THE DIMENSIONS WITHIN EXERCISES APPROACH: A NOMOLOGICAL NETWORK ANALYSIS OF ASSESSMENT CENTER VALIDITY

by

ALEXANDER CHRISTOPHER LOPILATO

(Under the direction of Brian J. Hoffman)

ABSTRACT

The present study extends earlier research that has examined the extent to which performance on assessment center (AC) dimensions can be differentiated by exercise. That is, previous research has found that distinct AC dimensions can be modeled within each AC exercise, supporting the notion that differences in AC performance across exercises do not reflect measurement bias. However, the relationships between the dimension within exercise factors and theoretically relevant external variables are still unknown. Thus, this paper extends AC research by examining the nomological network of the dimension within exercise ratings. Specifically, this paper compares how the relationships between dimensions within exercise ratings and both personality and mental ability variables (a) differ by the dimension and exercise combination being examined and (b) differ from more commonly used AC rating composites: across exercise dimension ratings, across dimension exercise rating, and the overall assessment rating (OAR).

INDEX WORDS: assessment centers, construct-related validity, nomological network

THE DIMENSIONS WITHIN EXERCISES APPROACH: A NOMOLOGICAL NETWORK ANALYSIS OF ASSESSMENT CENTER VALIDITY

by

ALEXANDER CHRISTOPHER LOPILATO

B.S. University of Georgia, 2010

A Thesis Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

MASTER OF SCIENCE

ATHENS, GEORGIA

2014

THE DIMENSIONS WITHIN EXERCISES APPROACH: A NOMOLOGICAL NETWORK ANALYSIS OF ASSESSMENT CENTER VALIDITY

by

ALEXANDER CHRISTOPHER LOPILATO

Major Professor:	Brian J. Hoffman
Committee:	Nathan T. Carter
	W. Keith Campbell

Electronic Version Approved:

Julie Coffield
Interim Dean of the Graduate School
The University of Georgia
December 2014

TABLE OF CONTENTS

LIST OF TABLES

Page

CHAPTER 1

INTRODUCTION

Assessment centers (ACs) present test takers with a series of moderate–to–high fidelity simulations of the tasks that are important to performance on a focal job. These tasks are designed to provide insight into how the test taker will behave on the job by eliciting behaviors pertinent to skills needed for effective performance on the job (Thornton & Mueller-Hanson, 2004). Consistent with this expectation, numerous studies support the criterion-related validity of ACs for the prediction of managerial performance (Arthur, Day, McNelly, & Edens, 2003; Gaugler, Rosenthal, Thornton, & Bentson, 1987; Meriac, Hoffman, Woehr, & Fleisher, 2008).

Despite this supportive evidence, the mechanisms that account for the criterion-related validity of ACs are as of yet unclear. Research and practice have traditionally focused on the dimensions as the focal unit of design, scoring, and interpretation. Yet, findings that exercises explain more variance than focal dimensions (Lance, Lambert, Gewin, Lievens, & Conway, 2004; Sackett and Dreher, 1982) have cast doubt on dimension-based interpretations of ACs (Jackson, Stillman, & Englert, 2010). More recently, an alternative interpretation of ACs proposes that exercises should be the focal unit of analysis (Lance, Newbolt, Gatewood, Foster, French, & Smith, 2000). This interpretation proposes that the tendency of AC ratings to conform to a structure of multiple exercises rather than multiple dimensions should not be interpreted as evidence that ACs are rife with systematic method bias, but instead that exercise effects reflect situational specificity in assessee performance in different simulation exercises (Jackson et al.,

2010; Lance et al., 2000). The last decade has witnessed a lively debate on task-based and dimension-based interpretations of ACs (For a review see Jackson, Lance, & Hoffman, 2012).

More recently, Hoffman (2012) argued that considering dimensions or tasks alone oversimplifies the nature of information that can be obtained from ACs. He proposes multiple sources of information can be gleaned from ACs, including dimensional performance, exercise performance, general performance and dimension-specific exercise performance. Moreover, recent research supports these assertions, with evidence for the criterion-related validity and nomological network of dimensions, exercises, and general performance factors (Jackson et al., 2010; Hoffman, Monahan, Kennedy, LoPilato, Sutton, Lance, Under Review; Meriac et al., 2008). Yet, limited systematic attention has been devoted to understanding the final source of information proposed by the multifaceted perspective on AC performance: dimension-specific performance within exercises. Dimension-specific performance within exercises represents the variance of a given candidate's performance on a specific dimension within a specific exercise (Hoffman & Meade, 2012). The paucity of research is a key gap, as Hoffman and Meade (2012) showed that an empirical structure specifying dimensions within exercises (DWE) provided an acceptable fit to AC data. Furthermore, Putka and Hoffman (2013) showed that this aspect of performance was the second largest source of variance present in AC ratings, and practitioners sometimes consider DWE ratings. Although these recent studies have documented the existence of this source of variance, the substantive underpinnings of DWE are as of yet unexplored.

The present study draws from interactionist perspectives of behavior (Lievens, Chasteen, Day, & Christiansen, 2006) to provide an empirical foundation for the meaning of the DWE scoring approach proposed by recent research exploring multifaceted interpretation of AC performance (Hoffman, 2012; Hoffman & Meade, 2012; Putka & Hoffman, 2013). Specifically,

we propose that historically weak levels of convergence between individual differences and AC dimensions (Meriac et al., 2008) and exercises (Hoffman et al., Under Review; Jackson et al., 2010) stem from the failure of past research to consider dimension-specific exercise performance. To test this proposition, we first attempt to replicate Hoffman and Meade's (2012) results by supporting distinguishable dimensions within exercise in multiple samples of AC ratings. Then we compare the nomological network of individual differences related to dimensions and exercises to that of DWEs in two AC samples.

Past Internal Structure Research

In the thirty years since Sackett and Dreher's (1982) application of the multitrait-multimethod (MTMM) method to dimension rated following each exercise (post exercise dimension ratings; PEDRs), numerous increasingly sophisticated statistical models have been applied to the analysis of PEDRs (Woehr, Meriac, & Bowler, 2012). Specifically, confirmatory factor analytic (CFA) models have been used to model both AC dimension and exercise factors (Lance, Noble, & Scullen, 2002; Lievens & Conway, 2001; Monahan, Hoffman, Lance, Jackson, & Foster, 2014; Woehr et al., 2012). Using these models AC dimension factors have been compared to trait factors and AC exercise factors have been compared to method factors (Woehr et al., 2012). With its focus on the MTMM approach to construct validation, research using CFA models has tended to interpret dimensions as the only source of true score variance in AC ratings and exercises as a source of performance irrelevant bias (Lance, Baranik, Lau, & Scharlau, 2009).

However, this assumption has been the subject of increasing criticism (Lance, 2008; Lance, 2012). Recent studies have shown that exercise variance does not necessarily reflect assessor bias (Putka & Hoffman, 2013) and that exercise factors correlate meaningfully with

criterion variables (Lance et al., 2000; Hoffman, Melchers, Blair, Kleinmann, & Ladd, 2011). That is, research has demonstrated the criterion-related validity of exercise factors (Lance et al., 2000; Lance, Foster, Gentry, & Thoresen, 2004; Hoffman et al., 2011). Specifically, research has found that exercise performance is moderately correlated with an assessee's job knowledge (Lance, Foster, Nemeth, Gentry, Drollinger, 2007), supervisor ratings of job performance (Hoffman et al., Under Review; Lance et al., 2007) and weakly to moderately correlated with personality and cognitive ability (Jackson et al., 2010; Hoffman, et al., Under Review). However, consistent with the MTMM model, this research has exclusively examined exercise performance on the basis of a single score that summarizes information within the exercises.

Despite the increasing support of considering exercise variance when scoring and interpreting AC performance, the use of exercise scores continues to be criticized because performance in a simulation exercise is conceptually ambiguous (Arthur, Day, & Woehr, 2008; Rupp, Thornton, & Gibbons, 2008). In addition, although some evidence has simultaneously examined the construct validity of both dimensions and exercises, these units have been examined separately, with little consideration of their interplay as a meaningful component of assessee performance.

The Interactionist Perspective

Although an implicit assumption of the MTMM approach is that only dimension variance is meaningful, this interpretation is inconsistent with the guiding rationale of using multiple simulation exercises (Howard, 2008). That is, AC exercises are often explicitly designed so that different exercises elicit different performance behaviors on the same AC dimension (Hoffman, 2012; Howard, 2008). For instance, when assessing one's proficiency on an interpersonal dimension such as leadership, the behaviors elicited by a leaderless group discussion exercise

and a role play exercise could be very different, but both valid behavioral realizations of one's proficiency as a leader. Thus, consistent with the design of ACs, it is common practice to report "split" ratings in developmental feedback, which are dimension ratings that are separated by the exercise they were measured by. From this research has begun to test whether dimensional performance that is distinguishable within an exercise but differs across exercises is a meaningful unit of analysis in AC ratings (Putka & Hoffman, 2013; Hoffman & Meade, 2012).

