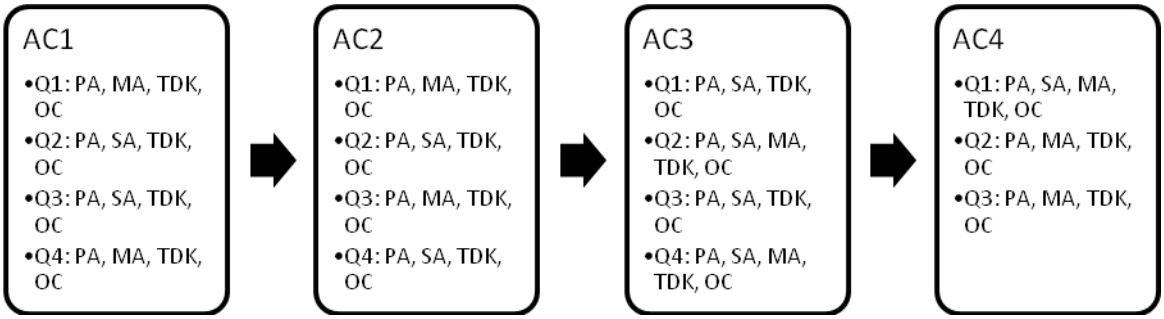
**Management Ability**
The extent to which work is effectively planned, organized, and coordinated for the efficient accomplishment of specified goals. This includes proper assignment of personnel, appropriate allocation and management of resources, recognition of resource limitations, and enforcement of policies. Management Ability differs from Supervisory Ability in that Management Ability is concerned with allocating personnel and equipment to meet Division or Unit work responsibilities or assignments; whereas Supervisory Ability focuses on the work performance and professional development of individuals in one's area of responsibility.

**Technical & Departmental Knowledge**
Demonstrates knowledge and understanding of departmental policies, procedures, and rules and regulations in planning work, monitoring employee performance, disciplining employees, making decisions, giving advice, and responding to situations. This includes utilizing knowledge of the departmental organization to find solutions to problems.

**Oral Communication**
The clear, unambiguous, and effective expression of oneself through oral means to individuals such as co-workers, other agency employees, the general public, and community groups to ensure the accurate and/or persuasive exchange of information. This includes receiving and comprehending information from another individual in order to respond appropriately.

**Human Relations**
The use of appropriate interpersonal skills which indicate a consideration of the feelings, interests and needs of employees, representatives of other agencies and the general public. This includes using tact, building and maintaining rapport and morale, recognizing stress symptoms in others when interacting in one-on-one situations or with groups to resolve interpersonal conflicts, and address complaints.

Performance Dimensions Measured in Structured Interview Questions in each AC Administration




PA = Problem Analysis
SA = Supervisory Ability
MA = Management Ability
TDK = Technical and Departmental Knowledge
OC = Oral Communication

# APPENDIX D

| | Structured Interview Question | | | |
|---|---|---|---|---|
| **AC Performance Dimension** | **1** | **2** | **3** | **4** |
| Problem Analysis (PA) | 6 | 4 | 5 | 4 |
| Management Ability (MA) | 6 | not measured | 4 | not measured |
| Technical & Departmental Knowledge (TDK) | 5 | 3 | 4 | 6 |
| Oral Communication | 4 | 5 | 6 | 4 |

Rating scale: 1-7, Range = 6

Step 1. A consistency index is calculated for each AC performance dimension by calculating difference scores for each pair of scores on a dimension measured in each exercise. This value is then divided by the number of pairs and then the square root is obtained:

$C_{PA} = \sqrt{(6\text{-}4)^2 + (6\text{-}5)^2 + (6\text{-}4)^2 + (4\text{-}5)^2 + (4\text{-}4)^2 + (5\text{-}4)^2/6} = .55$

* Although the present study only calculated consistency based on scores from the structured interview component of an AC, this table could be expanded to include all exercises in the AC.

Step 2.

Next, subtract from 1 and divide by the range of the scale used to score each exercise.

$C_{PA} = 1\text{-} (.55/6) = .91$

The closer the index is to 1, the more consistent the candidate was on that particular dimension.

Consistency indices are calculated for each performance dimension in the AC.

An overall consistency value for the AC is obtained by adding together the consistency index for each performance dimension:

$$C_{PA} + C_{MA} + C_{TDK} + C_{OC} = C_{Total}$$

Figure 1a. Example of how C-index is calculated.

Abstract Example:

Step 1: Construct an individual dimension x exercise matrix for each candidate:

| | Exercises | | | | |
|---|---|---|---|---|---|
| Dimensions: | 1 | 2 | 3 | … | K |
| 1 | $X_{11}$ | $X_{12}$ | $X_{13}$ | … | $X_{1K}$ |
| 2 | $X_{21}$ | $X_{22}$ | $X_{23}$ | … | $X_{2K}$ |
| 3 | $X_{31}$ | $X_{32}$ | $X_{33}$ | … | $X_{3K}$ |
| … | … | … | … | … | … |
| J | $X_{J1}$ | $X_{J2}$ | $X_{J3}$ | … | $X_{JK}$ |

Step 2: Pairwise correspondence indices $c_{ik}$ between pairs of **exercises (columns)** indicate **consistency**:

$$c_{12} = \sqrt{\frac{(X_{11} - X_{12})^2 + (X_{21} - X_{22})^2 + (X_{31} - X_{32})^2 + \ldots + (X_{J1} - X_{J2})^2}{J}}$$

$$c_{13} = \sqrt{\frac{(X_{11} - X_{13})^2 + (X_{21} - X_{23})^2 + (X_{31} - X_{33})^2 + \ldots + (X_{J1} - X_{J3})^2}{J}}$$

Step 3: Repeat for all possible pairs of exercises: ${}_K C_2$ or $\left(\frac{J(J-1)}{2}\right)$

Step 4: The **overall consistency index** $C = 1 - \frac{\overline{c_{ik}}}{R}$

Where $R$ = the range of the rating scale.

