

CONSTRUCT VALIDATION OF ASSESSMENT CENTERS
USING EXTERNAL CRITERIA

Alecia J. Billington

A thesis submitted in partial fulfillment of
the requirements for the degree of
Master of Arts

Department of Psychology

Central Michigan University
Mount Pleasant, Michigan
November, 2010

Accepted by the Faculty of the College of Graduate Studies,
Central Michigan University, in partial fulfillment of
the requirements for the master's degree

Thesis Committee:

Neil Christiansen, Ph.D.

Committee Chair

Stuart Quirk, Ph.D.

Faculty Member

Kevin Love, Ph.D.

Faculty Member

October 15, 2010

Date of Defense

Roger Coles, Ph.D.

Dean
College of Graduate Studies

March 16, 2011

Approved by the
College of Graduate Studies

This is dedicated to my sister Adeline who still loves me even though she has to share me with graduate school. May you someday feel the freedom to pursue your dreams wherever they take you. When that time comes, I will be as supportive to you as you have been to me.

ACKNOWLEDGEMENTS

I want to thank the members of my thesis committee for their hard work and encouragement throughout the life of this project: Dr. Neil Christiansen (committee chair and professor of psychology), Dr. Kevin Love (professor of management), and Dr. Stuart Quirk (professor of psychology). These committee members provided feedback and support through numerous revisions of my thesis manuscript. Especially Dr. Christiansen, who took time out of his busy schedule to assist me in the many of the analyses that were performed on this dataset. I also wish to acknowledge the support of Central Michigan University in producing this work.

ABSTRACT

CONSTRUCT VALIDATION OF ASSESSMENT CENTERS USING EXTERNAL CRITERIA

by Alecia J. Billington

The purpose of the present study was to examine Assessment Center (AC) validity by using external criteria such as individual difference measures and Multi-Source Feedback (MSF) ratings. Typical AC validation research examines ratings of performance dimensions across exercises with the expectation that similar dimensions should be more highly correlated across different exercises than with different dimensions within the same exercise. This study looked to information outside of the AC in order to validate the ratings.

The main hypotheses tested were that AC dimensions would be more related to their corresponding MSF dimensions than they would be to the non-corresponding MSF dimensions. It was also expected that AC dimensions which were linked to specific individual difference variables would be more related to those variables than would other, less related dimensions. Therefore the association between AC dimensions and their related individual differences such as personality and cognitive ability were also examined for validity evidence.

Data were collected from upper and middle level managers ($N=274$) from a wide range of industries who participated in a developmental AC, were evaluated through MSF ratings and completed self-report Five Factor Model (FFM) of personality

inventories, and cognitive ability measures. The hypotheses were tested using correlational and chi-square analyses. The results showed that, in general, corresponding AC and MSF dimensions were more related than were the non-corresponding dimensions.

When it came to the hypotheses regarding individual differences, however, the hypotheses were not supported. Those AC dimensions that were hypothesized to be related to personality traits were not significantly more related than those that were not. The only exception was that AC dimensions linked to cognitive ability were more related to cognitive ability than were other dimensions. The amount of variance in AC ratings due to exercise performance was also examined in order to determine if removing this variance would change the validity of the AC overall. In general, removing exercise variance did decrease the criterion validity of the AC.

Based on the results of this study, it may be useful in the future to use external criteria such as MSF ratings and personality as a form of validation for ACs. This could aid researchers by providing a fuller picture of AC validity. In this particular AC, the validity was quite low and most of the variance in dimension ratings could not be explained fully by personality variables. Future implications such as designing ACs more likely to tap into personality constructs related to job performance are discussed.

TABLE OF CONTENTS

LIST OF TABLES	ix
CHAPTER	
I. INTRODUCTION	1
II. LITERATURE REVIEW	5
<i>History of the Assessment Center</i>	5
<i>The “Typical” Assessment Center</i>	6
<i>Face Validity</i>	8
<i>Predictive Validity</i>	9
<i>Construct Validity</i>	10
<i>External Construct Validity</i>	14
<i>Present Study and Hypotheses</i>	16
III. METHOD	23
<i>Sample Characteristics</i>	23
<i>Measures</i>	23
IV. RESULTS	25
<i>Assessment Center and Multi-Source Feedback Ratings</i>	25
<i>Assessment Center PEDRs and Individual Differences</i>	33
<i>Confirmatory Factor Analyses</i>	38
<i>Removing Exercise Variance</i>	38
V. DISCUSSION	42
Limitations	45
REFERENCES	48

LIST OF TABLES

TABLE	PAGE
1. <i>Dimensions Used in Present Study</i>	18
2. <i>Assessment Center Post Exercise Dimension Ratings</i>	27
3. <i>Relationships Between Assessment Center and Multi-Source Feedback Dimensions</i>	30
4. <i>Average Correlations Between Assessment Center and Multi-Source Feedback Dimensions</i>	31
5. <i>Correlations Between Corresponding Assessment Center Dimension Ratings and Multi-Source Feedback Dimensions</i>	33
6. <i>Correlations Between Assessment Center Dimensions and Individual Difference Variables</i>	35
7. <i>Average Correlations Between Assessment Center Ratings and Individual Difference Variables</i>	36
8. <i>Correlations Between Exercise Factors, Cognitive Ability, Personality and Performance Ratings</i>	39

CHAPTER I

INTRODUCTION

Assessment Centers (ACs) have long been used to select job candidates for various positions, especially those interested in management and leadership careers. They have become increasingly popular and well regarded in the business world due in no small part to their longstanding ability to predict job performance and career advancement (Byham, 1970; Gaugler, Rosenthal, Thornton, & Bentson, 1987). In stark contrast to industry's widespread acceptance of the assessment center, many within the academic community have expressed reservations about the construct validity of this method.

Researchers have put considerable time and effort into explaining the variance in scores accounted for by assessment center dimensions and exercises in the hopes of better understanding what exactly assessment centers are measuring (Bowler & Woehr, 2006; Collins, Schmidt, Sanchez-Ku, Thomas, McDaniel, & Le, 2003; Jones & Born, 2008; Kaufman, Jex, Love, & Libkuman, 1993; Lance, Newbolt, Gatewood, Foster, French, & Smith, 2000). The main methods used in the research literature for ascertaining the construct validity of the AC are to examine internal information about the structure of the dimensions and exercises (Bowler & Woehr, 2006; Sackett & Dreher, 1982; Woher & Arthur, 2003). Instead of considering possible external correlates of AC scores as important sources of construct validity, research in this area has focused almost exclusively on using multitrait-multimethod (MTMM) matrices to analyze the intercorrelations between dimensions and exercises.

Despite the dominance of these two methods of collecting and examining assessment center data, there is still widespread confusion and debate about what is being measured. This is evidence that there is room for a different approach to conceptualizing AC construct validity. In fact, there have already been calls to abandon the traditional MTMM approach when dealing with ACs because of its inability to shed light on a more precise understanding of what assessment centers measure (Lance et al., 2000).

Another major weakness of many studies in the assessment center literature is reliance on limited information and absence of external, job related criteria. This is not consistent with established approaches to construct validity which is a fixture of psychological science (Cronbach & Meehl, 1955). This broader approach to validity is concerned with including additional constructs within the AC and job performance nomological networks (Hoffman & Woehr, 2009; Lance et al., 2008). By doing this, predicting relevant outcomes *and* understanding psychometric properties of the assessment center can be achieved.

This viewpoint is in line with Lance et al. (2000) who suggested that researchers turn their attention towards “extending the nomological network of assessment center construct validity” (p.1). By examining construct validity using external correlates (such as ratings of actual job performance) it may be easier to understand what is being measured by assessment centers. This will also allow for a clearer picture of the relative predictive power of individual AC dimensions which is an important part of understanding multi-dimensional constructs such as AC performance (Wong et al., 2008).

The present study will integrate these recent advances in viewing AC construct validity by examining relationships between individual difference measures, such as

personality and cognitive ability, and matching assessment center dimensions and multi-source feedback dimension ratings. This is somewhat similar to studies in the multi-source feedback literature which have sought to validate self, peer, and supervisor reports of performance with assessment center performance (Atkins & Wood, 2002; Hagan, Konopaske, Bernardin, & Tyler, 2006; Hoffman & Woehr, 2009). These studies, however, either excluded individual difference variables as part of the construct validation or did not always line up the same dimensions that had been tested in an assessment center with those that were collected through multi-source feedback. Relationships between individual difference measures, such as personality and cognitive ability, and assessment center dimensions ratings will also be examined.

Multi-source feedback instruments are an ideal source of external information because of their centrality in many important organizational processes (London & Smither, 1995). They are often used in applied settings in order to provide feedback about job performance and can contribute useful information when it is time to make decisions about promotional and developmental opportunities (Church & Bracken, 1997). This runs parallel to the goals of an assessment center which are also focused on providing quality individual assessments for personnel decisions. Because of these similarities, a comparison between ratings on both methods provides an opportunity to test the construct validity of assessment centers.

By using an external validation approach to test construct validity, this research will be better equipped to highlight the predictive power of assessment center dimensions. Instead of solely relying on analysis of internal data and comparing ratings between assessors, other measurement methods and sources can be examined. This will

continue the work of expanding our understanding of assessment center validity and provide new avenues for future research.