Theoretically, the DWE interpretation of AC performance most closely follows the interactionist perspective that has been used to understand performance in ACs (Lievens et al., 2006; Lievens & Christiansen, 2012). Specifically, it has been argued that AC research should move away from determining ways to fix ACs (e.g. improved rater training and AC design) and towards a theory as to why assessees' performance differs by exercise (Haaland & Christiansen, 2002; Lievens et al., 2006). To explain these differences research has turned to Trait Activation Theory (TAT). TAT theorizes that behavior is a function of the person-situation interaction in which behavior is a response to trait-relevant cues provided by the situation (Lievens et al., 2006; Lievens & Christiansen, 2012; Tett & Burnett, 2003; Tett & Guterman, 2000).

Consistent with TAT, evidence suggests that weak convergence across exercises may be due to the fact that exercises vary in their ability to elicit similar performance behaviors from the same underlying dimension (Lievens et al., 2006). Specifically, exercises provide trait cues that inform assessees as to what behavior they should display. Even when measuring the same dimension, however, different exercises provide different information regarding the appropriate trait-relevant behavior (Haaland & Christiansen, 2002; Lievens et al., 2006; Tett & Burnett, 2003). Accordingly, behavior on the same dimension will vary across exercises as "trait activation theory predicts divergent results between methods which are dissimilar in trait

activation potential [TAP]" (Lievens, De Koster, & Schollaert, 2008, p. 231). Thus, compared to the overall dimension and overall exercise ratings, interpreting DWE ratings is more consistent with the interactionist perspective because it investigates variance in performance due to the person-situation interaction (Hoffman, 2012).

To date, only two studies have examined DWE ratings (Hoffman & Meade, 2012; Putka & Hoffman, 2013). Both studies examined the internal structure of the ratings. First, Hoffman and Meade (2012) used CFA to compare competing models of dimensions within exercises. Specifically, they classified six narrow dimensions that were measured in three exercises into three broader dimensions: (a) Communication and Interpersonal Facilitation, (b) Leadership and Supervision, and (c) Technical Activities and Mechanics of Management. Using CFA, they compared multiple models that differed by the number of dimension factors within each exercise (e.g. three dimensions within each exercise vs. one dimension within each exercise).They found that the best fitting model was one that modeled three broad performance dimensions within each exercise (Hoffman & Meade, 2012). Their findings indicated that performance within an exercise is multifaceted.

Next, Putka and Hoffman (2013) used linear mixed models to isolate 15 sources of variance in AC ratings. In accordance with previous AC research, they found both a large exercise effect and a weak dimension effect (Putka & Hoffman, 2013). That is, they determined that of the total observed variance in AC ratings 42.8%-52% was due to how assessees' performance varied across exercises and 0.5%-1.8% was due to how assessees' performance varied across dimensions. Of importance to this study, they also found that across all three samples the second largest source of variance in AC ratings was due to the variability of an

assessee's performance across different dimension-exercise combinations, which ranged from 28.6%-31.3%.

Together these two studies demonstrate that it is possible to model broad AC dimension factors within an exercise and that this source of variance largely contributes to the overall variance in AC ratings. However, previous research has not examined the nomological network of DWE ratings. That is, by examining the relationships between the DWE ratings and theoretically relevant external variables, or the nomological network, we provide a greater understanding of the construct-related validity of DWE factors (Cronbach & Meehl, 1955). Thus, in the present study, our focus will be on the nomological network of the DWE ratings.

The Construct-Related Validity of DWE Ratings

Previous research has found weak support for the nomological networks of both AC dimensions and exercises (Hoffman et al., Under Review; Jackson et al., 2010; Meriac et al., Under Review; Meriac et al., 2008). One explanation for these findings is that, past studies have restricted their focus to the dimension-level or occasionally, the exercise-level. Specifically, this research has collapsed across all exercises in forming overall dimensions or, when examining exercises, collapsed across all dimensions to form an overall exercise score.

In contrast to previous research, this study examines whether stronger support for the nomological network of AC ratings is provided when both dimensions and exercises are taken into account. To do so, we will compare the correlations that DWE ratings demonstrate with the Five Factor Model (FFM) of personality (Emotional stability, Agreeableness, Conscientiousness, Extraversion, and Openness to experience) and general mental ability (GMA) to the correlations that across exercise dimension ratings, across dimension exercise ratings, and overall assessment ratings (OARs) have with the FFM and GMA. However, before we make our hypotheses

regarding the nomological network of the DWE ratings we will first review the nomological networks of both AC dimensions and exercises.

AC dimension ratings and external correlates. Research on the nomological network of AC dimensions has established two findings: (a) the large quantity of narrow dimensions used by primary studies can be condensed into a smaller, more meaningful set of broader dimensions and (b) these broader dimensions exhibit weak to moderate relationships with conceptually similar traits (Arthur et al., 2003; Meriac et al., 2008; Meriac et al., Under Review). Specifically, there have been two large scale quantitative reviews that have examined the nomological networks associated with each AC dimension.

The first review (Meriac et al., 2008) examined how the seven AC dimensions identified by Arthur et al.'s (2003) dimension framework: (a) Consideration/Awareness of others, (b) Communication, (c) Drive, (d) Influencing others, (e) Organizing and planning, (f) Problem solving, and (g) Stress tolerance related to the FFM and GMA. They found corrected correlations ranging in absolute magnitude from .01 to .29, for the FFM and .22 to .36, for GMA (Meriac et al., 2008). However, a clear pattern of correlations between the AC dimensions and the FFM did not emerge for every AC dimension. Rather, some dimensions correlated strongly with conceptually similar traits, whereas others did not. For example, both drive and influencing others correlated strongly with extraversion (ρ = .29 & .21, respectively), but organizing and planning correlated weakly with conscientiousness (ρ = .07). Regarding GMA, it correlated more strongly with cognitive dimensions such as problem solving (ρ = .34) than it did with more interpersonal dimensions such as consideration/awareness for others (ρ = .22).

A second, more recent review, was conducted with the intention of providing a stronger theoretical framework for conceptualizing AC dimensions and their relationships with the FFM

and GMA (Meriac et al., Under Review). Using CFA, this review first created several different AC dimension taxonomies that posited fewer, broader dimensions than Arthur et al.'s (2003) taxonomy (Meriac et al., Under Review). These updated taxonomies were based upon previous performance behavior taxonomies that typically consist of two (Organ, 1988) to three (Borman & Brush, 1993) dimensions. They used CFA to establish that a three dimension taxonomy consisting of a relational skills dimension, a drive dimension, and an administrative dimension (Meriac et al., Under Review).

With a taxonomy established, they used socioanalytic theory (SAT; Hogan, 2013; Hogan & Holland, 2003), which is a theoretical framework that describes behavior as being motivated by the need to get along and the need to get ahead, to derive the nomological networks for each of the three AC dimensions. Specifically, they predicted that the relational skills dimension would be more related to emotional stability, agreeableness, and conscientiousness; the drive dimension would more related to extraversion and openness to experience; the administrative skills dimension would be more related to GMA. However, similar to the first review, they only found support for a subset of the predicted relationships (Drive & Extraversion, $\rho = .25$; Administrative skills & GMA, $\rho = .30$). The rationale for their predictions will be provided in a following section.

AC exercise ratings and external correlates. Regarding research investigating the empirical relationships between AC exercises and external correlates, similar to research on AC dimensions, there exists only equivocal support (Christiansen, Hoffman, Lievens, & Speer, 2013). Continuing the similarity between the two streams of research there have been two large scale reviews of these relationships. The first review characterized the TAP of six commonly used AC exercises: In-basket (IB), Competitive Leaderless group discussion (LGD), Cooperative

LGD, Role-play (RP), Case analysis (CA), and Oral presentation (OP) (Lievens et al., 2006). Specifically, subject matter experts (SMEs) rated the extent to which each AC exercise provided the opportunity for an assessor to observe a range of trait-relevant behaviors (Lievens et al., 2006). The higher an exercise's TAP rating, the more likely it is to provide an opportunity for a variety of trait-relevant behavior (Christiansen, et al., 2013). While this review provided an initial framework for how exercises should relate to personality traits, more empirical work is needed to support their framework.

More recently, Hoffman et al. (Under Review) conducted a meta-analysis in which they summarized the relationships between five AC exercises: IB, LGD (collapsed competitive and cooperative), RP, CA, and OP and the FFM and GMA. Although the correlations they found among the different exercises and FFM traits and GMA were mostly weak, they did find evidence that exercises exhibit differential relationships with external correlates while also providing some support the TAP ratings found in Lievens et al. (2006). Specifically, they found moderate, positive correlations between the more technical exercises (IB and CA) and GMA (Hoffman et al., Under Review). And, in support of Lievens et al. (2006), they found that performance on the LGD correlated the strongest with emotional stability ($\rho = .09$), extraversion ($\rho = .16$), and openness to experience ($\rho = .08$); performance on the IB exercise correlated the strongest with conscientiousness ($\rho = .15$).