Figure 1b. Reproduced from Gibbons (2007). Example of how Gibbons originally proposed to calculate consistency.

# APPENDIX E

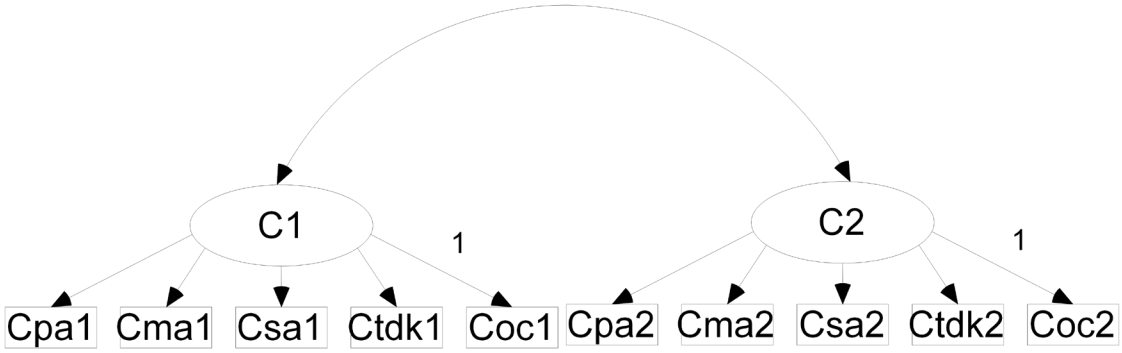



*Figure 2.* Proposed model to test Hypothesis 1-temporal stability of AC performance consistency.

C – Total consistency in AC*
Cpa – Problem analysis consistency in AC*
Cma – Management ability consistency in AC*
Csa – Supervisory ability in AC*
Ctdk – Technical/Departmental knowledge consistency in AC*
Coc – Oral communication consistency in AC*

*All of the variables listed above were measured in AC1 and AC2
** Supervisory ability was not included in subsequent analyses because it was only measured once in the most recent AC administration and consistency indices could not be calculated

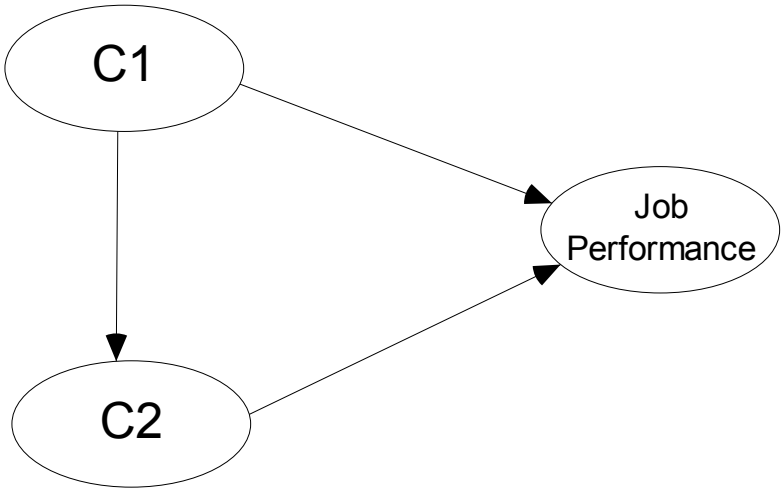



*Figure 3.* Model to test Hypotheses 2a through 2e and 3a through 3e – testing the relationship between AC performance consistency and ratings of job performance.

C1 – Total consistency in AC1
C2 – Total consistency in AC2

APPENDIX F

Table 1

*Means, Standard Deviations, and Intercorrelations for AC Dimension Consistency Indices at Time 1 and Time 2*

| | M | SD | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1. C-PA 1 | .77 | .10 | - | | | | | | | | | |
| 2. C-PA 2 | .79 | .10 | -.11 | - | | | | | | | | |
| 3. C- MA 1 | .84 | .11 | -.06 | .00 | - | | | | | | | |
| 4. C- MA 2 | .81 | .17 | .09 | .10 | .14 | - | | | | | | |
| 5. C-TDK 1 | .77 | .09 | .18* | .10 | .07 | .12 | - | | | | | |
| 6. C- TDK 2 | .78 | .11 | -.08 | .27** | .06 | .08 | -.03 | - | | | | |
| 7. C- OC 1 | .79 | .09 | .01 | -.05 | .21* | .02 | .04 | -.02 | - | | | |
| 8. C- OC 2 | .78 | .10 | .09 | .07 | .11 | .12 | .10 | .25** | -.01 | - | | |
| 9. Overall C 1 | 3.93 | .72 | -.01 | .11 | -.06 | -.05 | .05 | .02 | -.07 | .07 | - | |
| 10. Overall C 2 | 4.25 | .76 | .07 | .04 | .06 | .05 | .04 | -.09 | .05 | .07 | .44** | - |