CHAPTER II

LITERATURE REVIEW

Despite various debates in the literature surrounding the assessment center, there is a general consensus that obtaining a better understanding of exactly what it is measuring is valuable. The purpose of this section is to outline key issues in the construct validity of assessment centers. This involves both historical and theoretical analysis of the methods used in assessment centers as well as relevant streams of research in personality, cognitive, and situational measurement. A better understanding of the development of assessment centers, their implementation, and the methodological questions surrounding them, will inform the lines of inquiry in the present study.

History of the Assessment Center

Assessment centers have a long history being used as selection and development tools in applied settings. They first grew out of work by Henry Murray and his time spent at Harvard University Psychological Clinic in the 1930s (Murray, 1938). This venture was not originally focused on personnel selection but rather obtaining an assessment of a person's entire personality from several different psychological angles. There, Murray and his colleagues worked to develop an intensive battery of tests and assessments on a small group of male undergraduates. Each colleague would construct assessments based on their areas of expertise, administer them to the young men, and then the group would get together and discuss. These discussions were focused on reporting findings and impressions regarding each subject and would result in some type of collective agreement on the personalities of each of the subjects. From this Murray

and other colleagues eventually developed the first assessment center devoted to personnel selection (Bray, 1982).

This original assessment center devoted to personnel selection was performed for the United States' Office of Strategic Services (OSS) in order to select intelligence agents. It eventually included group activities and work simulations and was a product of collaboration between selection teams of both the US and British defense departments (Bray, 1982). This was one of the earliest assessment centers used for personnel selection and is the predecessor to today's most commonly used assessment centers.

American Telephone and Telegraph (AT&T) also contributed to the development and prominence of the assessment center methodology by undertaking a decades long longitudinal study of managerial lives. This research program relied heavily on an initial three day assessment center of each manager and was later supplemented by follow up interviews, surveys, and assessments. The full assessment center data was compared from year one to year eight and each were rated on the same dimensions which allowed observation of changes on these dimensions over time. This battery of longitudinal assessments was shown to be predictive of management potential even after candidates had been in the workplace for over twenty years (Bray, 1982). There has been a wealth of research based on these original studies and their encouraging findings about the predictive ability of assessment centers.

The "Typical" Assessment Center

Today, assessment centers are most commonly used as selection tools in order to identify job candidates for a wide range of occupations (Woehr & Arthur, 2003). Arthur, Day, McNelly, and Edens (2003), conducted a meta-analytic review of 34 independent

assessment centers and found that the dimensions on which most candidates were assessed could be sorted into one of six common dimensions. These six most common dimensions consist of: consideration and awareness of others, communication, drive, influencing others, organization and planning, and problem solving. Most of the assessment centers included in this study measured some permutation of these constructs and were often assessed at different levels of analysis with some being more specific and focused (e.g., aggression) and others very broad (e.g., drive). This is typical of most assessment centers which are used as a method to measure psychological constructs at varying degrees of specificity.

Assessment centers can vary substantially in the methods of measurement as well as the constructs measured, however there are similarities in exercises used to measure these constructs (Thornton & Rupp, 2006). The simulation exercise is the defining feature of the assessment center and often takes the form of: group tasks and discussions, oral presentations, in-baskets, interaction simulations, and interviews. Other forms of assessment typical to ACs are some type of psychological and cognitive ability test. The most commonly used exercises in the AC literature are leaderless group discussions, in-basket simulation, oral presentations, case analysis, behavioral interview, and psychological testing (Thornton & Rupp, 2006).

Although initially developed for the purposes of employee selection and classification, assessment centers have more recently become popular tools for employee development (Tillema, 1998). In fact, these types of ACs are so popular that it is estimated that 40% of organizations which use them do so for the purpose of employee development (Spsychalski et al., 1997). These assessment centers focus on capturing

dimensions and areas in which candidates have the potential for growth and are often the first step in determining an individualized development plan for assessees. Thus, the exercises are often focused on increasing learning, self-reflection, and include feedback and coaching activities (Thornton & Rupp, 2006). Instead of selecting employees out, they work to assist managers in the advancement of their skills through training and further assessment.

Face Validity

The longstanding popularity of ACs is due in no small part to the widespread acceptance among applicants, organizations, and the legal community of their face validity. This is likely the result of the exercises found in the typical assessment center which act as simulations of many of the tasks important for successful performance in the given position (Thornton & Rupp, 2006). In contrast with personality tests, which are often viewed as opaque and irrelevant to many laypeople, assessment centers offer an opportunity in which to display tangibly job related competencies.

Assessment centers are also viewed favorably by the courts and could serve as a way for organizations to protect themselves against claims of unfair treatment and discrimination (Coulton & Field, 1995). This is because their predictive validity does not differ across race or gender unlike other selection devices such as tests of cognitive ability (Huck & Bray, 1976). This resistance to legal challenges and acceptance by both employers and applicants are evidence for the face validity of ACs.

Predictive Validity

Perhaps the most compelling reason ACs are a popular fixture in the world of management selection owes to the consensus on their predictive validity. Research has shown that Overall Assessment Ratings (OAR) have an average predictive validity of .29 and are related to several important variables such as job performance, ratings of career potential, and subsequent career advancement (Gaugler, Rosenthal, Thornton, & Bentson, 1987; Byham, 1970). Longitudinal studies of assessment centers have also found evidence supporting their ability to predict future career success (Jansen & Vinkenburg, 2005). Ratings of college students have been found to be predictive of subsequent promotion, even 16 years after the initial ratings were collected (Bray, Campbell, & Grant, 1974).

Supervisory ratings of performance are one of the most commonly used methods of assessing job performance and any widely used selection device should be predictive of performance on these ratings. In a meta-analysis of over 5,000 assessment centers, the OAR had a corrected correlation of .28 with supervisory ratings of performance (Hermelin, Lievens, & Robertson, 2007). It is likely that this – and other – estimates of OAR correlations to performance measures are conservative due to the presence of indirect range restriction. This is because assessees are often pre-selected before completing an AC and then further restriction occurs during the selection process before the performance ratings can even be collected.

Performance in a training program is another outcome of which the AC ratings display predictive and even incremental validity over cognitive ability and personality. In research conducted on managers completing a cross-cultural training program, AC

ratings of various dimensions were found to have incremental validity over cognitive ability and conceptually related personality traits (Lievens, Harris, Van Keer, & Bisqueret, 2003). Even more traditional assessment centers focused on prediction of promotion potential have found AC ratings to provide incremental validity over the standard self-report personality test. This phenomenon could be explained by conceptualizing AC ratings and personality traits as sampling different domains due to their frequent lack of correlations between one another (Goffin, Rothstein, & Johnston, 1996).

These evidences of predictive validity are important and encouraging indicators to organizations who hope to use assessment centers as selection devices. Because ACs seem to tap into behaviors and competencies related to advancement, businesses feel certain that they truly are getting job candidates with promise.

Construct Validity

Although assessment centers possess predictive validity, general legal acceptance, and face validity they do come with a set of challenges. In fact, it is unclear exactly what assessment centers are measuring. This lack of construct validity is one of the most common criticisms of assessment centers (Klimoski & Brickner, 1987). More specifically, assessment centers often appear to lack two of the key components of construct validity: convergent and discriminant validity. This is most often observed in correlations between AC dimension and exercise ratings.

For the past few decades, ratings of different behavioral dimensions within the same exercise have consistently been shown to be more highly correlated with one another than with similar dimensions measured across exercises (e.g., Sackett & Dreher,

1982). This phenomenon, until recently, has almost exclusively been studied using the internal MTMM approach which examines the intercorrelations between dimensions and exercises within a given assessment center. This method treats variance due to exercises as invalid error and a sign of poor construct validity within ACs (Lance et al., 2000; Sackett & Dreher, 1982).

Based on the original conceptualization of the MTMM approach developed by Campbell and Fiske (1959), dimensions are viewed as psychological traits and exercises as measurement methods. This leads to the assumption that variance attributable to dimensions is evident of trait variance and that variance due to exercises is method bias and should be regarded as error (Sackett & Dreher, 1982). Proponents of this view see such patterns of variance as being due to contamination of the method. These findings of “exercise effects” are widespread in the AC literature and are evidenced by lack of convergent validity between dimensions across exercises, lack of divergent validity between ratings of different dimensions on the same exercise, and the presence of a factor structure that can be explained by exercise variance (Lance et al., 2000).

However, there are compelling alternative explanations for these patterns of variance which do not discount the construct validity of assessment centers. The most promising approach involves viewing the tendency for the majority of variance to be explained by exercise factors as important, systematic variance which is related to relevant predictor constructs (Lance et al., 2000). This is in stark contrast to the pessimism surrounding the traditional MTMM approach to the validity of assessment center dimensions.

In fact, there are several streams of research in this area which point to the interpretation that variance across exercises can be explained by cross situational specificity in assessee behavior (Lance et al, 2000; Bowler & Woehr, 2006). These differences in action from one exercise to another can be explained by the interactionist approach, which emphasizes the interplay between characteristics of the person and the situation (Pervin, 1989). Under this framework, it would make more sense to view exercise variance as a product of situational differences in skill requirements, task characteristics, and an individual's differing levels of such necessary knowledge, skills, and abilities. This could be a more plausible explanation for exercise variance and open up new lines of investigation into assessment centers.

Indeed, performance related factors such as job knowledge and cognitive ability have been shown to be related to exercise variance in assessment centers (Jackson, Stillman, & Atkins, 2005; Lance Foster, Nemeth, Gentry, & Drollinger, 2004; Lance et al., 2000). This is more credible evidence that such variance is not random error but systematically related to important predictors of job performance.