DWE ratings and external correlates. Thus, based upon previous research that has shown that (a) AC dimensions and exercises differentially relate to both the FFM traits and GMA (Hoffman, et al., Under Review; Meriac et al., 2008; Meriac et al., Under Review) and (b) exercises differ in their ability to elicit trait-relevant behaviors (Lievens et al., 2006) we expect that certain DWE combinations will yield higher correlations with certain FFM traits and GMA

than will other DWE combinations. That is, we do not expect the DWE ratings to have uniformly higher correlations with every FFM trait and GMA when compared to the correlations between the different aggregation levels of AC ratings and the FFM and GMA; nor do we expect the relationships between DWE ratings and the external correlates expected to be similar across all of the DWE combinations. To investigate these claims we utilized the framework provided by Meriac et al. (Under Review) to categorize AC dimensions into broader factors.

Relational skills dimension. The relational skills dimension encompasses interpersonal behaviors that have a strong pro-social or communal orientation (e.g., organizational citizenship; Meriac et al., Under Review). Research has argued that the relational skills dimension is related to three of the FFM traits: emotional stability, agreeableness, and conscientiousness (Meriac et al., Under Review) as these traits are thought to be indicative of getting along or communal behaviors (Blickle et al., 2009; Chiaburu et al., 2011; DeYoung, 2006; Digman, 1997; Hogan & Holland, 2003). While previous empirical investigations have found weak relationships between the relational skills dimension and the relevant FFM traits (Meriac et al., Under Review) these investigations have not examined the relationships between the relational skills dimension within certain exercises and the relevant FFM traits. That is, exercises that provide trait-cues for emotional stability, agreeableness, or conscientiousness are likely to provide better measures of the relational skills dimension. Specifically, Lievens et al. (2006) determined that the LGD, OP, RP, and IB exercises provide strong trait-cues for emotional, stability, agreeableness, and conscientiousness. However, Hoffman et al. (Under Review) found mixed support for the relationships among those exercises and FFM traits. Emotional stability was found to be only significantly related to the LGD exercise; agreeableness was not significantly related to performance on any of the exercises and only displayed a positive correlation with RP;

conscientiousness was significantly and positively related to LGD, OP, and IB exercises, with its strongest correlation occurring with IB. Following, we hypothesize that:

Hypothesis 1a: Emotional stability will be more strongly related to performance on the relational-skills dimension when it is measured by the LGD exercise relative to the other dimension-exercise combinations, dimension-level ratings, exercise-level ratings, and OARs.

Hypothesis 1b: Agreeableness will be more strongly related to performance on the relational-skills dimension when it is measured by the RP exercise relative to the other dimension-exercise combinations, dimension-level ratings, exercise-level ratings, and OARs.

Hypothesis 1c: Conscientiousness will be more strongly related to performance on the relational-skills dimension when it is measured by the LGD, OP, and IB exercises relative to the other dimension-exercise combinations, dimension-level ratings, exercise-level ratings, and OARs.

Drive dimension. The drive dimension encompasses interpersonal behavior that "produces results and advances an individual within the group and the group within its competition" (Hogan & Holland, 2003, p. 103). Following this, research has argued that the drive dimension is related to two of the FFM traits: extraversion and openness (Meriac et al., Under Review) as these traits are thought to be indicative of status striving, agentic behaviors (Blickle et al., 2009; Chiaburu et al., 2011; DeYoung, 2006; Digman, 1997; Hogan, 2013; Hogan & Holland, 2003). Similar to the first hypothesis, it is likely that we can gain a better measure of the drive dimension by looking at ratings of drive within exercises that activate either extraversion or openness. Two AC exercises have been determined to activate these traits: LGD

and RP (Hoffman et al., Under Review; Lievens et al., 2006). Moreover, Hoffman et al. (Under Review) found that extraversion was significantly and positively correlated with the LGD, RP, and OP exercises and that openness to experience significantly and positively correlated with the LGD, RP, OP, and CA exercises. Thus, we hypothesize that:

Hypothesis 2a: Extraversion will be more strongly related to performance on the drive dimension when it is measured by the LGD, RP, and OP exercises relative to the other dimension-exercise combinations, dimension-level ratings, exercise-level ratings, and OARs.

Hypothesis 2b: Openness to experience will be more strongly related to performance on the drive dimension when it is measured by the LGD, RP, OP, and CA exercises relative to the other dimension-exercise combinations, dimension-level ratings, exercise-level ratings, and OARs.

Administrative skills dimension. The administrative skills dimension encompasses technical behaviors that closely resemble task performance behaviors that are explicitly required by the job. Given that these behaviors are not interpersonally oriented SAT does not provide clear guidance on its nomological network. However, the broader performance literature has shown that task performance behaviors moderately to strongly relate to GMA (Borman & Motowidlo, 1997; Hoffman & Woehr, 2009). In addition, cognitively loaded dimensions such as the administrative skills dimension have been found to be more related to GMA (Meriac et al., 2008; Meriac et al., Under Review). Moreover, GMA is relevant to task-based exercises that require assessees to problem solve and make administrative decisions such as the IB and CA exercises (Hoffman et al., Under Review). Because of this the IB and CA exercises will provide a better measure of the administrative skills dimension. Accordingly, we hypothesize that:

Hypothesis 3: GMA will be more strongly related to performance on the Administrative dimension when it is measured by either the IB or CA exercises relative to the other dimension-exercise combinations, dimension-level ratings, exercise-level ratings, and OARs.

CHAPTER 2

METHOD

To investigate our hypotheses we used two independent AC samples consisting of participants from different careers.

Participants

Sample 1. Data were collected for 204 participants across 31 organizations in the United States. The mean age of the participants was 41.11 years (73.04% male). The majority of the sample (83%) was Caucasian. All participants were midlevel managers from various industries (e.g. banking, engineering, & manufacturing).

Sample 2. Data were collected for 214 participants from an organization that specializes in postal, insurance, credit, banking, and administrative services. The mean age of the participants was 45.53 years (53.74% male). The majority of the sample (72.90%) was Caucasian. All participants were managers.

Measures and Procedure

Sample 1. The AC measured 16 dimensions, all of which are listed in Appendix A. Moreover, it consisted of four exercises: a) IB, b) LGD, c) RP, and d) a Behavioral Interview (BI). Participants were rated on the AC dimensions after the completion of each exercise. Assessors had at least a master's degree in Industrial-Organizational Psychology and previous rating experience. Before combining ratings to form DWEs, across exercise dimension ratings, across dimension exercise ratings, and an OAR, dimensions were first categorized into broader dimensions according to the framework provided by Meriac et al. (Under Review; see Appendix

A). Once the AC dimensions were recoded DWE ratings were created by averaging together ratings of the same dimension within an exercise. That is, because the narrower dimensions were collapsed into broader dimensions some exercises measured the same dimensions more than once. Next, to create the across exercise dimension ratings we averaged dimension ratings across exercises. For example, if both the LGD and RP exercises measured drive an across exercise rating of drive was created by averaging together the drive rating from the LGD and the drive rating from the RP. Further, across dimension exercise ratings were created by averaging dimension ratings within an exercise. That is, if the RP measured both the drive and relational skills dimensions then an average RP rating was created by averaging together both the drive and relational skills ratings. Finally, an OAR was created for each participant by averaging together all of their dimension ratings across every exercise.

In addition to completing the AC, participants also completed the Global Personality Inventory (GPI; Schmit, Kihm, & Robie, 2000) and the Watson Glaser Critical Thinking Appraisal (Watson & Glaser, 1980). Using the GPI, we created scale level measures of the FFM following the steps outlined by Schmit et al. (2000).

Sample 2. The AC measured 15 dimensions, all of which are listed in Appendix A. Additionally, the participants completed four exercises: a) a group discussion and oral presentation on managing new staff, b) a group discussion and oral presentation based on selecting new staff, c) a case analysis that involved identifying problems in shop layouts in pictures and d) a role play that involved a coaching discussion with an employee. Assessors rated the participants' performance on the AC dimensions after the completion of each exercise. The assessors consisted of higher-level managers (n = 19) and psychologists (n =4). All of which went through a two day training period.

Identical to the Sample 1 procedure, we recoded the AC dimensions to fit within the framework provided by Meriac et al. (Under Review; See Appendix A). Once the dimensions were recoded we proceeded through the steps outlined in previous section to create the DWE ratings, across exercise dimension ratings, across dimension exercise ratings, and the OARs for each participant.

In addition to completing the AC, participants also completed the Fifteen Factor Questionnaire (Psytech International, 2002) and the General Reasoning Test (Psytech International, 1991). To create scale level measures of the FFM we followed the steps outlined by the Fifteen Factor Questionnaire test manual (Psytech International, 2002). Further, GMA was measured by the General Reasoning Test and scored according to its test manual (Psytech International, 1991).