**Correlation is significant at the .01 level (2-tailed)

*Correlation is significant at the .05 level (2-tailed)

Key: C-PA 1 is consistency of problem analysis at time 1
C-PA 2 is consistency of problem analysis at time 2
C-MA 1 is consistency of management ability at time 1
C-MA 2 is consistency of management ability at time 2
C-TDK 1 is consistency of technical departmental knowledge at time 1
C-TDK 2 is consistency of technical departmental knowledge at time 2
C-OC 1 is consistency of oral communication at time 1
C-OC 2 is consistency of oral communication at time 2

Table 2a

*Means, Standard Deviations and Intercorrelations for Consistency Indices of AC Dimensions and Supervisor Ratings of Job Performance*

| | M | SD | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1. C-PA 1 | .77 | .10 | - | | | | | | | | | | | | | | | |
| 2. C-PA 2 | .79 | .10 | -.11 | - | | | | | | | | | | | | | | |
| 3. C-MA 1 | .84 | .11 | -.06 | .00 | - | | | | | | | | | | | | | |
| 4. C-MA 2 | .81 | .17 | .09 | .10 | .14 | - | | | | | | | | | | | | |
| 5. C-TDK 1 | .77 | .09 | .18* | .10 | .07 | .12 | - | | | | | | | | | | | |
| 6. C-TDK 2 | .78 | .11 | -.08 | .27** | .06 | .08 | -.03 | - | | | | | | | | | | |
| 7. C-OC 1 | .79 | .09 | .01 | -.05 | .21* | .02 | .04 | -.02 | - | | | | | | | | | |
| 8. C-OC 2 | .78 | .10 | .09 | .07 | .11 | .12 | .10 | .25** | -.01 | - | | | | | | | | |
| 9. Total C 1 | 3.93 | .72 | -.01 | .11 | -.06 | -05 | .05 | .02 | -.07 | .07 | - | | | | | | | |
| 10. Total C 2 | 4.25 | .76 | .07 | .04 | .06 | .05 | .04 | -.09 | .05 | .07 | .44** | - | | | | | | |
| 11. Customer Service | 3.74 | .74 | .21* | -.23* | -.07 | -.12 | .00 | -.20* | .07 | -.05 | .02 | .11 | - | | | | | |
| 12.Work Quantity | 3.75 | .90 | .19* | -.17 | .00 | -.12 | -.08 | -.12 | .15 | .15 | .08 | .07 | .61** | - | | | | |
| 13. Work Quality | 3.88 | .86 | .23* | -.23* | -.02 | -.01 | -.02 | -.16 | .12 | .13 | .06 | .08 | .63** | .87** | - | | | |
| 14. Job Knowledge | 3.99 | .74 | .21* | -.26* | -.01 | -.07 | -.06 | -.20* | .01 | .08 | .01 | .05 | .65** | .73** | .83** | - | | |
| 15. Teamwork | 3.99 | .85 | .14 | -.20 | .00 | -.14 | -.01 | -.19* | .10 | .07 | .05 | .14 | .66** | .79** | .78** | .76** | - | |
| 16. Total Job Performance | 19.34 | 3.62 | .22* | -.24** | -.01 | -.10 | -.04 | -.19* | .10 | .09 | .05 | .10 | .79** | .91** | .93** | .89** | .90** | - |

**Correlation is significant at the .01 level (2-tailed)

*Correlation is significant at the .05 level (2-tailed)

Table 2b

*Means, Standard Deviations and Intercorrelations for Consistency Indices of AC Dimensions and Peer Ratings of Job Performance*

| | M | SD | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1. C-PA 1 | .77 | .10 | - | | | | | | | | | | | | | | | |
| 2. C-PA 2 | .79 | .10 | -.11 | - | | | | | | | | | | | | | | |
| 3. C-MA 1 | .84 | .11 | -.06 | .00 | - | | | | | | | | | | | | | |
| 4. C-MA 2 | .81 | .17 | .09 | .10 | .14 | - | | | | | | | | | | | | |
| 5. C-TDK 1 | .77 | .09 | .18* | .10 | .07 | .12 | - | | | | | | | | | | | |
| 6. C-TDK 2 | .78 | .11 | -.08 | .27** | .06 | .08 | -.03 | - | | | | | | | | | | |
| 7. C-OC 1 | .79 | .09 | .01 | -.05 | .21* | .02 | .04 | -.02 | - | | | | | | | | | |
| 8. C-OC 2 | .78 | .10 | .09 | .07 | .11 | .12 | .10 | .25** | -.01 | - | | | | | | | | |
| 9. Total C 1 | 3.93 | .72 | -.01 | .11 | -.06 | -05 | .05 | .02 | -.07 | .07 | - | | | | | | | |
| 10. Total C 2 | 4.25 | .76 | .07 | .04 | .06 | .05 | .04 | -.09 | .05 | .07 | .44** | - | | | | | | |
| 11. Customer Service | 3.98 | .72 | .12 | .29* | -.02 | .02 | .02 | .07 | -.03 | .06 | .04 | .18 | - | | | | | |
| 12.Work Quantity | 3.79 | .89 | .12 | .05 | .02 | .17 | .04 | -.14 | -.05 | .10 | -.02 | .08 | .43** | - | | | | |
| 13. Work Quality | 3.88 | .79 | .13 | .11 | .13 | .18 | .10 | -.22 | .02 | .19 | .04 | .14 | .36** | .77** | - | | | |
| 14. Job Knowledge | 3.90 | .73 | .12 | .07 | .01 | .30* | .11 | -.06 | -.07 | .18 | .01 | -.02 | .19 | .56** | .76** | - | | |
| 15. Teamwork | 4.12 | .84 | .16 | .20 | .02 | .05 | .13 | -.08 | .01 | .19 | .00 | .27* | .64** | .66** | .57** | .33** | - | |
| 16. Total Job Performance | 19.66 | 3.14 | .17 | .18 | .04 | .18 | .10 | -.11 | -.03 | .18 | .02 | .17 | .66** | .88** | .88** | .71** | .82** | - |