At the foundation of this new way of viewing exercise variance, is the denial of the assumption that candidate behaviors within an assessment center should be cross-situational. It is possible to expect this lack of consistency since AC exercises provide unique opportunities to observe distinct behaviors (Howard, 1997). A more modern way to look at exercise variance would be through the lens of Trait Activation Theory (TAT), which would not expect that dimensions correlate across exercises unless those exercises have similar "activation potential" for the given trait (Tett & Guterman, 2000; Haaland & Christiansen, 2002). For example, a trait such as extraversion may manifest itself

differently depending on the exercise in which it is observed. It is also possible that an exercise such as an in – basket would provide more cues for dimensions related to the trait of conscientiousness than a leaderless group discussion and thus we should not expect dimensions related to this to correlate highly with one another across such exercises (Haaland & Christiansen, 2002).

The basis of this theory can be traced back to interactionist theories of personality which stress the importance of *both* the person and the situation in predicting future performance (Pervin, 1989). This draws upon research which emphasizes the impact of situational cues on subsequent behavior and does not expect cross-situational consistency unless cues for behavior are also similar (Wright & Mischel, 1988). Since assessment center exercises are created in order to be divergent from one another in the type of behavior they call for and later sample (Sackett & Harris, 1988), it should not be expected that dimensions display convergent validity across exercises (Haaland & Christiansen, 2002). Instead of viewing exercises as repeated measures of the same trait, they can be thought of as measurement methods intended to evoke divergent behaviors (Lievens et al., 2006).

Indeed, when the Trait Activation Potential (TAP) for each dimension is assessed by exercise, ratings from exercises with similar TAPs display higher convergence than those with dissimilar TAPs (Haaland & Christiansen, 2002; Lievens et al., 2006). This is evidence for the argument that the traditional MTMM Model is not an appropriate lens through which to view the construct validity of ACs since it places unrealistic expectations on correlations across exercises (Lance et al., 2000). Thus, instead of viewing variance due to exercises as error, researchers should focus their efforts on new

methods of demonstrating construct validity that do not depend on the traditional use of MTMM matrices.

Such information begs the question: What about assessment center dimensions? Some researchers have gone so far as to suggest that ratings should take place only on tasks within exercises instead of the traditional ratings of dimensions (Jackson, Stillman, & Atkins, 2005; Lance et al., 2000; Lance, 2008). If realized, this suggestion may lead to assessment center professionals “throwing the baby out with the bathwater”.

Before dimensions are discounted as faulty measures, it is important to understand what job relevant outcomes they may be predicting outside of an assessment center. Examining the dimensions that compose overall assessment center performance will allow for a better understanding of the relationships between dimensions and external constructs (Wong, Law, & Huang, 2008). This is because it is possible that there are certain dimensions which could more strongly predict external constructs, or vice versa. These external constructs could include information relevant to the job at hand and include ratings of job performance, job knowledge, and cognitive ability.

External Construct Validity

As mentioned previously, including external constructs in the validation framework is a longstanding tradition within psychological science (Cronbach & Meehl, 1955). In this approach, a nomological network is “elaborated” in which the construct of interest resides and related constructs are also identified which allows for a better understanding of the relationships between constructs of interest and the “laws” governing their interconnections. By using a more holistic approach to validating ACs, it is more likely that researchers will be better equipped to “develop a theory of

performance that explains how an individual can (or will) meet the demands of a particular job” (Landy, 1986, p.1188).

More specifically within the AC domain, external construct validation would presuppose convergent validity between constructs within the AC and similar constructs in the job setting. Likewise, there should be divergent validity between dissimilar dimensions in the AC and job (Atkins & Wood, 2002). This follows assertions that convergent validity should not be expected unless it is based on appropriate theory (Cronbach & Meehl, 1955).

The nomological approach to construct validity has been used in the literature concerned with the validation of Multi-Source Feedback (MSF) by use of assessment center ratings and individual difference variables (Hoffman & Woehr, 2009; Beehr, Ivanitskaya, Hansen, Erofeev, & Gudanowski, 2001). MSF systems are valid indicators of on the job performance because of their broad focus on a variety of performance indicators. This is where the similarity between MSF and ACs is clear, since each of them focus on measuring a broad sample of behaviors on job relevant dimensions beyond traditional indicators of performance like salary, advancement, or only supervisory ratings (Atkins & Wood, 2002). There are differences between MSF and ACs to be sure, in regards to the type of measurement used, rater training, and range of observable behaviors. However, because of the parallel nature of these measurements, the assessment center and multi-source feedback instruments are uniquely suited to be used for validation projects against one another.

Using this concept, Atkins and Wood (2002) conducted a study in order to examine the relationships between ratings of performance on an MSF instrument and an

AC. They found that ratings of performance on an MSF, aggregated across raters (supervisors, co-workers, and subordinates), were significantly related to overall ratings of performance on an assessment center. Others have also compared ratings on ACs and MSF as a form of criterion validity, comparing overall ratings between each method. MSF ratings have been shown to be related to overall performance on assessment centers. In regards to construct validity, there have also been promising findings which indicate that measures of competencies in an MSF are significantly correlated with similar competencies in an AC (Hagan et al., 2006). Both of these types of ratings have been shown to have incremental validity over and above supervisory ratings of performance.

These forays into validation using both MSF and AC ratings highlight the need for a more comprehensive understanding of the relationships between the two. It is also important that construct validity evidence be examined by comparing ratings on the same competencies from both sources of information. This will allow researchers to have additional data regarding the construct validity of assessment center dimensions that does not rely on the traditional, internally focused MTMM approach. By using external sources of information, it will be possible to have a deeper understanding of what exactly AC dimensions are measuring and how related they are to specific measures of job relevant performance.

Present Study and Hypotheses

The present study used ratings from a MSF instrument, personality test, and cognitive ability indicators to evaluate the construct validity of an assessment center. This will serve as a way to examine evidence of convergent and divergent validity between similar AC and MSF dimensions. The study also attempted to demonstrate

convergent and divergent validity between AC dimensions and conceptually related personality and cognitive ability traits.

The dimensions that were measured in both the MSF and AC are listed in Table 1 and are organized under the overarching factor of work behavior of which they are most indicative. The factors represented are leadership, administrative, communication, and thinking. Because behaviors are largely cross-situational and stable between similar environments, and an assessment center is a high fidelity simulation, it is expected that convergent validity will be demonstrated. More specifically, convergent validity will be measured by comparing ratings of the AC and MSF with those that are similar being the most strongly related to one another (Thornton & Rupp, 2006). This builds on previous research which has found that MSF ratings and AC performance have positive relationships with one another (Hoffman, et al., 2008; Hagan et al., 2006; Atkins & Wood, 2002). Taking AC research in a new direction, it is expected that there will be a high degree of overlap between similar dimensions measured by both AC and MSF ratings. This external validation process will serve as a new way to conceptualize AC dimension validity, while still examining convergent and divergent validity estimates.

Hypothesis 1: Dimension ratings from the assessment center will be more positively related to ratings on the same dimensions of multi-source feedback instruments than different multi-source feedback dimensions.

Table 1. *Dimensions Used in Present Study*

<i>Dimension</i>	<i>Factor</i>	<i>Hypothesized Related Individual Difference Trait</i>	<i>Description</i>
Analyze Issues	Thinking	Conscientiousness, Cognitive Ability, Openness	Approaches issues from a broad perspective, considering a wide range of information and factors; grasps complexities and perceives relationships among problems or issues.
Use Sound Judgment	Thinking	Conscientiousness Cognitive Ability	Applies logic and experience to make timely, wise judgments.
Manage Execution	Administrative	Conscientiousness Cognitive Ability	Organizes and prioritizes work activities; delegates responsibility; monitors progress.
Lead Courageously	Leadership	Extraversion	Steps forward to address difficult issues; stands firm on behalf of the organization and key stakeholders.
Influences Others	Leadership	Extraversion Agreeableness	Persuades others, gains their support and commitment
Foster Teamwork	Leadership	Agreeableness, Openness, Extraversion	Uses terms and an empowering, collaborative approach on appropriate issues; fosters collaboration among teams and team members
Coach and Develop	Leadership	Extraversion, Agreeableness	Accurately assesses employees' strengths and development needs; provides feedback, coaching, and opportunities to develop.
Build Relationships	Leadership	Extraversion, Agreeableness	Initiates and develops relationships with a wide variety of people based on trust; shows interest in an understanding of others' needs and concerns.
Manage Disagreements	Leadership	Agreeableness, Emotional Stability	Brings substantive conflicts and disagreements into the open and attempts to resolve them collaboratively.
Foster Open Communication	Communication	Extraversion, Agreeableness	Ensures a smooth flow of information between self and others through clear speaking and writing, encouragement of open expression of ideas and effective listening.

Note: All dimensions were measured in both the assessment center and multi-source feedback inventory.

Although the main focus of this paper is the interrelationships between the MSF and AC ratings, another important method of validation should be the examination of convergent validity between AC dimensions and conceptually related constructs. These constructs which should be related to performance are personality and cognitive ability. It has been shown that personality traits are related to AC performance and since the dimensions in the present study are based on a competency model with descriptions of mid-level behaviors such as “learns new information quickly”, it follows that dimensions should be positively related to similar personality traits (Goffin et al., 1996). Indeed, it has been shown that certain AC dimensions provide the opportunity to observe behaviors related to important personality traits (Haaland & Christiansen, 2002). It is expected that AC dimensions related to a particular personality trait will be more highly related to their corresponding trait than other important individual difference traits.