AC Model Structure

We used Mplus 7 (Muthén & Muthén, 2012) to test a series of CFA models that specified latent dimension factors within each exercise. Consistent with previous research (Hoffman et al., 2011; Hoffman & Meade, 2012; Meriac et al., 2014; Rupp et al., 2006), we grouped the manifest AC dimensions into broad performance categories (Appendix A contains a table that shows how the narrow dimensions from both AC samples were matched to the broader dimensions contained in the above models). To determine whether performance within an exercise could be meaningfully differentiated, we parameterized three AC models that differed in the number of latent dimension factors within each exercise (Hoffman & Meade, 2012; Meriac et al., 2014; Rupp et al., 2006).

The first model estimated three latent dimension factors within each exercise, which is consistent with the model Hoffman & Meade (2012) supported. However, in contrast to Hoffman

& Meade (2012), we relied on the dimension taxonomy provided by Meriac et al. (2014). Specifically, we parameterized a relational skills dimension, a drive dimension, and an administrative skills dimension. The relational skills dimension consists of the communication, consideration/awareness of others, influence others, and stress tolerance dimensions described by Arthur et al.'s (2003) AC dimension taxonomy. The drive dimension is identical to the drive dimension described by Arthur et al.'s (2003) AC dimension taxonomy. The administrative skills dimension consists of the two remaining dimensions: problem solving and planning and organizing (Arthur et al., 2003).

Next, the two-factor model DWE model consisted of a performance dimension, which was identical to the administrative factor estimated in the previous model, and an interpersonal dimension, which subsumed both the drive and relational factors in the previous model. The third model consisted of a general AC performance factor within each exercise.

Models that estimated improper parameter values (e.g. negative variances or correlations greater than 1) were considered to be non-convergent and were not evaluated (Lance et al., 2000; Marsh, 1994). Models that converged to a proper solution were evaluated using the chi-square statistic, the standardized root mean squared residual (SRMSR), comparative fit index (CFI; Bentler, 1990), and root mean square error of approximation (RMSEA; Steiger, 1990). Models were deemed to adequately fit the data if RMSEA values were below 0.06 and CFI values were greater than or equal to 0.95 (Hu & Bentler, 1999). Additionally, because each model was nested within the previous model we used a likelihood ratio test to compare each model.

Testing the Nomological Network

To test Hypotheses 1 through 3 we computed a correlation matrix for each sample. Each correlation matrix contained the four different AC rating variants (DWE, Dimensions-level,

Exercise-level, and OARs) in addition to the FFM and GMA variables. Next, in accordance with our hypotheses, we tested differences among the correlations between the DWE ratings and the FFM and GMA variables to the correlations obtained between the different AC rating variants and the FFM and GMA variables using the formulas provided by Meng, Rosenthal, and Rubin (1992). That is, to determine whether a rating on a given dimension as measured by a given exercise was more valid than any of the other AC ratings (Dimension-level ratings, Exercise-level ratings, and OARs), we compared the magnitude of the correlations between the DWE ratings and the theoretically relevant external variables (FFM and GMA) to the magnitude of the correlations between the dimension-level ratings, exercise-level ratings, and OARs and those same theoretically relevant external variables.

CHAPTER 3

RESULTS

AC Structure

Before testing the substantive hypotheses, it was necessary to examine the structure underlying the AC ratings using the methods outlines above. The results of these analyses are presented in Tables 1 and 2.

Sample 1. We first tested the a priori models against one another (Table 1). The first model we fit was the target three factor DWE model, which included three dimensions factors within each exercise. This model fit the data well with fit indices within the range of their respective cut-offs (TLI = .91, CFI = .94, SRMR = .05, and RMSEA = .06). Following this we fit the two factor DWE model, which specified a performance and interpersonal factor within each exercise, and used a chi-square difference test to compare it to the target. The two factor DWE model fit the data significantly worse than the target model ($\Delta \chi^2 = 82.95$, $\Delta = 18$, $p < .01$). Next, we tested the one factor DWE model, which specified a single exercise factor for each exercise. The one factor model also fit the data significantly worse than the target model ($\Delta \chi^2 = 119.45$, $\Delta = 30$, $p < .01$) Thus, the three factor DWE model was supported over the two competing models. However, because of the large correlations between the different AC dimensions within the same exercise (.51 - .94), we also fit a series of models where within a given exercise the AC dimensions were collapsed into the two dimensional and then one dimensional structure while retaining the three dimensional structure within the remaining two exercises. For example, the three factor DWE model was compared to a model that specified a performance and

interpersonal factor in the RP while retaining the three factor structure in the remaining exercises. Following this, it was found that the more parsimonious model 3A (Table 1), which specified a single AC dimension within the IB and three dimensions within the RP and LGD exercises, fit the data as well as the target model ($\Delta\chi^2 = 15.85$, $\Delta df = 14$, $p > .05$) and thus on the basis of model parsimony it was supported over the target model.

Sample 2. The target three factor DWE model, which specified the three broad AC dimensions for three of the four exercises and the administrative and relational dimensions for the RP exercise did not converge to an admissible solution (Table 2). However, we continued testing the competing theoretical models. The two factor DWE model (Model 2), which specified two dimensions (a performance and an interpersonal dimension) within each exercise also failed to converge to an admissible solution. Next, we tested the one factor DWE model (Model 3), which specified a single exercise factor for each exercise. This model converged to an admissible solution ($\chi^2 = 1072.34$, df = 623, TLI = .92, CFI = .92, SRMR = .06, RMSEA = .06). We were unable to use a nested model chi-square difference test to determine the best fitting model because neither the three factor DWE model nor two factor DWE model converged to an admissible solution. However, because the study depends on examining the nomological network of different dimensions within an exercise, we proceeded with tests of differential correlations among the dimensions within exercises. The Sample 1 and 2 correlations we tested can be found in tables three and four, respectively.

Overall Dimensions and Exercise Factors

Before reporting the nomological network results for the DWE ratings, we will first examine the relationships between the three broad AC dimensions, exercise factors, the OAR and the external variables (Tables 3, 4, 5, & 6).

Sample 1. The administrative skills dimension was only significantly related to agreeableness (r = -.14) and was weakly related to the remaining personality correlates with correlations ranging from -.13 with conscientiousness to .03 with extraversion. The administrative skills dimension was significantly and positively correlated with GMA (r = .39). The relational skills dimension did not correlate significantly with any of the personality factors with correlations ranging from -.08 with agreeableness to .05 with extraversion. The relational skills dimension also correlated significantly with GMA (r = .35). The drive dimension was only significantly correlated with openness (r = .15). It was weakly correlated with the remaining personality factors with correlations ranging from -.11 with conscientiousness to .12 with extraversion. The drive dimension significantly correlated with GMA (r = .23).

Next, the IB exercise factor did not significantly correlate with any of the personality factors with correlations ranging from -.09 with agreeableness and .11 with extraversion. The IB exercise factor correlated significantly with GMA (r = .33). The RP exercise factor did not significantly correlate with any of the personality factors with correlations ranging from -.04 with openness to .07 with agreeableness. The RP exercise factor was also the only exercise factor that did not significantly correlate with GMA (r = .09). The LGD exercise factor significantly and negatively correlated with both conscientiousness (r = -.18) and agreeableness (r = -.15). It was not significantly related to any of the remaining personality factors with correlations ranging from -.03 with emotional stability to .03 with openness. The LGD exercise factor was also significantly and positively correlated with GMA (r = .31).

Finally, the OAR was not significantly related to any of the personality factors with correlations ranging from -.11 with both conscientiousness and agreeableness to .06 with extraversion. However, the OAR was the strongest correlate of GMA (r = .41, $p < .01$).

Sample 2. The administrative skills dimension correlated significantly with openness (r = .14), extraversion (r = .25), and agreeableness (r = -.17). It was weakly related to both conscientiousness (r = -.03) and emotional stability (r = .09). The administrative skills dimension was significantly and positively correlated with GMA (r = .32). Similarly, the relational skills dimension correlated significantly with openness (r = .15), extraversion (r = .25), and agreeableness (r = -.15) and was weakly related to conscientiousness (r = .00) and emotional stability (r = .09). It was also significantly and positively correlated with GMA (r = .31). The drive dimension correlated significantly with extraversion (r = .27) and agreeableness (r = -.19) and it correlated non-significantly with openness (r = .12), conscientiousness (r = -.02), and emotional stability (r = .11). It also significantly and positively correlated with GMA (r = .29).

Next, the first OP exercise factor correlated significantly with extraversion (r = .21) and agreeableness (r = -.15) and it correlated non-significantly with openness (r = .13), conscientiousness (r = -.02), and emotional stability (r = .10). It also significantly and positively correlated with GMA (r = .26). The second OP exercise factor only significantly correlated with extraversion (r = .18) and correlated non-significantly with the remaining personality factors with correlations ranging from -.13 with agreeableness to .11 with conscientiousness. It also correlated significantly with GMA (r = .26). Similar to the first OP exercise factor, the CA exercise factor correlated significantly with extraversion (r = .26) and agreeableness (r = -.16) and it correlated non-significantly with openness (r = .08), conscientiousness (r = -.05), and emotional stability (r = .10). It also significantly and positively correlated with GMA (r = .22). The RP exercise factor significantly correlated with extraversion (r = .18) and correlated non-significantly with the remaining personality factors with correlations ranging from -.11 with agreeableness to .13 with conscientiousness. It also correlated significantly with GMA (r = .26).