**Correlation is significant at the .01 level (2-tailed)

*Correlation is significant at the .05 level (2-tailed)

Table 3a

*Means, Standard Deviations, and Intercorrelations for Consistency Indices and Standard Deviations for AC Dimensions at Time 1 and Time 2*

| | M | SD | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1. C-PA 1 | .77 | .10 | - | | | | | | | | | | | | | | | |
| 2. C-PA 2 | .79 | .10 | -.11 | - | | | | | | | | | | | | | | |
| 3. C-MA 1 | .84 | .11 | -.06 | .00 | - | | | | | | | | | | | | | |
| 4. C-MA 2 | .81 | .17 | .09 | .10 | .14 | - | | | | | | | | | | | | |
| 5. C-TDK 1 | .77 | .09 | .18* | .10 | .07 | .12 | - | | | | | | | | | | | |
| 6. C-TDK 2 | .78 | .11 | -.08 | .27** | .06 | .08 | -.03 | - | | | | | | | | | | |
| 7. C-OC 1 | .79 | .09 | .01 | -.05 | .21* | .02 | .04 | -.02 | - | | | | | | | | | |
| 8. C-OC 2 | .78 | .10 | .09 | .07 | .11 | .12 | .10 | .25** | -.01 | - | | | | | | | | |
| 9. SD-PA 1 | .98 | .44 | 1.00** | .11 | .06 | -.09 | -.18* | .08 | -.01 | -.09 | - | | | | | | | |
| 10. SD-PA 2 | .95 | .44 | .11 | -1.00** | .00 | -.10 | -.10 | -.27** | .05 | -.07 | -.11 | - | | | | | | |
| 11. SD-MA 1 | .65 | .45 | .07 | .01 | -.98** | -.16* | -.08 | -.08 | -.21** | -.09 | -.07 | -.01 | - | | | | | |
| 12. SD-MA 2 | .73 | .53 | -.08 | -.12 | -.09 | -.83** | -.13 | -.03 | -.09 | -.10 | .08 | .12 | .12 | - | | | | |
| 13. SD-TDK 1 | .96 | .38 | -.18* | -.10 | -.07 | -.12 | -1.00** | .03 | -.04 | -.10 | .18* | .10 | .08 | .13 | - | | | |
| 14. SD-TDK 2 | .95 | .46 | .08 | -.27** | -.06 | -.08 | .03 | -1.00** | .02 | -.25** | -.08 | .27** | .08 | .03 | -.03 | - | | |
| 15. SD-OC 1 | .90 | .37 | -.01 | .05 | -.21* | -.02 | -.04 | .02 | -1.00** | .01 | .01 | -.05 | .21** | .09 | .04 | -.02 | - | |
| 16. SD-OC 2 | .92 | .42 | -.09 | -.07 | -.11 | -.12 | -.10 | -.25** | .01 | -1.00** | .09 | .07 | .09 | .10 | .10 | .25** | -.01 | - |

**Correlation is significant at the .01 level (2-tailed)

*Correlation is significant at the .05 level (2-tailed)

Key: SD-PA 1 is standard deviation of problem analysis at time 1
SD-PA 2 is standard deviation of problem analysis at time 2
SD-MA 1 is standard deviation of management ability at time 1
SD-MA 2 is standard deviation of management ability at time 2
SD-TDK 1 is standard deviation of technical departmental knowledge at time 1
SD-TDK 2 is standard deviation of technical departmental knowledge at time 2
SD-OC 1 is standard deviation of oral communication at time 1
SD-OC 2 is standard deviation of oral communication at time 2