In the present study, the sample in question is a group of managers currently involved in directing others. The data of the current study is organized heavily around factors involving leadership behaviors like the abilities to persuade others to get behind important tasks, providing feedback and coaching opportunities, and leading teams collaboratively. Since leadership measurement and prediction is one of the primary goals of the AC and MSF tools, it makes sense to examine the literature on leadership performance and personality predictors alongside the traditional personality and job performance literature. This will help to provide information regarding which AC dimensions are most likely to be related to which Five-Factor Model personality traits. With regards to job performance, Conscientiousness is one of the traits most consistently shown to be related to job performance across jobs (Barrick & Mount, 1991). This is

likely because conscientious individuals are usually described as diligent, dependable, and hard working. These behaviors are important in driving job performance broadly and have also emerged as particularly relevant for managers (Barrick & Mount, 1991; Judge et al., 2002). Conceptually, conscientiousness should be related to managerial job performance because of conscientious individuals' persistence and initiative which is necessary when directing others (Kirkpatrick & Locke, 1991). It is therefore expected that AC dimensions measuring behaviors related to Conscientiousness will be more closely related to this FFM trait than others.

Hypothesis 2: Assessment center dimensions most closely related to Conscientiousness will be more highly related to self-reports of Conscientiousness than other FFM traits.

Extraversion has also been shown to be related to job performance, although not necessarily in all jobs (Barrick & Mount, 1991). Instead, Extraversion appears to be uniquely predictive of managerial and leadership effectiveness (Judge et al., 2002). This owes largely to the fact that individuals described as extraverted are usually energetic, assertive, and engaging. These characteristics are important functions of leadership roles as well as teamwork and interpersonal competencies (Barrick, Stewart, Neubert, & Mount, 1998). Thus, it is expected that within the present sample of managerial participants that behaviors measured in the AC which are indicative of Extraversion will be related to self-report scores of Extraversion.

Hypothesis 3: Assessment center dimensions most closely related to Extraversion will be more highly related to self-reports of Extraversion than other FFM traits.

Interacting with team members and subordinates is a common function of many managerial positions. Emotional Stability is usually defined as the propensity to experience positive emotional states and to remain calm and secure under pressure.

These tendencies are important in building and maintaining close working relationships and have been found to be related to job performance while working within teams (Mount et al., 1998). These behaviors are indicative of dimensions measured in this study and it seems that individuals who possess high levels of emotional stability should flourish in difficult social interactions and high pressure environments often tapped by AC and MSF dimensions.

Hypothesis 4: Assessment center dimensions most closely related to Emotional Stability will be more highly related to self-reports of Emotional Stability than other FFM traits.

Along with the reasoning that leadership demands a level of teamwork and cooperation, it follows that Agreeableness would also be related to interpersonal, team work related dimensions. Agreeable individuals are the most likely to be seen as warm, helpful, and tolerant of others. These are all characteristics which would be expected of managers during both the assessment center and every day work situations. Indeed, individuals high in agreeableness have been found to be better performers at group leadership tasks (Barrick et al., 1998). Thus, it is expected that self-reports of Agreeableness will be related to conceptually similar dimensions

Hypothesis 5: Assessment center dimensions most closely related to Agreeableness will be more highly related to self-reports of Agreeableness than other FFM traits.

Openness to Experience has been shown to be related to effectiveness of managers, especially in business settings (Judge et al., 2002). This can most likely be explained by the need for individuals in business settings to be open to new, creative ways of solving problems. Indeed, individuals who are high in Openness are more likely to engage in divergent thinking when searching for solutions, are more likely to adapt to

change, both of which are important for dynamic business settings (McCrae, 1987). This suggests that measuring Openness to Experience is a valid method for which to examine the construct validity of an AC.

Hypothesis 6: Assessment center dimensions most closely related to Openness to Experience will be more highly related to self-reports of Openness to Experience than other FFM traits.

The predictive validity of cognitive ability is one of the most robust findings in the job performance literature. It has consistently been shown to be related to managerial success in organization, training performance, and career advancement (Bray & Grant, 1966; Hunter & Hunter, 1984). Because of these findings on the importance of cognitive ability in predicting managerial performance, this construct is especially suited to be used in a construct validation.

Hypothesis 7: Assessment center dimensions most closely related to cognitive ability will be more highly related to scores of cognitive ability than other constructs.

Together, these hypotheses specify interrelationships between assessment center and multi-source feedback ratings that hinge on findings in the personality and job performance literatures. This builds on evidence from other studies which have indicated that personality and cognitive ability are related to AC dimension ratings (Spector, Schneider, Vance, & Hezlett, 2000). By utilizing an external MTMM approach, it is hoped that examples of divergent and convergent validity between AC, MSF, personality, and cognitive ability scores will provide construct validity evidence for the controversial AC dimension.

CHAPTER III

METHOD

Sample Characteristics

The subjects ($N=274$) were a combination of middle and upper level managers from a variety of industries. The average age of participants was roughly 42 years old, 75% were male, and over 83% were Caucasian. The largest number of participants belonged to the retail industry (70%) with the rest hailing from a wide variety of industries ranging from manufacturing to transportation to healthcare. The average amount of organizational tenure per participant was approximately 10 years.

Measures

Assessment Center. All participants took part in an assessment center used for developmental purposes which was composed of a variety of exercises and were rated on a common set of dimensions (see Table 1). These exercises included:

1. **Behavioral interview:** participants responded to a series of behaviorally based interview questions.
2. **In-basket:** participants completed a task in which they received a variety of memos and other pieces of common workplace communications and were required to respond in a predetermined amount of time.
3. **Direct report:** participants were required to counsel a subordinate who came to them with a set of problems and challenges.
4. **Task force:** subjects participated in a group problem solving activity.

Multi-Source Feedback. Ratings of targets on the multi-source feedback instruments were collected on participants as a method of performance appraisal. The data were analyzed including supervisor, peer, and subordinate reports of performance.

Ratings were gathered on a variety of dimensions relating to constructs such as communication, leadership, analytical thinking and interpersonal skill (see Table 1).

Personality. Participants completed the Global Personality Inventory® (GPI), a multidimensional commercial personality test designed for use in a work context (Schmit, Kihm, & Robie, 2000). The GPI consists of 300 items that are rated on agreement using a 5-point Likert scale. Of the 37 facet scales, 30 have been linked to the FFM dimensions, with past research indicating adequate reliability and a replicable factor structure (Schmit et al., 2000). Composites were computed based on a combination of the factor structure described in the manual as well as past research conducted on the structure of the FFM (Christiansen, 2008).

Participants also completed the Wesman Personnel Classification Test (Pearson, 2007) in order to test general cognitive ability. This test is comprised of 40 items designed to measure both verbal and inductive problem solving. This test has been shown to be related to other measures of cognitive ability like the Watson-Glaser Critical Thinking Appraisal and the Wonderlic Personnel Test (Watson & Glaser, 1980).

Trait – Dimension Linkages. Each AC and MSF dimension was assigned to a individual difference trait it was hypothesized to best represented. These judgments were based on looking through the dimension descriptions, thinking of behavioral examples and linking this information back to personality descriptions found in the GPI. The hypothesized linkages can be found in Table 1.

CHAPTER IV

RESULTS

Assessment Center and Multi-Source Feedback Ratings

Traditional MTMM analyses were performed (see Table 2) as is typical of the AC validity literature (i.e., Bowler & Woehr, 2006; Gaugler, 1987). The data from these analyses showed moderate to weak relationships between PEDRs across exercises (average $r=.11$) and were lower than what is typically found ($r=.25$; Bowler & Woehr, 2006). Correlations between PEDRs of non-corresponding dimensions within the same exercises ($r=.37$) were also lower than past research ($r=.53$; Bowler & Woehr, 2006). Likewise, correlations between different dimensions across exercises ($r=.12$) were notably lower than what has been found in meta-analyses of similar studies ($r=.20$).

For the AC, post exercise dimension ratings (PEDRs) of participant performance were collapsed across exercises and reported as overall dimension scores. Similarly, MSF ratings were collapsed across rater source and reported as overall behavioral dimensions. Comparing correlations between these ratings, it was possible to directly test Hypothesis 1 which predicted that AC dimensions would be more highly correlated with their corresponding MSF dimensions than with unrelated dimensions.

Table 3 displays the results of these correlational analyses. In general, the correlations between AC and MSF dimensions were modest with correlations between corresponding dimensions ranging from $r=-.02$ to $r=.19$. It should also be noted that some of the dimensions that were not corresponding were actually slightly more highly correlated than were corresponding dimensions. For example, the AC dimension of

Manage Disagreements was correlated with the MSF dimension of Influence Others ($r=.04$) more highly than the MSF dimension of Manage Disagreements ($r=-.02$).