Finally, the OAR significantly correlated with openness (r = .15), extraversion (r = .27), and agreeableness (r = -.18) and only weakly and non-significantly correlated with conscientiousness (r = -.02) and emotional stability (r = .10). The OAR was also found to be the strongest correlate of GMA (r = .33, $p < .01$).

Across Samples 1 and 2 it was found that the nomological networks for the AC dimensions and exercise factors were similar. That is, the dimension, exercise, and OAR scores were all weakly related to the FFM variables, with the largest correlations occurring with either openness or extraversion and moderately related to GMA. This pattern of results has been found in previous meta-analyses (Hoffman, et al., 2014; Meriac et al., 2014; Meriac et al., 2008). Next we examined the nomological networks of the DWE ratings to determine if they are more strongly related to the external variables than the broad AC dimensions, exercise factors, and OARs.

Relational Skills Dimension

Sample 1. To examine the nomological network of the DWE ratings we compared their correlations with external variables (FFM and GMA). We then compared the nomological network of DWEs to that of dimension ratings, exercise ratings, and OARs. Hypothesis 1a posited that the strongest AC correlate of emotional stability would be the relational skills dimension when measured by the LGD exercise; however this correlation was weak (r = -.02) and did not differ significantly from other correlations between AC ratings and emotional stability (i.e. other DWE ratings, overall dimension ratings, overall exercise ratings, and OAR; See Table 7). Thus, Hypothesis 1a was not supported.

Next, Hypothesis 1b posited that the strongest AC correlate of agreeableness would be the relational skills dimension when measured by the RP exercise. Again, these two measures

were not-significantly correlated (r = .07), but this correlation was significantly more strong than nine out of the fifteen correlations it was compared to (see Table 8). Specifically, when measured by the RP exercise the relational skills dimension correlated more positively and strongly with agreeableness than did the OAR, the overall relational skills dimension, overall administrative skills dimension, the overall LGD exercise factor, the overall IB exercise factor, the relational skills dimension when measured by the LGD, the drive dimension when measured by the IB, and the administrative skills dimension when measured by both the IB and LGD. However, it was not significantly different than the correlations between agreeableness and the drive or administrative performance dimension from the role play exercise or the overall role play exercise. Together, these findings provide do not support Hypothesis 1b.

Finally, Hypothesis 1c posited that the strongest AC correlates of conscientiousness would be the relational skills dimensions when measured by the LGD, OP, or IB exercises. Because Sample 1 did not include an OP exercise, we were only able to examine the LGD and IB exercises. Specifically, when measured by the LGD exercise, the relational skills dimension was found to be significantly and negatively related to conscientiousness (r = -.15) and when measured by the IB exercise, the relational skills dimension was found to be non-significantly and positively related to conscientiousness (r = .05). Regarding the correlation between the relational skills dimension measured by the IB and conscientiousness our results found that it was not significantly different from than any of the overall dimension scores, overall exercise scores, or the OAR.

As for the relationship between conscientiousness and the relational skills dimensions measured by the IB exercise, it was significantly different from the relationship between conscientiousness and the administrative skills dimension when measured by the LGD exercise,

as well as all three overall AC dimension ratings, the overall LGD exercise ratings, and the OAR (see Table 9). As such, Hypothesis 1c was only partially supported.

Sample 2. Contrary to Hypothesis 1a, agreeableness correlated negatively and non-significantly with the relational skills dimension when measured by the RP exercise (r = -.12, see Table 7). Moreover, this correlation did not differ significantly from the other correlations between agreeableness and the other AC ratings, as agreeableness was weakly and negatively correlated with the majority of the various AC ratings. As such, the results for Sample 2 did not support Hypothesis 1b.

As for Hypothesis 1c, the correlations between conscientiousness and both the relational skills dimension when measured by the first and second OP exercises were both non-significant (r = -.01 and r = .08, respectively). Specifically, the correlation between conscientiousness and the relational skills dimension when measured by the first OP exercise was not significantly different than the correlations between conscientiousness and the other AC ratings (See Table 9). However, the correlation between conscientiousness and the relational skills dimension when measured by the second OP exercise was found to be significantly more positive than the correlations between conscientiousness and the administrative skills dimension when measured by both the OP1 and CA exercises, the drive dimensions when measured by both the OP2 and CA exercises, the overall administrative skills rating the overall drive rating, the overall CA rating, and the OAR because conscientiousness was actually negatively correlated with these ratings. On the basis of the differential correlations, the Sample 2 results provided modest support for Hypothesis 1c.

Drive Dimension

Sample 1. Hypothesis 2a posited that the strongest AC correlate of extraversion is the drive dimension when measured by the LGD and RP. However, extraversion was non-significantly and positively correlated with both the drive dimension when measured by the LGD exercise (r = .04) and when measured by the RP exercise (r = .05). Next, extraversion was only significantly more strongly related to the drive dimension when measured by the LGD than the administrative skills dimension when measured by the LGD. However, the correlation between drive measured by the LGD and extraversion was not different from the correlations between extraversion and any of the overall dimension scores, exercise scores, or the OAR. Moreover, extraversion did not correlate with drive when drive was measured under the role play, nor was it significantly different from the relationships between extraversion and the other AC ratings (See Table 10). Thus, Hypothesis 2a was not supported.

Next, Hypothesis 2b posited that drive dimension when measured by the LGD, RP, OP, and CA exercises would correlate more strongly with openness than would other AC scores. Unfortunately, openness was not significantly correlated with the either the drive dimension when measured by the LGD exercise (r = .12) or the drive dimension when measured by the RP exercise (r = .09). But, the correlation between openness and the drive dimension when measured by the LGD exercise was significantly different than the correlations between openness and the administrative skills dimension when measured by the IB exercise, openness and the relational skills dimension when measured by the RP exercise, as well as openness and the relational skills dimension when measured by the LGD exercise. Moreover, the correlation was significantly greater than the correlation between openness and the overall relational skills dimension rating, the overall RP exercise rating, and the OAR.

As for the relationship between openness and the drive dimension when measured by the RP exercise, it was only significantly more positive than the relationship between openness and the relational dimension when measured by the RP exercise. The correlation did not differ from the correlations seen between openness and any of the overall dimensions, exercises, or the OAR. Thus, the results only partially supported Hypothesis 2b.

Sample 2. Focusing on Hypothesis 2a, the correlations between extraversion and the drive dimension when measured by both the first and second OP exercises were significant and positive ($r = .24$ and $r = .16$, respectively). However, similar to the Sample 1 results, the correlation between conscientiousness and the drive dimension when measured by the first OP exercise did not differ significantly from the majority of correlations between extraversion and the other AC scoring units. Next, the correlation between extraversion and the drive dimension when measured by the second OP exercise was not significantly larger than the other correlations found between extraversion and the different AC ratings. Thus, the Sample 2 results did not support Hypothesis 2a.

Concerning Hypothesis 2b, the correlations found between openness and drive as measured by the CA exercise ($r = .08$), openness and drive as measured by the first OP exercise ($r = .14$), and openness and drive as measured by the second OP exercise ($r = .07$) were relatively weak and the only significant correlation was between openness and drive from the first OP exercise. Moreover, none of the above correlations with openness were significantly different from the correlations found between openness and the other AC ratings (see Table 11). Consequently, Hypothesis 2b was not supported by the Sample 2 results.

Administrative Skills Dimension

Sample 1. Hypothesis 3 posited that the strongest, positive AC correlate of GMA is the administrative skills dimension when measured by either the IB or CA exercises. The results showed that GMA significantly and moderately correlated with the administrative skills dimension when measured by the IB exercise (r = .32). As can be seen in Table 12, this relationship was significantly more positive than the correlation exhibited between GMA and all three AC dimensions when measured by the RP exercise as well as the drive dimension when measured by the LGD exercise. However, this correlation was not significantly more positive than the overall administrative skills dimension rating, the overall IB exercise rating, or the OAR (see Table 12). Thus, the results partially supported Hypothesis 3.

Sample 2. Similar to the Sample 1 results, the correlation found between GMA and the administrative skills dimension when measured by the CA exercise was significant and moderate (r = .22). However, this correlation was not significantly different from the correlations found between GMA and the other AC ratings including the overall dimension ratings, exercise ratings, and the OAR (see Table 12). Thus, Hypothesis 3 was not supported by the Sample 2 results.

Table 1. Sample 1 Confirmatory Factor Analysis Results

Model	χ^2	df	Δ χ2	Δ df	TLI	CFI	SRMR	RMSEA
1. Nine factor - Target Model (Baseline)	213.30	119			.91	.94	.05	.06
Six Factor Model								
2. Two factors for all exercises	296.25	137	82.95**	18	.87	.90	.06	.08
2a. Two factors for IB	222.59	126	9.29	7	.92	.94	.05	.06
2b. Two factors for RP	247.19	126	33.89**	7	.89	.92	.06	.07
2c. Two factors for LGD	255.22	126	41.92**	7	.89	.92	.05	.07
Three Factor Model								
3. One factor for each exercise	332.75	149	119.45**	30	.86	.88	.06	.08
3a. One factor for IB	229.15	133	15.85	14	.92	.94	.05	.06
3b. One factor for RP	254.90	133	41.60**	14	.90	.92	.06	.07
3c. One factor for LGD	283.87	133	70.57**	14	.87	.90	.08	.05

Note: *$p \leq .05$; **$p \leq .01$.