Table 3b

*Means, Standard Deviations, and Intercorrelations for Consistency Indices and Ranges for AC Dimensions at Time 1 and Time 2*

| | M | SD | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1. C-PA 1 | .77 | .10 | - | | | | | | | | | | | | | | | |
| 2. C-PA 2 | .79 | .10 | -.11 | - | | | | | | | | | | | | | | |
| 3. C-MA 1 | .84 | .11 | -.06 | .00 | - | | | | | | | | | | | | | |
| 4. C-MA 2 | .81 | .17 | .09 | .10 | .14 | - | | | | | | | | | | | | |
| 5. C-TDK 1 | .77 | .09 | .18* | .10 | .07 | .12 | - | | | | | | | | | | | |
| 6. C-TDK 2 | .78 | .11 | -.08 | .27** | .06 | .08 | -.03 | - | | | | | | | | | | |
| 7. C-OC 1 | .79 | .09 | .01 | -.05 | .21* | .02 | .04 | -.02 | - | | | | | | | | | |
| 8. C-OC 2 | .78 | .10 | .09 | .07 | .11 | .12 | .10 | .25** | -.01 | - | | | | | | | | |
| 9. R-PA 1 | 2.18 | 1.01 | -.98** | .12 | .05 | -.08 | -.18* | .08 | -.03 | -.09 | - | | | | | | | |
| 10. R-PA 2 | 2.05 | 1.00 | .10 | -.98** | .02 | -.09 | -.08 | -.27** | .05 | -.05 | -.11 | - | | | | | | |
| 11. R-MA 1 | .92 | .64 | .07 | .01 | -.98** | -.16* | -.08 | -.08 | -.21** | -.09 | -.06 | -.02 | - | | | | | |
| 12. R-MA 2 | 1.09 | .81 | -.07 | -.11 | -.08 | -.80** | -.14 | -.01 | -.09 | -.09 | .08 | .09 | .10 | - | | | | |
| 13. R-TDK 1 | 2.11 | .84 | -.15 | -.11 | -.05 | -.13 | -.98** | .04 | -.04 | -.11 | .15 | .09 | .07 | .15 | - | | | |
| 14. R-TDK 2 | 2.04 | 1.04 | .11 | -.26** | -.07 | -.08 | .05 | -.98** | .01 | -.25** | -.11 | .27** | .09 | -.03 | -.06 | - | | |
| 15. R-OC 1 | 1.99 | .86 | -.04 | .03 | -.20* | -.03 | -.06 | .03 | -.98** | -.02 | .06 | -.03 | .20* | .09 | .05 | -.02 | - | |
| 16. R-OC 2 | 1.97 | .95 | -.08 | -.05 | -.10 | -.13 | -.11 | -.26** | .03 | -.98** | .08 | .05 | .10 | .07 | .11 | .27** | .01 | - |

**Correlation is significant at the .01 level (2-tailed)

*Correlation is significant at the .05 level (2-tailed)

Key: R-PA 1 is range of problem analysis at time 1
R-PA 2 is range of problem analysis at time 2
R-MA 1 is range of management ability at time 1
R-MA 2 is range of management ability at time 2
R-TDK 1 is range of technical departmental knowledge at time 1
R-TDK 2 is range of technical departmental knowledge at time 2
R-OC 1 is range of oral communication at time 1
R-OC 2 is range of oral communication at time 2

Table 3c

*Means, Standard Deviations, and Intercorrelations for AC Dimension Standard Deviations and Supervisor Ratings of Job Performance*

| | M | SD | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1. SD-PA 1 | .98 | .44 | - | | | | | | | | | | | | | |
| 2. SD-PA 2 | .95 | .44 | -.11 | - | | | | | | | | | | | | |
| 3. SD-MA 1 | .65 | .45 | -.07 | -.01 | - | | | | | | | | | | | |
| 4. SD-MA 2 | .73 | .53 | .08 | .12 | .12 | - | | | | | | | | | | |
| 5. SD-TDK 1 | .96 | .38 | .18* | .10 | .08 | .13 | - | | | | | | | | | |
| 6. SD-TDK 2 | .95 | .46 | -.08 | .27** | .08 | .03 | -.03 | - | | | | | | | | |
| 7. SD-OC 1 | .90 | .37 | .01 | -.05 | .21** | .09 | .04 | -.02 | - | | | | | | | |
| 8. SD-OC 2 | .92 | .42 | .09 | .07 | .09 | .10 | .10 | .25** | -.01 | - | | | | | | |
| 9. Customer Service | 3.74 | .74 | -.21* | .23* | .09 | .10 | .00 | .20* | -.07 | .05 | - | | | | | |
| 10. Work Quantity | 3.75 | .90 | -.19* | .17 | .02 | .03 | .08 | .12 | -.15 | -.15 | .61** | - | | | | |
| 11. Work Quality | 3.88 | .86 | -.23* | .23* | -.01 | -.03 | .02 | .16 | -.12 | -.13 | .63** | .87** | - | | | |
| 12. Job Knowledge | 3.99 | .74 | -.21* | .26* | .01 | .02 | .06 | .20* | -.01 | -.08 | .65** | .73** | .83** | - | | |
| 13. Teamwork | 3.99 | .85 | -.14 | .20* | .01 | .07 | .01 | .19* | -.10 | -.07 | .66** | .79** | .78** | .76** | - | |
| 14. Total Job Performance | 19.34 | 3.62 | -.22* | .24** | .02 | .04 | .04 | .19* | -.10 | -.09 | .79** | .91** | .93** | .89** | .90** | - |

**Correlation is significant at the .01 level (2-tailed)

*Correlation is significant at the .05 level (2-tailed)