Table 2. *Assessment Center Post Exercise Dimension Ratings*

	1	2	3	4	5	6	7	8	9
1. IB-Analyze	--								
2. IB-Judgment	.65	--							
3. IB-Execution	.41	.66	--						
4. IB-Courage	.44	.64	.55	--					
5. IB-Influence	.50	.55	.42	.54	--				
6. IB-Coach	.40	.38	.39	.31	.38	--			
7. IB-Team	.33	.31	.24	.23	.33	.42	--		
8. IB-Relationships	.26	.27	.21	.11	.36	.48	.27	--	
9. IB-Communication	.49	.41	.37	.32	.44	.51	.60	.46	--

27

Table 2. *Assessment Center Post Exercise Dimension Ratings (continued)*

	1	2	3	4	5	6	7	8	9	10	11
10. DR-Analyze	.06	-.07	-.07	-.03	.13	.04	.08	.13	.07	--	--
11. DR-Judgment	-.01	-.12	-.12	-.03	.03	.04	-.02	.01	-.02	.62	--
12. DR-Execution	.09	-.02	-.02	-.02	.02	.04	-.01	.00	.02	.31	.61
13. DR-Courage	.02	-.02	-.12	.08	.02	-.03	-.09	-.11	-.11	.24	.49
14. DR-Influence	.09	.05	.02	.02	.05	.04	.07	.07	.07	.37	.40
15. DR-Coach	-.01	.02	.03	.04	.05	.04	-.07	-.01	-.01	.26	.28
16. DR-Relationship	.03	.02	.04	.03	.12	.10	.08	.14	.14	.24	.14
17. DR-Disagree	.06	.07	.06	.08	.09	.08	-.05	.04	.04	.39	.39
18. DR-Communication	.10	.13	.07	.04	.14	.08	.08	.12	.12	.32	.13

Table 2. *Assessment Center Post Exercise Dimension Ratings (continued)*

	1	2	3	4	5	6	7	8	9	10	11
19. TF-Analyze	.11	.11	.08	.09	.12	.12	.06	-.04	.11	.14	.13
20. TF-Judgment	.07	.03	.05	.07	.08	.07	-.02	-.10	.01	.21	.19
21. TF-Execution	.09	.11	.11	.18	.10	.09	-.01	-.07	.01	.11	.08
22. TF-Courage	.02	-.03	.01	.06	.07	.02	.02	-.10	.02	.23	.23
23. TF-Influence	.14	.12	.12	.12	.14	.14	.10	.05	.11	.22	.11
24. TF-Teamwork	.09	.09	.02	.14	.14	.09	.05	-.06	-.02	.08	.07
25. TF-Relationship	.05	.06	.02	.07	.07	.05	-.03	-.01	-.02	.11	.14
26. TF-Disagree	.09	.07	.03	.08	.08	.09	.01	-.04	.06	.19	.15
27. TF-Communicate	.06	.10	.06	.12	.12	.06	.02	-.01	.00	.12	.08

28

Table 2. *Assessment Center Post Exercise Dimension Ratings (continued)*

	12	13	14	15	16	17	18	19	20	21	22
10. DR-Analyze	--										
11. DR-Judgment	--										
12. DR-Execution	--										
13. DR-Courage	.50	--									
14. DR-Influence	.24	.28	--								
15. DR-Coach	.21	.20	.40	--							
16. DR-Relationship	.00	-.27	.44	.40	--						
17. DR-Disagree	.16	.22	.49	.35	.41	--					
18. DR-Communication	-.07	-.24	.44	.33	.68	.44	--				

Table 2. *Assessment Center Post Exercise Dimension Ratings (continued)*

	12	13	14	15	16	17	18	19	20	21	22
19. TF-Analyze	.05	.01	.05	-.06	-.04	-.01	.03	--			
20. TF-Judgment	.11	.16	.24	.06	.01	.13	.08	.65	--		
21. TF-Execution	.05	.09	.15	.06	-.01	.02	.05	.39	.63	--	
22. TF-Courage	.13	.21	.21	.19	.05	.11	.05	.20	.50	.38	--
23. TF-Influence	.08	.23	.23	.08	.06	.16	.05	.51	.60	.39	.35
24. TF-Teamwork	.05	.12	.12	.06	.02	.13	.03	.43	.42	.11	.05
25. TF-Relationship	.09	.18	.18	.14	.13	.17	.11	.27	.20	.11	-.11
26. TF-Disagree	.10	.21	.21	.10	.04	.19	.04	.46	.53	.33	.39
27. TF-Communication	.08	.07	.07	.04	.08	.14	.07	.47	.37	.21	-.02

29

Table 2. *Assessment Center Post Exercise Dimension Ratings (continued)*

	23	24	25	26	27
19. TF-Analyze	--				
20. TF-Judgment	--				
21. TF-Execution	--				
22. TF-Courage	--				
23. TF-Influence	--				
24. TF-Teamwork	.49	--			
25. TF-Relationship	.37	.61	--		
26. TF-Disagree	.63	.59	.44	--	
27. TF-Communication	.48	.65	.69	.49	--

Table 3. *Relationship between Assessment Center and Multi-Source Feedback Dimensions*

Assessment Center Dimension	Multi-Source Feedback Dimension				
	Analyze Issues	Use Sound Judgement	Manage Execution	Lead Courageously	Influence Others
Analyze Issues	.08	.08	.00	.07	.05
Use Sound Judgment	.06	.06	.05	.07	.05
Manage Execution	.13*	.11	.12*	.09	.10
Lead Courageously	.10	.20**	.18**	.19**	.18**
Influence Others	.05	.02	-.01	.03	.06
Coach and Develop	-.01	-.04	.01	-.04	.01
Foster Teamwork	.07	-.01	.03	-.01	.04
Build Relationships	.01	-.04	.00	-.02	.03
Foster Open Communication	.12*	.07	.05	.01	.06
Manage Disagreements	.01	-.01	-.03	.01	.04

Table 3. *Relationship between Assessment Center and Multi-Source Feedback Dimensions (continued)*

Assessment Center Dimension	Multi-Source Feedback Dimension				
	Coach and Develop	Foster Teamwork	Build Relationships	Foster Open Communication	Manage Disagreements
Analyze Issues	.04	.03	.04	.04	.04
Use Sound Judgment	.03	.02	.00	.02	.01
Manage Execution	.07	.04	.01	.07	.02
Lead Courageously	.19**	.18**	.12*	.18**	.14*
Influence Others	.04	.06	.03	.06	.06
Coach and Develop	.03	.05	.04	.01	.08
Foster Teamwork	.00	.11	.11	.05	.11
Build Relationships	.02	.11	.10	.08	.12
Foster Open Communication	.04	.11	.11	.11	.13*
Manage Disagreements	.02	.00	-.02	.02	-.02

Note: $N=274$, $*=p<.05$, $**=p<.01$. Relationships that are bolded were hypothesized to be higher.

Although these individual correlations do provide some, marginal evidence of convergence between similar dimensions, further analyses were undertaken in order to better understand the general patterns of relationships. Averages of the correlations were next computed in order to determine the general patterns of correlations between corresponding and non-corresponding AC and MSF dimensions and are presented in Table 4. The mean ($r=.08$) and median ($r=.09$) of the correlations between corresponding AC and MSF dimensions were slightly higher than the mean ($r=.05$) and median ($r=.05$) correlations between non-corresponding dimensions. The correlations examined for Hypothesis 1 were next subjected to a chi-square test in order to determine if the average correlation between related AC and MSF dimensions was significantly different than from the averages of dissimilar dimensions.

Table 4. *Average Correlations between Assessment Center and Multi-Source Feedback Dimensions*

<i>Assessment Center Dimensions</i>	<i>Multi-Source Feedback Dimensions</i>		
	<u>Corresponding</u> Mean r	<u>Non-Corresponding</u> Mean r Median r	
Analyze Issues	.08	.06	.06
Use Sound Judgment	.06	.04	.02
Manage Execution	.12	.02	.01
Lead Courageously	.19	.02	.01
Influence Others	.06	.06	.05
Coach and Develop	.03	.05	.03
Foster Teamwork	.11	.07	.05
Build Relationships	.10	.05	.04
Foster Open Communication	.11	.06	.05
Manage Disagreements	-.02	.08	.08
Total	.08	.05	.05

Note: Median $r=.09$ for Corresponding AC Dimensions, $\chi^2=34.68$, $p<.01$, $k=10$ for Corresponding Dimensions, $k=90$ for Non-Corresponding Dimensions.

Because the observed correlations were not independent and were not expected to reflect a normal distribution, a chi-square test with one degree of freedom was used. This

is similar to analyses conducted in similar studies (Haaland & Christiansen, 2002; Lievens et al., 2006). Pairwise comparisons were made between correlations of related and unrelated dimensions and individual difference traits. The equation for this test was used as described in Welkowitz, Ewen, and Cohen (1982) such that the chi-square value equals the number of positive difference values (f_p) minus the number of negative values (f_m) squared and divided by the total number of pairs (N).

$$\chi^2 = \frac{(f_p - f_m)^2}{N}$$

The results of these analyses are reported in Table 4 and show that the averages for the related AC and MSF dimensions (mean $r=.08$, median $r=.09$) were significantly larger ($\chi^2=34.68$, $p<.01$) than the averages of the unrelated dimensions (mean $r=.05$, median $r=.05$). Based the results of these analyses, Hypothesis 1 was generally supported.

Although Hypothesis 1 was supported, the correlations between AC dimensions and MSF ratings were low when compared with the rest of the AC literature. For example, another research study found the correlation between OAR and overall performance ratings to be $r=.25$ (Gaugler et al., 1987) whereas the present study found that relationship to be notably lower ($r=.10$). Not surprisingly, these findings echo the inability of the dimensions to predict overall performance and can be found in Table 5. The predictive validity of the AC dimensions was weak with correlations ranging from $r=.00$ to $r=.09$. The only dimension that had a significant, moderate relationship with overall performance was Lead Courageously, $r=.19$.