Table 2. Sample 2 Confirmatory Factor Analysis Results

Model	χ^2	df	$\Delta \chi 2$	Δ df	TLI	CFI	SRMR	RMSEA
1. Target Model[a]	995.11	574			.92	.93	.05	.06
Eight Factor Model								
2. Two factors for each exercise[a]	1053.20	600			.91	.92	.06	.06
2a. Two factors for OP1[a]	1020.57	584			.92	.93	.06	.06
2b. Two factors for OP2[a]	1014.87	584			.92	.93	.05	.06
2c. Two factors for CA[a]	1008.64	584			.92	.93	.05	.06
Four Factor Model								
3. One factor for each exercise	1072.34	623	13.70	15	.92	.92	.06	.06
3a. One factor for OP1[a]	1055.43	608			.92	.92	.06	.06
3b. One factor for OP2	1058.64	608			.92	.92	.06	.06
3c. One factor for CA[a]	1055.89	608			.92	.92	.06	.06
3d. One factor for RP[a]	1064.14	608			.92	.92	.06	.06

Note. *Note*: *$p \leq .05$; **$p \leq .01$; [a]Inadmissible Solution.

Table 3. Sample One Correlation Matrix: Part One

	1. IB-AD	2. IB-REL	3. IB-DR	4. RP-AD	5. RP-REL	6. RP-DR	7. LGD-AD	8. LGD-REL	9. LGD-DR	10. AD	11. REL
1.	1.00										
2.	***.74***	1.00									
3.	.62	***.57***	1.00								
4.	.09	.08	-.01	1.00							
5.	.12	**.17**	.06	***.42***	1.00						
6.	-.02	-.04	-.05	***.33***	-.03	1.00					
7.	.16	.02	.13	.07	-.01	.07	1.00				
8.	***.21***	.11	**.15**	.06	.03	.01	***.65***	1.00			
9.	.07	.05	.10	.06	.02	.03	***.55***	***.38***	1.00		
10.	***.73***	***.49***	***.45***	***.49***	***.24***	**.15**	***.66***	***.51***	***.36***	1.00	
11.	.54	***.62***	***.38***	.27	***.55***	-.02	***.43***	***.72***	***.28***	***.66***	1.00
12.	***.35***	***.31***	***.56***	***.22***	.02	***.56***	***.43***	***.31***	***.66***	***.53***	***.35***
13.	***.93***	***.92***	***.74***	.08	**.15**	-.04	.11	***.18***	.07	***.66***	***.61***
14.	.11	.14	.02	***.80***	***.83***	***.40***	.05	.05	.05	***.41***	***.47***
15.	***.20***	.09	**.16**	.08	.02	.04	***.87***	***.92***	***.62***	***.62***	***.65***
16.	***.68***	***.61***	***.51***	***.40***	***.41***	**.16**	***.60***	***.67***	***.43***	***.89***	***.91***
17.	-.03	-.01	.06	.06	-.13	.09	.02	-.01	.12	.02	-.07
18.	-.02	.05	-.04	-.03	.04	-.10	***-.20***	**-.15**	-.05	-.13	-.06
19.	.07	.11	.12	.05	-.01	.05	-.07	.00	.04	.03	.05
20.	-.12	-.04	-.11	.04	.07	.02	**-.16**	**-.14**	-.02	**-.14**	-.08
21.	-.08	.03	-.02	.07	.01	.11	-.06	-.03	.07	-.05	.00
22.	***.32***	***.30***	***.23***	.07	.08	.02	***.30***	***.28***	**.17**	***.39***	***.35***

Note: Bold values denote $p < .05$; Bold and italicized values denote $p < .01$; IB = In-Basket; CA = Case Analysis; RP = Role Play; AD = Administrative Skills; REL = Relational Skills; DR = Drive; OAR = Overall Assessment Rating; O = Openness; C = Conscientiousness; E = Extraversion; A = Agreeableness; ES = Emotional Stability; GMA = General Mental Ability.

Table 4. Sample One Correlation Matrix: Part Two

	12. DR	13. IB	14. RP	15. LGD	16. OAR	17. O	18. C	19. E	20. A	21. ES	22. GMA
12.	1.00										
13.	***.41***	1.00									
14.	***.27***	.12	1.00								
15.	***.47***	**.16**	.06	1.00							
16.	***.61***	***.70***	***.49***	***.71***	1.00						
17.	**.15**	-.01	-.04	.03	.00	1.00					
18.	-.11	.01	-.01	**-.18**	-.11	***.22***	1.00				
19.	.12	.11	.03	-.02	.06	***.65***	***.26***	1.00			
20.	-.06	-.09	.07	**-.15**	-.11	***.46***	***.44***	***.41***	1.00		
21.	.09	-.02	.06	-.03	.00	***.48***	***.31***	***.56***	***.41***	1.00	
22.	***.23***	***.33***	.09	***.31***	***.41***	**.14**	-.12	.05	-.08	.02	1.00

Note: Bold values denote $p < .05$; Bold and italicized values denote $p < .01$; IB = In-Basket; CA = Case Analysis; RP = Role Play; AD = Administrative Skills; REL = Relational Skills; DR = Drive; OAR = Overall Assessment Rating; O = Openness; C = Conscientiousness; E = Extraversion; A = Agreeableness; ES = Emotional Stability; GMA = General Mental Ability.

Table 5. Sample Two Correlation Matrix: Part One

	1. OP1-AD	2. OP1-REL	3. OP1-DR	4. OP2-AD	5. OP2-REL	6. OP2-DR	7. CA-AD	8. CA-REL	9. CA-DR	10. RP-AD	11. RP-REL	12. AD
1.	1.00											
2.	***.72***	1.00										
3.	***.84***	***.72***	1.00									
4.	***.45***	***.50***	***.49***	1.00								
5.	***.50***	***.54***	***.49***	***.79***	1.00							
6.	***.45***	***.49***	***.48***	***.83***	***.82***	1.00						
7.	***.41***	***.42***	***.42***	***.34***	***.45***	***.36***	1.00					
8.	***.44***	***.41***	***.43***	***.35***	***.45***	***.37***	***.81***	1.00				
9.	***.46***	***.43***	***.45***	***.32***	***.44***	***.36***	***.86***	***.87***	1.00			
10.	***.46***	***.48***	***.45***	***.35***	***.41***	***.38***	***.27***	***.29***	***.30***	1.00		
11.	***.42***	***.48***	***.43***	***.38***	***.41***	***.40***	***.26***	***.29***	***.28***	***.86***	1.00	
12.	***.81***	***.73***	***.75***	***.70***	***.71***	***.66***	***.65***	***.62***	***.63***	***.77***	***.70***	1.00
13.	***.63***	***.72***	***.63***	***.61***	***.70***	***.62***	***.54***	***.60***	***.57***	***.80***	***.89***	***.89***
14.	***.74***	***.69***	***.83***	***.66***	***.71***	***.73***	***.72***	***.73***	***.80***	***.47***	***.46***	***.86***
15.	***.95***	***.84***	***.95***	***.51***	***.54***	***.50***	***.45***	***.46***	***.49***	***.49***	***.47***	***.83***
16.	***.49***	***.54***	***.52***	***.94***	***.91***	***.95***	***.40***	***.41***	***.39***	***.40***	***.43***	***.73***
17.	***.46***	***.45***	***.47***	***.35***	***.47***	***.38***	***.94***	***.93***	***.97***	***.31***	***.29***	***.67***
18.	***.45***	***.50***	***.45***	***.38***	***.43***	***.41***	***.28***	***.30***	***.30***	***.95***	***.98***	***.75***
19.	***.77***	***.76***	***.78***	***.69***	***.75***	***.71***	***.67***	***.68***	***.70***	***.72***	***.73***	***.97***
20.	.09	.13	**.14**	**.15**	.08	.07	.04	.11	.08	.13	.12	**.14**
21.	-.06	-.01	.02	.04	.08	.01	-.06	.00	-.07	-.01	-.03	-.03
22.	**.14**	***.23***	***.24***	***.23***	.10	**.16**	***.24***	***.28***	***.24***	**.16**	***.19***	***.25***
23.	-.12	-.07	-.19	**-.16**	-.11	-.10	**-.16**	**-.17**	**-.14**	-.09	-.12	**-.17**
24.	.09	.09	.09	.04	.07	.06	.10	.09	.10	.04	.05	.09
25.	***.26***	***.26***	***.21***	***.20***	***.27***	***.28***	***.22***	***.19***	***.21***	***.26***	***.24***	***.32***

Note: Bold values denote $p < .05$; Bold and italicized values denote $p < .01$; OP1 = Oral Presentation 1; OP2 = Oral Presentation 2; CA = Case Analysis; RP = Role Play; AD = Administrative Skills; REL = Relational Skills; DR = Drive; OAR = Overall Assessment Rating; O = Openness; C = Conscientiousness; E = Extraversion; A = Agreeableness; ES = Emotional Stability; GMA = General Mental Ability.