Table 3d

*Means, Standard Deviations, and Intercorrelations for AC Dimension Standard Deviations and Peer Ratings of Job Performance*

| | M | SD | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1. SD-PA 1 | .98 | .44 | - | | | | | | | | | | | | | |
| 2. SD-PA 2 | .95 | .44 | -.11 | - | | | | | | | | | | | | |
| 3. SD-MA 1 | .65 | .45 | -.07 | -.01 | - | | | | | | | | | | | |
| 4. SD-MA 2 | .73 | .53 | .08 | .12 | .12 | - | | | | | | | | | | |
| 5. SD-TDK 1 | .96 | .38 | .18* | .10 | .08 | .13 | - | | | | | | | | | |
| 6. SD-TDK 2 | .95 | .46 | -.08 | .27** | .08 | .03 | -.03 | - | | | | | | | | |
| 7. SD-OC 1 | .90 | .37 | .01 | -.05 | .21** | .09 | .04 | -.02 | - | | | | | | | |
| 8. SD-OC 2 | .92 | .42 | .09 | .07 | .09 | .10 | .10 | .25** | -.01 | - | | | | | | |
| 9. Customer Service | 3.98 | .72 | -.12 | -.29* | .05 | -.07 | -.02 | -.07 | .03 | -.06 | - | | | | | |
| 10. Work Quantity | 3.79 | .89 | -.12 | -.05 | -.01 | -.24* | -.04 | .14 | .05 | -.10 | .43** | - | | | | |
| 11. Work Quality | 3.88 | .79 | -.13 | -.11 | -.13 | -.29* | -.10 | .22 | -.02 | -.19 | .36** | .77** | - | | | |
| 12. Job Knowledge | 3.90 | .73 | -.12 | -.07 | -.07 | -.37** | -.11 | .06 | .07 | -.18 | .19 | .56** | .76** | - | | |
| 13. Teamwork | 4.12 | .84 | -.16 | -.20 | -.01 | -.19 | -.13 | .08 | -.01 | -.19 | .64** | .66** | .57** | .33** | - | |
| 14. Total Job Performance | 19.66 | 3.14 | -.17 | -.18 | -.04 | -.29* | -.10 | .11 | .03 | -.18 | .66** | .88** | .88** | .71** | .82** | - |

**Correlation is significant at the .01 level (2-tailed)

*Correlation is significant at the .05 level (2-tailed)

Table 3e

*Means, Standard Deviations, and Intercorrelations for AC Dimension Ranges and Supervisor Ratings of Job Performance*

| | M | SD | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1. R-PA 1 | 2.18 | 1.01 | - | | | | | | | | | | | | | |
| 2. R-PA 2 | 2.05 | 1.00 | -.11 | - | | | | | | | | | | | | |
| 3. R-MA 1 | .92 | .64 | -.06 | -.02 | - | | | | | | | | | | | |
| 4. R-MA 2 | 1.09 | .81 | .08 | .09 | .10 | - | | | | | | | | | | |
| 5. R-TDK 1 | 2.11 | .84 | .15 | .09 | .07 | .15 | - | | | | | | | | | |
| 6. R-TDK 2 | 2.04 | 1.04 | -.11 | .27** | .09 | -.03 | -.06 | - | | | | | | | | |
| 7. R-OC 1 | 1.99 | .86 | .06 | -.03 | .20* | .09 | .05 | -.02 | - | | | | | | | |
| 8. R-OC 2 | 1.97 | .95 | .08 | .05 | .10 | .07 | .11 | .27** | .01 | - | | | | | | |
| 9. Customer Service | 3.74 | .74 | -.21* | .21* | .09 | .09 | .01 | .19* | -.11 | .07 | - | | | | | |
| 10. Work Quantity | 3.75 | .90 | -.21* | .16 | -.01 | -.01 | .07 | .11 | -.19* | -.12 | .61** | - | | | | |
| 11. Work Quality | 3.88 | .86 | -.24* | .22* | -.07 | -.07 | .02 | .16 | -.16 | -.10 | .63** | .87** | - | | | |
| 12. Job Knowledge | 3.99 | .74 | -.21* | .25** | -.03 | -.03 | .07 | .20* | -.05 | -.07 | .65** | .73** | .83** | - | | |
| 13. Teamwork | 3.99 | .85 | -.14 | .19* | .01 | .06 | .02 | .19* | -.15 | -.06 | .66** | .79** | .78** | .76** | - | |
| 14. Total Job Performance | 19.34 | 3.62 | -.23* | .23* | .02 | .01 | .04 | .19* | -.15 | -.07 | .79** | .91** | .93** | .89** | .90** | - |

**Correlation is significant at the .01 level (2-tailed)

*Correlation is significant at the .05 level (2-tailed)