Table 5. *Correlations between Corresponding Assessment Center Dimension Ratings and Multisource Feedback*

Assessment Center Dimension Score	Relationship with Corresponding Performance Dimension	Relationship with Overall Performance	Semi-Partial <i>r</i> Controlling for Exercise Variance
	Zero-Order Correlation	Zero-Order Correlation	
Analyze Issues	.10	.05	-.04
Use Sound Judgment	.02	.04	-.05
Manage Execution	.08	.08	.02
Lead Courageously	.20*	.19*	.06*
Influence Others	.06	.05	-.08
Foster Teamwork	.11	.06	-.09
Coach and Develop	.02	.02	-.01
Build Relationships	.09	.05	-.02
Manage Disagreements	-.01	.00	.02
Openness to Communication	.09	.09	-.04
OAR	--	.10	-.08

Note: *N*=274, OAR=Overall Assessment Rating, *=*p*<.05

Assessment Center PEDRs and Individual Differences

Hypotheses 2-7 stated that AC PEDRs judged to be related to FFM and cognitive ability measures would be more correlated with one another than with PEDRs measuring unrelated individual difference measures. The results of these correlational analyses are reported in Table 6.

Correlations between PEDRs thought to be related to Agreeableness ranged from $r=-.02$ to $r=.04$ while those designated as unrelated ranged from $r=-.05$ to $r=.08$. The AC dimension of Coach and Develop was the only dimension hypothesized to be related to Openness to Experience and it was moderately related, $r=.14$ while those unrelated to this dimension ranged from $r=.03$ to $r=.16$. PEDRs thought to be related to Extraversion ranged from $r=-.02$ to $r=.12$ to with the unrelated dimensions ranging from $r=-.03$ to $r=.08$. All of the dimensions related to Conscientiousness had negative relationships

ranging from $r=-.12$ to $r=-.07$ while those thought to be unrelated also had negative relationships with Conscientiousness ranging from $r=-.12$ to $r=-.09$. Dimensions thought to be related to cognitive ability had the highest relationships, with correlations as high as $r=.40$ while the unrelated dimensions ranged from $r=.06$ to $r=.30$.

In order to make better sense of this multitude of intercorrelations, mean and medians of the correlations were computed and the average correlations between AC ratings and individual difference variables are reported in Table 7. The results of these analyses indicate that AC dimensions hypothesized to be related to individual difference traits did slightly better on average (grand mean $r=.07$) when compared with dimensions thought to be unrelated (grand mean $r=.06$). This striking similarity in averages between related and unrelated dimensions may be due to certain individual difference variables which were shown to be related in the opposite direction than expected. For example, Conscientiousness was expected to be positively related to performance on relevant AC dimensions and was actually found to be negatively related (median $r=-.10$).

Table 6. Correlations between Assessment Center Dimensions and Individual Difference Measures

Assessment Center Dimension	Individual Difference Variable					
	Agreeableness	Openness	Extraversion	Emotional Stability	Conscientiousness	Cognitive Ability
Analyze Issues	-.05	.11	-.03	.06	-.12*	.40**
Use Sound Judgment	.00	.16*	.07	.02	-.10	.31**
Manage Execution	.01	.12*	.02	.13	-.07	.27**
Lead Courageously	.08	.12*	.12	.03	-.02	.10
Influence Others	-.02	.12	.06	.13	-.09	.30**
Coach and Develop	.03	.14*	.08	.10	-.05	.06
Foster Teamwork	.08	.07	.06	.12	-.02	.25**
Build Relationships	.04	.03	-.02	.08	-.10	.15
Foster Open Communication	.04	.07	-.03	.08	-.03	.24**
Manage Disagreements	.04	.10	.05	.06	-.12*	.10

Note: *= $p < .05$, **= $p < .05$, $N=274$, Correlations that are bolded were hypothesized to be higher than other relationships.

Table 7. Average Correlations between Assessment Center Ratings and Individual Differences

	Corresponding		Non-Corresponding		
	<i>k</i>	Mean <i>r</i>	<i>k</i>	Mean <i>r</i>	Median <i>r</i>
<i>Agreeableness</i>	6	.04	24	.08	.08
Influence Others	-	-.02	4	.12	.13
Foster Teamwork	-	.08	4	.12	.12
Coach and Development	-	.03	4	.06	.08
Building Relationships	-	.04	4	.04	.06
Manage Disagreements	-	.04	4	.04	.06
Foster Open Communication	-	.04	4	.09	.07
<i>Openness to Experience</i>	2	.09	6	.06	.05
Analyze Issues	-	.11	3	-.01	-.03
Foster Teamwork	-	.07	3	.12	.12
<i>Extraversion</i>	6	.05	25	.08	.08
Lead Courageously	-	.12	5	.06	.08
Influence Others	-	.06	4	.12	.13
Foster Teamwork	-	.06	4	.12	.12
Coach and Develop	-	.08	4	.06	.08
Build Relationships	-	-.02	4	.04	.06
Foster Open Communication	-	-.03	4	.09	.07
<i>Emotional Stability</i>	1	.06	4	.03	.08
Manage Disagreements	-	.06	4	.03	.08
<i>Conscientiousness</i>	3	-.10	11	.04	.05
Analyze Issues	-	-.12	3	-.01	-.03
Use Sound Judgment	-	-.10	4	.06	.05
Manage Execution	-	-.07	4	.07	.07
<i>Cognitive Ability</i>	3	.33	11	.04	.05
Analyze Issues	-	.40	3	-.01	-.03
Use Sound Judgment	-	.31	4	.06	.05
Manage Execution	-	.27	4	.07	.07
Total	21	.07	81	.06	.07

Note: Median $r = .04$ for Agreeableness, $r = .14$ for Openness, $r = .08$ for Emotional Stability, $r = .10$ for Conscientiousness, $r = .31$ for Cognitive Ability, $\chi^2 = 2.25$, $p > .05$.

Next, chi-square analyses were performed in order to determine if the differences between PEDRs related to individual difference traits and those unrelated to the traits were statistically significant. Once again, normality was not assumed and the analysis was performed using one degree of freedom. The results suggest that the average AC ratings for corresponding individual difference traits (grand mean $r=.07$) were not significantly different ($\chi^2=2.25, p>.05$) than those AC ratings for non-corresponding individual difference traits (grand mean $r=.06$).

The one exception to this conclusion is cognitive ability. A careful reexamination of Table 6 shows that the averages of the AC ratings thought to be related to this ability (mean $r=.33$) are stronger than those thought to be unrelated (mean $r=.05$). Based on these findings, Hypothesis 7 seems to be partially supported at least regarding cognitive ability.

This first chi-square analysis was computed using the cognitive ability as an individual difference comparison. Since cognitive ability was one of the most robust individual difference predictors, an additional chi-square analysis was calculated on only the personality variables. This was done in order to determine if the cognitive ability scores skewed the analyses in any way. The results of this second chi-square analysis show that the average AC ratings for corresponding individual difference traits ($\chi^2=3.84, p<.05$) were significantly different than those AC ratings for non-corresponding individual difference traits. However, the differences were in the wrong direction and are therefore difficult to interpret.

Confirmatory Factory Analyses

At the outset of this project, confirmatory factor analysis (CFA) was used in order to examine the data. For these analyses, covariance matrices were read into LISREL version 8.5 (Joreskog & Sorbom, 1996) for factor extraction. Unfortunately, this did not produce an admissible solution (several factor correlations were found to be >1.0) and some models failed to converge. Because of these difficulties common in the assessment center literature (Lance et al., 2000), the testing of the hypotheses in this study was based solely on the aforementioned correlational and chi-square analyses.

Removing Exercise Variance

Since it was not possible to run CFA analyses, semi-partial correlations were computed as a way to remove sources of exercise variance. This was an important set of analyses to undertake because of their ability to uncover the amount of variance between AC and MSF ratings that was due to exercise effects. The first step in this process was to examine the relationships between exercise variance and MSF ratings dimensions, individual difference variables, and rater source effects. By doing this, it will be possible to understand how removing exercise variance could improve or harm the content validity of the present AC. Before exercise variance was removed, the relationships between exercise factors and dimensions were generally positive and can be found in Table 8.

Table 8. *Correlations between Exercise Factors, Cognitive Ability, Personality, and Performance Ratings*

External Variable	Exercise Factors		
	In-Basket	Subordinate Roleplay	Task-Force Simulation
<i>Cognitive Ability</i>	.25**	.05	.23**
<i>Personality</i>			
Extraversion	.02	.10	.00
Agreeableness	.00	.08	.01
Conscientiousness	-.02	-.06	-.13*
Emotional Stability	.02	.09	.12*
Openness to Experience	.09	.11	.11
<i>Performance Dimension Ratings</i>			
Analyzes Issues	.11	.01	.07
Uses Sound Judgment	.10	.01	.04
Manages Execution	.12*	.02	.00
Leads Courageously	.05	.05	.04
Influences Others	.08	.08	.04
Fosters Teamwork	.18*	.08	.01
Coaches and Develops	.08	.18*	.00
Builds Relationships	.18*	.18*	-.02
Manages Disagreements	.21**	.21*	-.04
Openness to Communication	.12*	.12*	.03
Overall Performance	.14*	.05	.02
<i>Performance Rating Source Factors</i>			
Supervisor	.14*	.07	.05
Co-worker	.07	.09	.06
Subordinate	.10	-.03	-.06

Note. N=274 *= $p < .05$, **= $p < .01$

The In-Basket exercise, for example, had the strongest relationship with MSF dimensions with its highest correlations with dimensions like Manages Disagreements ($r=.21$), Fosters Teamwork ($r=.18$), and Builds Relationships ($r=.18$) and the lowest correlations between the In-Basket and the dimensions of Leads Courageously ($r=.05$) and Influences Others ($r=.08$). Regarding overall job performance, the In-Basket ($r=.14$) was the only exercise that was significantly related to the composite total of MSF scores.