Table 6. Sample Two Correlation Matrix: Part Two

	13. REL	14. DR	15. OP1	16. OP2	17. CA	18. RP	19. OAR	20. O	21. C	22. E	23. A	24. ES	25. GMA
13.	1.00												
14.	***.76***	1.00											
15.	***.70***	***.83***	1.00										
16.	***.68***	***.74***	***.55***	1.00									
17.	***.59***	***.80***	***.50***	***.42***	1.00								
18.	***.88***	***.48***	***.50***	***.43***	***.31***	1.00							
19.	***.94***	***.92***	***.83***	***.76***	***.72***	***.75***	1.00						
20.	**.15**	.12	.13	.11	.08	.13	**.15**	1.00					
21.	.00	-.02	-.02	.04	-.05	-.02	-.02	**-.14**	1.00				
22.	***.25***	***.27***	***.21***	***.18***	***.26***	***.18***	***.27***	***.47***	.07	1.00			
23.	**-.15**	***-.19***	**-.15**	-.13	**-.16**	-.11	***-.18***	***-.44***	.04	***-.43***	1.00		
24.	.09	.11	.10	.06	.10	.05	.10	.10	.09	***-.32***	-.12	1.00	
25.	***.31***	***.29***	***.26***	***.26***	***.22***	***.26***	***.33***	-.02	***-.27***	-.04	-.04	.05	1.00

Note: Bold values denote $p < .05$; Bold and italicized values denote $p < .01$; OP1 = Oral Presentation 1; OP2 = Oral Presentation 2; CA = Case Analysis;
RP = Role Play; AD = Administrative Skills; REL = Relational Skills; DR = Drive; OAR = Overall Assessment Rating; O = Openness; C = Conscientiousness; E = Extraversion; A = Agreeableness; ES = Emotional Stability; GMA = General Mental Ability.

Table 7. Hypothesis 1a – AC Ratings Correlated with Emotional Stability

	Sample 1 – Correlation Differences
Comparison	LGD-REL
IB-AD	.05 (-.12 - .22)
IB-REL	-.05 (-.24 - .13)
IB-DR	-.01 (-.19 - .17)
RP-AD	-.09 (-.28 - .10)
RP-REL	-.03 (-.23 - .16)
RP-DR	-.13 (-.33 - .06)
LGD-AD	.04 (-.08 - .16)
LGD-DR	-.10 (-.25 - .06)
AD	.03 (-.11 - .17)
REL	-.03 (-.13 - .08)
DR	-.12 (-.28 - .04)
IB	.00 (-.18 - .18)
RP	-.09 (-.28 - .10)
LGD	.00 (-.05 - .06)
OAR	-.02 (-.13 - .09)

Note: *$p \leq .05$; IB = In-Basket; RP = Role Play; LGD = Leaderless Group Discussion; AD = Administrative Skills; REL = Relational Skills; DR = Drive; OAR = Overall Assessment Rating

Table 8. Hypothesis 1b – AC Ratings Correlated with Agreeableness

Sample 1 – Correlation Differences		Sample 2 – Correlation Differences	
Comparison	RP-REL	Comparison	RP-REL
IB-AD	.19* (.00 - .37)	OP1-AD	.00 (-.14 - .15)
IB-REL	.11 (-.07 - .29)	OP1-REL	-.05 (-.19 - .09)
IB-DR	.18* (-.01 - .37)	OP1-DR	.07 (-.07 - .22)
RP-AD	.03 (-.12 - .18)	OP2-AD	.04 (-.11 - .19)
RP-DR	.05 (-.15 - .25)	OP2-REL	.00 (-.15 - .14)
LGD-AD	.23* (.03 - .43)	OP2-DR	-.01 (-.16 - .13)
LGD-REL	.21* (.01 - .40)	CA-AD	.04 (-.12 - .21)
LGD-DR	.09 (-.10 - .28)	CA-REL	.06 (-.11 - .22)
AD	.21* (.04 - .38)	CA-DR	.02 (-.14 - .19)
REL	.15* (.01 - .28)	RP-AD	-.03 (-.10 - .04)
DR	.13 (-.06 - .32)	AD	.05 (-.05 - .16)
IB	.16* (-.02 - .34)	REL	.04 (-.03 - .10)
RP	.00 (-.08 - .08)	DR	.07 (-.07 - .21)
LGD	.22* (.02 - .41)	OP1	.03 (-.11 - .17)
OAR	.18* (.03 - .33)	OP2	.02 (-.13 - .16)
	--	CA	.05 (-.12 - .21)
	--	RP	-.01 (-.04 - .02)
	--	OAR	.06 (-.04 - .16)

Note: $*p \leq .05$; IB = In-Basket; RP = Role Play; LGD = Leaderless Group Discussion; OP = Oral Presentation; CA = Case Analysis; AD = Administrative Skills; REL = Relational Skills; DR = Drive; OAR = Overall Assessment Rating.

Table 9. Hypothesis 1c – AC Ratings Correlated with Conscientiousness

	Sample 1 – Correlation Differences			Sample 2 – Correlation Differences	
Comparison	LGD-REL	IB-REL	Comparison	OP1-REL	OP2-REL
IB-AD	-.13 (-.31 - .04)	.07 (-.03 - .17)	OP1-AD	.05 (-.05 - .15)	.14* (.01 - .28)
IB-DR	-.11 (-.29 - .07)	.09 (-.04 - .22)	OP1-DR	-.03 (-.13 - .07)	.07 (-.07 - .20)
RP-AD	-.13 (-.32 - .06)	.08 (-.11 - .26)	OP2-AD	-.05 (-.18 - .08)	.04 (-.05 - .13)
RP-DR	-.20 (-.39 - .00)	.01 (-.17 - .18)	OP2-DR	-.02 (-.15 - .12)	.07* (-.01 - .16)
RP-REL	-.05 (-.25 - .14)	.15 (-.05 - .35)	CA-AD	.05 (-.09 - .20)	.14* (.00 - .29)
LGD-AD	.05 (-.06 - .17)	.26* (.06 - .45)	CA-REL	-.01 (-.16 - .13)	.08 (-.06 - .22)
LGD-DR	-.11 (-.26 - .05)	.10 (-.09 - .29)	CA-DR	.06 (-.09 - .20)	.15* (.01 - .29)
AD	-.02 (-.16 - .12)	.18* (.04 - .32)	RP-AD	-.01 (-.14 - .13)	.09 (-.06 - .23)
REL	-.10 (-.20 - .01)	.11* (-.02 - .23)	RP-REL	.02 (-.11 - .16)	.11 (-.03 - .26)
DR	-.05 (-.21 - .12)	.16* (-.01 - .32)	AD	.02 (-.08 - .12)	.11* (.01 - .22)
IB	-.16 (-.34 - .02)	.04 (-.01 - .10)	REL	-.01 (-.11 - .09)	.08 (-.02 - .19)
RP	-.14 (-.33 - .05)	.06 (-.12 - .24)	DR	.01 (-.09 - .12)	.10* (.00 - .21)
LGD	.03 (-.03 - .08)	.23* (.04 - .42)	OP1	.01 (-.07 - .09)	.10 (-.03 - .23)
OAR	-.05 (-.16 - .07)	.16* (.04 - .28)	OP2	-.05 (-.18 - .08)	.04 (-.02 - .10)
	--	--	CA	.04 (-.10 - .19)	.13* (-.01 - .27)
	--	--	RP	.01 (-.12 - .15)	.10 (-.04 - .25)
			OAR	.01 (-.08 - .10)	.10* (.01 - .20)

Note. *$p \leq .05$; IB = In-Basket; RP = Role Play; LGD = Leaderless Group Discussion; OP = Oral Presentation; CA = Case Analysis; AD = Administrative Skills; REL = Relational Skills; DR = Drive; OAR = Overall Assessment Rating.