Table 3f

*Means, Standard Deviations, and Intercorrelations for AC Dimension Ranges and Peer Ratings of Job Performance*

| | M | SD | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1. R-PA 1 | 2.18 | 1.01 | - | | | | | | | | | | | | | |
| 2. R-PA 2 | 2.05 | 1.00 | -.11 | - | | | | | | | | | | | | |
| 3. R-MA 1 | .92 | .64 | -.06 | -.02 | - | | | | | | | | | | | |
| 4. R-MA 2 | 1.09 | .81 | .08 | .09 | .10 | - | | | | | | | | | | |
| 5. R-TDK 1 | 2.11 | .84 | .15 | .09 | .07 | .15 | - | | | | | | | | | |
| 6. R-TDK 2 | 2.04 | 1.04 | -.11 | .27** | .09 | -.03 | -.06 | - | | | | | | | | |
| 7. R-OC 1 | 1.99 | .86 | .06 | -.03 | .20* | .09 | .05 | -.02 | - | | | | | | | |
| 8. R-OC 2 | 1.97 | .95 | .08 | .05 | .10 | .07 | .11 | .27** | .01 | - | | | | | | |
| 9. Customer Service | 3.98 | .72 | -.11 | -.31** | .05 | -.07 | -.07 | -.07 | .00 | -.04 | - | | | | | |
| 10. Work Quantity | 3.79 | .89 | -.12 | -.04 | -.01 | -.25* | -.06 | .16 | .01 | -.08 | .43** | - | | | | |
| 11. Work Quality | 3.88 | .79 | -.15 | -.11 | -.13 | -.27* | -.13 | .21 | -.05 | -.19 | .36** | .77** | - | | | |
| 12. Job Knowledge | 3.90 | .73 | -.15 | -.09 | -.04 | -.32** | -.13 | .05 | .07 | -.20 | .19 | .56** | .76** | - | | |
| 13. Teamwork | 4.12 | .84 | -.16 | -.20 | -.01 | -.20 | -.13 | .10 | -.06 | -.16 | .64** | .66** | .57** | .33** | - | |
| 14. Total Job Performance | 19.66 | 3.14 | -.17 | -.18 | -.04 | -.28* | -.13 | .12 | -.01 | -.17 | .66** | .88** | .88** | .71** | .82** | - |

**Correlation is significant at the .01 level (2-tailed)

*Correlation is significant at the .05 level (2-tailed)

Table 4a

*Means, Standard Deviations, and Intercorrelations for Mean Dimension and Overall AC scores and Supervisor Ratings of Job Performance*

| | M | SD | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1. PA 1 | 3.65 | .75 | - | | | | | | | | | | | | | | | |
| 2. PA 2 | 3.89 | .88 | .26** | - | | | | | | | | | | | | | | |
| 3. MA 1 | 4.38 | 1.01 | .60** | .31** | - | | | | | | | | | | | | | |
| 4. MA 2 | 4.53 | 1.07 | .25** | .74** | .26** | - | | | | | | | | | | | | |
| 5. TDK 1 | 3.55 | .81 | .63** | .21** | .45** | .18* | - | | | | | | | | | | | |
| 6. TDK 2 | 4.48 | .77 | .30** | .35** | .14 | .40** | .34** | - | | | | | | | | | | |
| 7. OC 1 | 4.13 | 1.11 | .38** | .47** | .46** | .53** | .41** | -.01 | - | | | | | | | | | |
| 8. OC 2 | 4.09 | .93 | .25** | .59** | .25** | .72** | .28** | .56** | .39** | - | | | | | | | | |
| 9. AC Total 1 | 79.64 | 4.65 | .20* | .42** | .33** | .28** | .24** | .02 | .49** | .25** | - | | | | | | | |
| 10. AC Total 2 | 81.39 | 4.74 | .21** | .16* | .12 | .21** | .23** | .54** | -.09 | .39** | .25** | - | | | | | | |
| 11. Customer Service | 3.74 | .74 | .03 | .12 | -.10 | .10 | .12 | .09 | .04 | .06 | -.11 | .12 | - | | | | | |
| 12. Work Quantity | 3.75 | .90 | .08 | .09 | .09 | .03 | .01 | .08 | .07 | .05 | .11 | .17 | .61** | - | | | | |
| 13. Work Quality | 3.88 | .86 | .04 | .07 | .10 | .01 | .03 | .12 | .03 | .07 | .08 | .25** | .63** | .87** | - | | | |
| 14. Job Knowledge | 3.99 | .74 | .00 | .12 | .02 | -.02 | -.01 | .02 | .02 | .05 | .09 | .18 | .65** | .73** | .83** | - | | |
| 15. Teamwork | 3.99 | .85 | .02 | .12 | .05 | .07 | .09 | .16 | .01 | .15 | .02 | .20* | .66** | .79** | .78** | .76** | - | |
| 16. Total Job Performance | 19.34 | 3.62 | .04 | .12 | .04 | .04 | .05 | .11 | .04 | .09 | .05 | .21* | .79** | .91** | .93** | .89** | .90** | - |

**Correlation is significant at the .01 level (2-tailed)

*Correlation is significant at the .05 level (2-tailed)