The Subordinate Roleplay exercise also had significant correlations with MSF performance dimensions like Manages Disagreements ($r=.21$), Builds Relationships

($r=.18$), and Coaches and Develops ($r=.18$) but there were also correlations which were close to zero such as Analyzes Issues ($r=.01$) and Uses Sound Judgment ($r=.01$). The Task-Force Simulation exercise had no significant correlations with ratings on the MSF instrument and some dimensions such as Builds Relationships ($r= -.02$) and Manages Disagreement ($r= -.04$) were even negatively related to exercise performance.

When it came to individual differences, cognitive ability was the most consistently related to exercise performance with correlations of .25 and .23 which were statistically significant. On the Task-Force Simulation, Conscientiousness ($r= -.13$) and Emotional Stability ($r=.12$) were the only other individual difference variables related to exercise performance.

As was mentioned earlier, none of the AC dimensions are highly correlated with overall job performance (Table 5). Only the dimension Lead Courageously ($r=.19$) was significantly related to overall performance. Due to the low correlations between individual AC dimensions and performance, OAR was not significantly correlated with overall performance either ($r=.10$). This correlation between OAR and overall performance is considerably weaker than the correlation of .25 which has been found in a similar study (Gaugler et al., 1987).

These weak relationships were only exacerbated when exercise variance was removed using semi-partial correlations (Table 8). Almost every single correlation between AC dimensions and overall job performance was reduced when exercise variance was removed. For example, the relationship between Influence Others and overall performance was originally positive ($r=.05$) but after exercise variance was

controlled for it was reduced to a negative correlation ($r=-.02$). This reduction was similar across a variety of AC dimensions.

CHAPTER V

DISCUSSION

This study attempted to use external criteria to determine AC validity, in contrast to traditional approaches which use only internal sources such as PEDRs (Bowler & Woehr, 2006; Sackett & Dreher, 1982; Woehr & Arthur, 2003). By examining correlations between AC ratings, job performance, and individual differences, it was possible to get a fuller picture of AC validity. Although the overall predictive validity of the assessment center was relatively low, the hypothesis that corresponding AC and MSF dimensions would be more related than non-corresponding dimensions was generally supported based on the chi-square analyses (Table 4). This method of analysis could still prove to be a promising way to test the construct validity of ACs because it demonstrates that even if the internal correlations of an AC are less than promising from a traditional MTMM perspective, an AC may still contain variance that can be shown to be related to job performance.

In addition to the hypotheses regarding corresponding AC and MSF dimensions, the role of AC dimensions related to specific personality and cognitive ability traits was also examined. In general, the self-reported FFM personality traits did not explain a significant amount of variance in AC performance (Tables 6-7). In fact, the highest average correlation with AC dimensions was actually Conscientiousness which had an unexpected negative correlation ($r=-.10$) with AC performance. Conscientiousness' low correlation with AC performance is in contrast to the significant relationship it had with overall ratings of job performance ($r=.12$).

There is precedent for Conscientiousness being negatively related to performance although it may seem counterintuitive to many (Furnham, Taylor, & Chamorro-Premuzian, 2008; Moutafi, Furnham, & Crump, 2006). This flies in the face of the common assertion amongst scholars of Industrial and Organizational Psychology that Conscientiousness is consistently tied to performance across almost all jobs and contexts (Barrick & Mount, 1991). One possible mechanism that may underlie this negative relationship between Conscientiousness and performance may be the moderating factor of fluid intelligence which refers to a person's ability to process information quickly and often under time pressure (Moutafi, Furnham, & Paltiel, 2004). One theory about this role in performance is that individuals high in Conscientiousness may work harder and more diligently to make up for a deficit in fluid intelligence by attempting to increase their levels of crystallized intelligence which is the ability to accumulate facts and information (Moutafi et al., 2004).

More specifically in relation to assessment centers, conscientiousness has been shown to be negatively related to performance on tasks such as an in-basket exercise (Tett, Steele, & Beauregard, 2003). This relationship may be due more to specific facets of Conscientiousness like methodicalness which may drive an individual high in this trait to focus narrowly on the quality of what they are producing versus the quantity (Tett et al., 2003; Tett & Burnett, 2003). Based on these findings about the importance of specificity in driving trait hypotheses, it would be beneficial for future studies to test more specific links between facets of personality and additional exercise performance (Tett & Burnett, 2003).

The contrasting relationships between the role of individual differences in the AC and the MSF were the most consistent findings of the study. This raises questions about the assessment center's ability to measure individual difference variables that have been consistently shown to be related to job performance in general and leader performance more specifically (Barrick & Mount, 1991; Judge et al., 2002). This AC's difficulty in tapping these personality traits could be an explanation for weak relationship between AC and job performance. It is possible that ACs are not currently designed to sufficiently measure personality trait like behavior that is related to job performance and because of this their predictive validity is reduced.

Conversely, cognitive ability was the dominant individual difference in predicting AC performance but not MSF ratings. These findings may explain the low correlations between personality traits and OAR, since most of the variance in performance can be attributed to cognitive ability. The importance of cognitive ability in AC performance is not surprising given other studies which have also shown cognitive ability to be a consistent predictor of AC performance, often above personality traits (Furnham et al., 2008; Furnham & Chamorro-Premuzic, 2004; Lance et al., 2007; Lance et al., 2000; Lievens et al., 2003; Spector et al., 2000).

On the other hand, cognitive ability was less related to overall job performance ($r=.01$). This is not in line with previous research which has shown its importance in predicting job performance (Schmidt & Hunter, 1998). This may be because the dimensions used to measure job performance in this study were not cognitively loaded and were in fact more related to interpersonal components of workplace behavior.

It is also possible that personality is more predictive of MSF ratings because these represent an example of “typical performance” whereas cognitive ability and AC ratings may be examples of “maximum performance” (Furnham, Taylor, & Chamorro-Premuzic, 2008; Sackett, Fogli, & Zedeck, 1988). The importance of personality in affecting typical performance has been shown in academic settings where cognitive ability is more highly correlated with test scores and personality is more closely related to performance over a semester long course (Furnham & Chamorro-Premuzic, 2004). Regarding AC performance more specifically, some view them as tests of “power” or maximum performance and thus better measures of what a participant can do, not necessarily what a participant will do (Furnham, Taylor, & Chamorro-Premuzic, 2008).

Limitations

There are several plausible reasons why the validity of this AC was remarkably low. First, the MSF ratings used for comparison in this study are not an exhaustive list of the dimensions on which participants were rated; they were simply the ones that overlapped with the AC dimensions. This was done in order to examine the validity of particular AC dimensions; however it is possible that AC dimensions or composite exercise factors were more related to MSF dimensions not examined in this study.

On the other hand, the vast amount of dimensions each individual was rated on in the MSF may have created rater fatigue. Variance of MSF ratings has been shown to be comprised mainly of rater effects and less related to actual job performance (Scullen, Mount, & Goff, 2000). This again calls into question the slicing and dicing of performance into seemingly endless amounts of dimensions when they may do little to explain job performance beyond a general factor (Hoffman & Woehr, 2009; Scullen et

al., 2000). Indeed the MSF dimensions ratings used in this particular study were highly correlated with one another.

Another area that could be revisited is the process through which the linking between personality traits and AC dimensions were performed. It is possible that other researchers would have linked the traits and dimensions differently and this would have affected the tests of the hypotheses. In future research, a process similar to what Haaland and Christiansen (2002) used in their linking of personality traits to exercises high in that trait's activation potential could be effective.

In conclusion, the present study was an attempt at a fresh approach to AC construct validity that was less concerned with conforming PEDRs to MTMM standards. The AC OAR had lower than expected predictive validity of overall job performance. However, the exercise scores were somewhat predictive of performance, especially the In-Basket and the Subordinate Roleplay. This was in contrast to the low validity of the AC dimensions when exercise variance was removed and suggests that the exercises were able to gauge a unique source of variance that was untapped by the AC dimensions.

These results, although less than ideal in the strength of their relationships with job performance do provide some support for using external criteria in AC validation. If only PEDRs were examined for their adherence to MTMM guidelines, the validity may have appeared to be even lower than it was. By examining external correlates of exercise factors as well dimensions, it is possible to get a fuller picture of the individual AC and areas that may need further study and development.

For example, the correlations between the exercises and job performance dimensions do indicate that the present AC dimensions could be failing to measure

important components of job performance. This could lead to revision of the AC dimensions or inclusion of other dimensions or exercises. Likewise, because personality was generally more related to job performance it is possible the AC is not adequately measuring personality related competencies that are driving job performance. Based on this information it is possible that revising the AC to tap these personality related competencies would increase its validity overall.

REFERENCES

- 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.
- Atkins, P.W. & Wood, R.E. (2002). Self-versus others' ratings as predictors of assessment center ratings: Validation evidence for 360-degree feedback programs. *Personnel Psychology*, *55*, 871-904.
- Barrick, M. & Mount, M.K. (1991). The big five personality dimensions and job performance: A meta-analysis. *Personnel Psychology*, *44*, 1-27.
- Barrick, M., Stewart, G.L., Neubert, M.J., & Mount, M.K. (1998). Relating member ability and personality to work-team processes and team effectiveness. *Journal of Applied Psychology*, *83*, 377-391.
- Beehr, T.A., Ivanitskaya, L., Hansen, C.P., Erofeev, D., & Gudanowski, D.M. (2001). Evaluation of 360 degree feedback ratings: Relationships with each other and with performance and selection predictors. *Journal of Organizational Behavior*, *22*, 775-788.
- Bowler, M.C. & Woehner, 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.
- Bray, D.W. (1982). The assessment center and the study of lives. *American Psychologist*, *37*, 180-189.
- Bray, D.W. & Grant, D.L. (1966). The assessment center in the measurement of potential for business management. *Psychological Monographs*, *80* (whole no.625).
- Bray, D.W., Campbell, R.J., & Grant, D.L. (1974). *Formative years in business: A long-term AT&T study of managerial lives*. New York: Wiley.
- Byham, W.C. (1970). Assessment center for spotting future managers. *Harvard Business Review*, *48*, 150-160.
- Campbell, D.T. & Fiske, D.W. (1959). Convergent and discriminant validation by multitrait-multimethod matrix. *Psychological Bulletin*, *56*, 81-105.
- Christiansen, N.D. (2008). *Further consideration of the use of narrow trait scales*. 23rd Annual Conference of the Society for Industrial and Organizational Psychology, San Francisco, CA.

- Church, A.H. & Allen, D.W. (1997). Advancing the state of the art of 360-degree feedback. *Group and Organization Management*, 22, 149-161.
- Collins, J.M., Schmidt, F.L., Sanchez-Ku, M., Thomas, L., McDaniel, M.A., & Le, H. (2003). Can basic individual differences shed lights on the construct meaning of assessment center evaluations? *International Journal of Selection and Assessment*, 11, 17-29.
- Coulton, G.F. & Field, H.S. (1995). Using assessment center centers in selecting entry-level police officers: Extravagance or justified expense? *Public Personnel Management*, 24, 223-254.
- Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52, 281–302.
- Fiske, D.W. (1987). Construct invalidity comes from method effects. *Educational and Psychological Measurement*, 14, 285-307.
- Furnham, A. & Chamorro-Premuzic, T. (2004). Personality and intelligence as predictors of statistics examination grades. *Personality and Individual Differences*, 37, 943-955.
- Furnham, A., Taylor, J. & Chamorro-Premuzic, T. (2008). Personality and intelligence correlates of assessment center exercises. *Individual Differences Research*, 6, 181-192.
- 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.
- Goffin, R.D., Rothstein, M.G., & Johnston, N. (1996). Personality testing and the assessment center: Incremental validity for managerial selection. *Journal of Applied Psychology*, 81, 746-756.
- 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.
- Hagan, C.M., Konopaske, R., Bernardin, H.J., & Tyler, C.L. (2006). Predicting assessment center performance with 360-degree, top-down, and customer-based competency assessments. *Human Resource Management*, 45, 357-390.
- Hermelin, E., Lievens, F., & Robertson, I. (2007). The validity of assessment centers for the prediction of supervisory performance ratings: A meta-analysis. *International Journal of Selection and Assessment*, 15, 405-411.

- Hoffman, B.J. & Woehr, D.J. (2009). Disentangling the meaning of multisource performance rating source and dimension factors. *Personnel Psychology*, 62, 735-765.
- Howard, A. (1997). A reassessment of assessment centers: Challenges for the 21st century. *Journal of Social Behavioral and Personality*, 12, 13-52.
- Huck, J.R. & Bray, D.W. (1976). Management assessment center evaluations and subsequent job performance of Black and White females. *Personnel Psychology*, 29, 13-30.
- Hunter, J.E. & Hunter, R.F. (1984). Validity and utility of alternative predictors of job performance. *Psychological Bulletin*, 96, 72-98.
- Jackson, D.J.R., Stillman, J.A., & Atkins, S.G. (2005). Rating tasks versus dimensions in assessment centers: A psychometric comparison. *Human Performance*, 18, 213-241.
- Jones, R.G. & Born, M.P. (2008). Assessor constructs in use as the missing component in validation of assessment center dimensions: A critique and directions for research. *International Journal of Selection and Assessment*, 16, 229-238.
- Joreskog, K.G. & Sorbom, D. (1996). LISREL 8.5: User's reference guide. Chicago: Scientific Software.
- Judge, T.A., Bono, J.E., Ilies, R., & Gerhardt, M.W. (2002). Personality and Leadership: A qualitative and quantitative review. *Journal of Applied Psychology*, 87, 765-780.
- Kauffman, J.R., Jex, S.M., Love, K.G., & Libkuman, T.M. (1993). The construct validity of assessment centre performance dimensions. *International Journal of Selection and Assessment*, 1, 213-223.
- Kirkpatrick, S.A. & Locke, E.A. (1991). Leadership: Do traits matter? *Academy of Management Executive*, 5, 48-59.
- Klimoski, R. & Brickner, M. (1987). Why do assessment centers work? The puzzle of assessment center validity. *Personnel Psychology*, 40, 243-260.
- Lance, C. E. (2008). Why assessment centers do not work the way they're supposed to. *Industrial and Organizational Psychology*, 1, 84-97.
- 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.R., & Smith, D.E. (2000). Assessment center exercise factors represent cross-situational specificity, not method bias. *Human Performance, 13*, 323-353.
- Landy, F. J. (1986). Stamp collecting versus science: Validation as hypothesis testing. *American Psychologist, 41*, 1183-1192.
- 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., 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 discrimination validity. *Journal of Applied Psychology, 91*, 247-258.
- Lievens, F., Harris, M.M., Van Keer, E., & Bisqueret, C. (2003). Predicting cross-cultural training performance: The validity of personality, cognitive ability, and dimensions measured by an assessment center and a behavior description interview. *Journal of Applied Psychology, 88*, 476-489.
- London, M. & Simther, J.W. (1995). Can multi-source feedback change perceptions of goal accomplishment, self-evaluations, and performance related outcomes? *Personnel Psychology, 48*, 803-839.
- McCrae, R. & Costa, P.T. (1987). Validation of the five-factor model of personality across instruments and observers. *Journal of Personality and Social Psychology, 52*, 81-90.
- 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.
- Moutafi, J., Furnham, A., & Crum, J. (2006). What facets of openness and conscientiousness predict fluid intelligence score? *Learning and Individual Differences, 16*, 31-42.
- Moutafi, J., Furnham, A., & Paltiel, L. (2004). Why is conscientiousness negatively correlated with intelligence? *Personality and Individual Differences, 37*, 1013-1022.
- Murray, H.A. (1938). *Explorations in personality*. New York: Oxford University Press.
- Pervin, L.A. (1989). Persons, situations, interactions: The history of a controversy and a discussion of theoretical models. *Academy of Management Review, 14*, 350-360.

- Podsakoff, P.M., MacKenzie, S.B., Lee, J.Y., & Podsakoff, N.P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology, 88*, 879-903.
- 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. & Harris, M.M. (1988). A further examination of the constructs underlying assessment center ratings. *Journal of Business and Psychology, 3*, 214-229.
- Sackett, P.R., Fogli, L., & Zedeck, S. (1988). Relations between measures of typical and maximum job performance. *Journal of Applied Psychology, 73*, 482-486.
- Schmidt, F.L. & Hunter, F.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.
- Schmit, M.J., Kihm, J.A., & Robie, C. (2000). Development of a Global Measure of Personality. *Personnel Psychology, 53*, 153-193.
- Scullen, S.E., Goff, M.E., & Mount, M.J. (2000). Understanding the latent structure of job performance ratings. *Journal of Applied Psychology, 85*, 956-970.
- Spector, P.E., Schneider, J.R., Vance, C.A., & Hezlett, S.A. (2000). The relations of cognitive ability and personality traits to assessment center performance. *Journal of Applied Social Psychology, 30*, 1474-1491.
- 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.
- Tett, R.P. & Burnett, D.D. (2003). A personality trait-based interactionist model of job performance. *Journal of Applied Psychology, 88*, 500-517.
- Tett, R.P., Steele, J.R., & Beauregard, R.S. (2003). Broad and narrow measures on both sides of the personality-job performance relationship. *Journal of Organizational Behavior, 24*, 335-356.
- Thornton, G.C. & Rupp, D.E. (2006). *Assessment centers in human resource management: Strategies for prediction, diagnosis, and development*. New Jersey: Lawrence Erlbaum Associates.
- Tillema, H.H. (1998). Assessment of potential, from assessment centers to development centers. *Assessment of Potential, 6*, 185-191.

- Watson, G. & Glaser, E.M. (1980). *Watson – Glaser Critical Thinking Appraisal Manual*. New York, NY: The Psychological Corporation.
- Welkowitz, J., Ewen, R.B., & Cohen, J. (1982). *Introductory statistics for the behavioral sciences*. Orlando: Academic Press, Inc.
- Wesman, A.G. (1965). *The Wesman personnel classification test* (manual). New York: The Psychological Corporation.
- Woehr, D.J. & Arthur, W. (2003). The construct-related validity of assessment center ratings: A review and meta-analysis of the role in methodological factors. *Journal of Management*, 29, 231-258.
- Wright, J.C. & Mischel, W. (1988). Conditional hedges and the intuitive psychology of traits. *Journal of Personality and Social Psychology*, 55, 454-469.