Table 4b

*Means, Standard Deviations, and Intercorrelations for Dimension and Overall AC scores and Peer Ratings of Job Performance*

| | M | SD | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1. PA 1 | 3.65 | .75 | - | | | | | | | | | | | | | | | |
| 2. PA 2 | 3.89 | .88 | .26** | - | | | | | | | | | | | | | | |
| 3. MA 1 | 4.38 | 1.01 | .60** | .31** | - | | | | | | | | | | | | | |
| 4. MA 2 | 4.53 | 1.07 | .25** | .74** | .26** | - | | | | | | | | | | | | |
| 5. TDK 1 | 3.55 | .81 | .63** | .21** | .45** | .18* | - | | | | | | | | | | | |
| 6. TDK 2 | 4.48 | .77 | .30** | .35** | .14 | .40** | .34** | - | | | | | | | | | | |
| 7. OC 1 | 4.13 | 1.11 | .38** | .47** | .46** | .53** | .41** | -.01 | - | | | | | | | | | |
| 8. OC 2 | 4.09 | .93 | .25** | .59** | .25** | .72** | .28** | .56** | .39** | - | | | | | | | | |
| 9. AC Total 1 | 79.64 | 4.65 | .20* | .42** | .33** | .28** | .24** | .02 | .49** | .25** | - | | | | | | | |
| 10. AC Total 2 | 81.39 | 4.74 | .21** | .16* | .12 | .21** | .23** | .54** | -.09 | .39** | .25** | - | | | | | | |
| 11. Customer Service | 3.98 | .72 | -.02 | .17 | -.07 | .17 | .12 | .13 | .09 | .10 | .07 | -.02 | - | | | | | |
| 12. Work Quantity | 3.79 | .89 | .00 | .10 | .02 | .05 | -.03 | .11 | -.04 | .02 | .05 | .05 | .43** | - | | | | |
| 13. Work Quality | 3.88 | .79 | .09 | .13 | .12 | .07 | -.02 | .24* | -.07 | .07 | .06 | .21 | .36** | .77** | - | | | |
| 14. Job Knowledge | 3.90 | .73 | .03 | -.07 | .10 | -.12 | .04 | .22 | -.12 | -.01 | -.01 | .25* | .19 | .56** | .76** | - | | |
| 15. Teamwork | 4.12 | .84 | .01 | .30** | -.08 | .27* | .01 | .12 | .06 | .15 | .12 | -.06 | .64** | .66** | .57** | .33** | - | |
| 16. Total Job Performance | 19.66 | 3.14 | .03 | .16 | .02 | .12 | .03 | .21 | -.02 | .08 | .07 | .11 | .66** | .88** | .88** | .71** | .82** | - |

**Correlation is significant at the .01 level (2-tailed)

*Correlation is significant at the .05 level (2-tailed)

Table 5

*Intercorrelations for Supervisor and Peer Ratings of Job Performance*

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1. Customer Service-S | - | | | | | | | | | | | |
| 2. Quantity-S | .61** | - | | | | | | | | | | |
| 3. Quality-S | .63** | .87** | - | | | | | | | | | |
| 4. Job Knowledge-S | .65** | .73** | .83** | - | | | | | | | | |
| 5. Teamwork-S | .66** | .79** | .78** | .76** | - | | | | | | | |
| 6. Total-S | .79** | .91** | .93** | .89** | .90** | - | | | | | | |
| 7. Customer Service-P | .38** | .34** | .27* | .19 | .23 | .33** | - | | | | | |
| 8. Quantity-P | .31** | .47** | .44** | .30* | .28* | .42** | .43** | - | | | | |
| 9. Quality-P | .22 | .42** | .40** | .26* | .27* | .37** | .36** | .77** | - | | | |
| 10. Job Knowledge-P | .13 | .21 | .24* | .14 | .16 | .21 | .19 | .56** | .76** | - | | |
| 11. Teamwork-P | .42** | .54** | .44** | .27* | .45** | .50** | .64** | .66** | .57** | .33** | - | |
| 12. Total-P | .37** | .51** | .46** | .30* | .36** | .47** | .66** | .88** | .88** | .71** | .82** | - |

**Correlation is significant at the .01 level (2-tailed)

*Correlation is significant at the .05 level (2-tailed)

Table 6

*Interrater Agreement Estimates for Supervisor and Peer Ratings of Job Performance*

| $r_{wg}$ for Supervisor Ratings of Job Performance | $r_{wg}$ for Peer Ratings of Job Performance |
|---|---|
| Customer Service = .726 | Customer Service = .738 |
| Quantity of Work = .598 | Quantity of Work = .604 |
| Quality of Work = .634 | Quality of Work = .697 |
| Job Knowledge = .726 | Job Knowledge = .736 |
| Teamwork/Cooperation = .642 | Teamwork/Cooperation = .648 |
| Aggregate Job Performance = .748 | Aggregate Job Performance = .810 |

APPENDIX G

| AC Dimension Score | Consistency Index | Standard Deviation |
|---|---|---|
| PA = 15 | .77 | .98 |
| MA = 18 | .84 | .65 |
| TDK = 15 | .77 | .96 |
| OC = 17 | .79 | .90 |

The table above represents hypothetical scores for a candidate in an AC. This figure is intended to serve as an example of how the standard deviation and/or the consistency index might be integrated into the calculation of overall AC dimension scores and presented to assessors in an operational AC.

1. The consistency index ranges from 0-1 with scores closer to 1 indicating a higher degree of consistency. If the consistency index for each AC dimension were to be included in overall AC scores, it should be added into the dimension scores or perhaps rescaled to reflect the scale of each dimension.

2. The standard deviation is the reverse of the consistency index and should not be simply added to AC dimension scores. Weighting overall AC dimension scores by the standard deviation for each dimension is one possible method of integrating the standard deviation. Additionally, the standard deviation for each dimension could be used to inform assessor ratings of consistency on each dimension using a likert scale format, similar to how the AC dimensions are scored.

Figure 4. Integrating Measures of Consistency into Assessment Center Scores