Table 10. Hypothesis 2a – AC Ratings Correlated with Extraversion

	Sample 1 – Correlation Differences			Sample 2 – Correlation Differences	
Comparison	LGD-DR	RP-DR	Comparison	OP1-DR	OP2-DR
IB-AD	-.03 (-.22 - .16)	-.02 (-.22 - .18)	OP1-AD	.10* (.02 - .18)	.02 (-.13 - .16)
IB-REL	-.07 (-.26 - .12)	-.06 (-.26 - .14)	OP1-REL	.00 (-.10 - .11)	-.08 (-.22 - .06)
IB-DR	-.08 (-.27 - .10)	-.08 (-.28 - .13)	OP2-AD	.01 (-.13 - .15)	-.07 (-.15 - .00)
RP-AD	-.01 (-.20 - .18)	.00 (-.16 - .16)	OP2-REL	.14* (.00 - .28)	.05 (-.03 - .14)
RP-REL	.06 (-.14 - .25)	.06 (-.14 - .26)	CA-AD	.00 (-.15 - .15)	-.09 (-.24 - .07)
LGD-AD	.11* (-.02 - .24)	.12 (-.07 - .30)	CA-REL	-.04 (-.19 - .11)	-.13 (-.28 - .03)
LGD-REL	.04 (-.11 - .19)	.05 (-.15 - .24)	CA-DR	.00 (-.14 - .14)	-.09 (-.24 - .07)
AD	.02 (-.14 - .17)	.02 (-.16 - .20)	RP-AD	.08 (-.07 - .22)	-.01 (-.16 - .15)
REL	-.01 (-.17 - .16)	.00 (-.20 - .20)	RP-REL	.05 (-.10 - .20)	-.03 (-.18 - .12)
DR	-.08 (-.19 - .04)	-.07 (-.20 - .06)	AD	-.02 (-.11 - .08)	-.10 (-.21 - .01)
IB	-.07 (-.26 - .12)	-.06 (-.26 - .14)	REL	-.02 (-.14 - .10)	-.10 (-.22 - .02)
RP	.02 (-.17 - .21)	.02 (-.13 - .17)	DR	-.04 (-.12 - .04)	-.12 (-.22 - -.02)
LGD	.06 (-.06 - .18)	.06 (-.13 - .26)	OP1	.02 (-.02 - .07)	-.06 (-.20 - .08)
OAR	-.02 (-.17 - .13)	-.01 (-.19 - .17)	OP2	.06 (-.08 - .20)	-.02 (-.07 - .02)
	--	--	CA	-.02 (-.17 - .12)	-.11 (-.26 - .04)
	--	--	RP	.06 (-.09 - .20)	-.03 (-.18 - .12)
	--	--	OAR	-.04 (-.13 - .05)	-.12 (-.23 - -.02)

Note: $*p \leq .05$; IB = In-Basket; RP = Role Play; LGD = Leaderless Group Discussion; OP = Oral Presentation; CA = Case Analysis; AD = Administrative Skills; REL = Relational Skills; DR = Drive; OAR = Overall Assessment Rating.

Table 11. Hypothesis 2b – AC Ratings Correlated with Openness

Sample 1 – Correlation Differences			Sample 2 – Correlation Differences			
Comparison	LGD-DR	RP-DR	Comparison	CA-DR	OP1-DR	OP2-DR
IB-AD	.16* (-.03 - .35)	.12 (-.08 - .32)	OP1-AD	-.01 (-.25 - .13)	.05 (-.03 - .13)	-.02 (-.16 - .13)
IB-REL	.14 (-.05 - .33)	.10 (-.10 - .30)	OP1-REL	-.05 (-.20 - .09)	.01 (-.10 - .11)	-.06 (-.20 - .07)
IB-DR	.07 (-.12 - .25)	.03 (-.17 - .23)	OP2-AD	-.07 (-.23 - .09)	-.01 (-.15 - .13)	-.08 (-.16 - .00)
RP-AD	.06 (-.13 - .25)	.02 (-.14 - .18)	OP2-REL	.00 (-.15 - .14)	.06 (-.08 - .19)	-.01 (-.09 - .07)
RP-REL	.26* (.06 - .45)	.22* (.02 - .42)	CA-AD	.03 (-.04 - .11)	.09 (-.05 - .24)	.03 (-.13 - .18)
LGD-AD	.10 (-.03 - .23)	.06 (-.13 - .25)	CA-REL	-.03 (-.10 - .04)	.03 (-.12 - .17)	-.04 (-.19 - .11)
LGD-REL	.13* (-.02 - .28)	.09 (-.10 - .29)	RP-AD	-.05 (-.21 - .11)	.01 (-.13 - .15)	-.06 (-.21 - .09)
AD	.11 (-.05 - .26)	.07 (-.11 - .25)	RP-REL	-.05 (-.21 - .12)	.01 (-.13 - .16)	-.05 (-.20 - .09)
REL	.19* (.03 - .36)	.16 (-.04 - .35)	AD	-.06 (-.18 - .05)	.00 (-.10 - .09)	-.07 (-.18 - .04)
DR	-.03 (-.14 - .09)	-.06 (-.19 - .07)	REL	-.07 (-.20 - .06)	-.01 (-.13 - .11)	-.08 (-.20 - .04)
IB	.14 (-.05 - .32)	.10 (-.10 - .30)	DR	-.04 (-.13 - .04)	.02 (-.06 - .10)	-.05 (-.15 - .05)
RP	.16* (-.03 - .35)	.12 (-.03 - .27)	OP1	-.05 (-.18 - .09)	.01 (-.03 - .06)	-.06 (-.19 - .08)
LGD	.10 (-.03 - .22)	.06 (-.13 - .25)	OP2	-.03 (-.18 - .12)	.03 (-.10 - .16)	-.04 (-.08 - .00)
OAR	.13* (-.02 - .27)	.09 (-.09 - .27)	CA	.00 (-.03 - .03)	.06 (-.08 - .20)	-.01 (-.16 - .14)
	--	--	RP	-.05 (-.21 - .11)	.01 (-.14 - .15)	-.06 (-.21 - .09)
	--	--	OAR	-.07 (-.17 - .04)	-.01 (-.10 - .08)	-.08 (-.18 - .03)

Note: $*p \leq .05$; IB = In-Basket; RP = Role Play; LGD = Leaderless Group Discussion; OP = Oral Presentation; CA = Case Analysis; AD = Administrative Skills; REL = Relational Skills; DR = Drive; OAR = Overall Assessment Rating.

Table 12. Hypothesis 3 – AC Ratings Correlated with GMA

	Sample 1 – Correlation Differences		Sample 2 – Correlation Differences
Comparison	IB-AD	Comparison	CA-AD
IB-REL	.02 (-.08 - .13)	OP1-AD	-.04 (-.19 - .11)
IB-DR	.10 (-.08 - .22)	OP1-REL	-.05 (-.19 - .10)
RP-AD	.26* (.07 - .45)	OP1-DR	.01 (-.14 - .16)
RP-REL	.25* (.06 - .44)	OP2-AD	.02 (-.13 - .18)
RP-DR	.31* (.11 - .51)	OP2-REL	-.05 (-.20 - .09)
LGD-AD	.02 (-.16 - .21)	OP2-DR	-.06 (-.22 - .09)
LGD-REL	.05 (-.13 - .23)	CA-REL	.03 (-.05 - .12)
LGD-DR	.16* (-.03 - .35)	CA-DR	.01 (-.06 - .09)
AD	-.08 (-.19 - .03)	RP-AD	-.04 (-.21 - .13)
REL	-.04 (-.18 - .10)	RP-REL	-.02 (-.19 - .14)
DR	.09 (-.07 - .26)	AD	-.11 (-.22 - .01)
IB	-.01 (-.07 - .04)	REL	-.10 (-.23 - .04)
RP	.24* (.05 - .43)	DR	-.07 (-.18 - .03)
LGD	.01 (-.17 - .19)	OP1	-.04 (-.18 - .11)
OAR	-.10 (-.22 - .02)	OP2	-.05 (-.20 - .10)
	--	CA	.00 (-.05 - .05)
	--	RP	-.04 (-.20 - .13)
	--	OAR	-.11 (-.23 - .00)

Note: *$p \leq .05$; IB = In-Basket; RP = Role Play; LGD = Leaderless Group Discussion; OP = Oral Presentation;
CA = Case Analysis; AD = Administrative Skills; REL = Relational Skills; DR = Drive; OAR = Overall Assessment Rating.

CHAPTER 4

DISCUSSION

The purpose of this paper was to provide the first examination of the nomological network of the DWE scoring approach to ACs and whether the DWE approach yields more favorable construct validity evidence relative to traditional scoring approaches. Although DWEs were supported in two of the exercises from Sample 1, DWEs were not supported in Sample 2. In addition, similar levels of nomological network evidence were found for DWEs, OARs, exercise scores, and dimension scores. Together, these findings suggest that although the DWE approach may hold applied value in reporting and interpreting AC results (Borman, 2012) the DWE approach may not enhance nomological network evidence of ACs relative to more traditional scoring strategies.

Main Findings

Although previous studies have recently found support for the DWE AC structure (Hoffman & Meade, 2012; Putka & Hoffman, 2013), the present study found mixed support for the DWE structure across two samples. In the first sample, both the two factor DWE model (Table 1, Model 2) and the one factor DWE model (Table 1, Model 3) fit the data significantly worse than the target three DWE model, which provided initial support for the distinguishability of multiple dimension within exercises. Notably, it was also found that, due to their strong intercorrelations, the dimensions within the IB exercise could be collapsed into a single IB exercise factor.

In the second sample, there was less evidence for distinguishable dimensions within exercises. As the CFA results demonstrated, Sample 2 was largely dominated by exercise variance and before the final four exercise factors model was tested results from earlier DWE models found correlations greater than .95 among different dimensions within the same exercise. In other words, within an exercise an assessee's behavior could not be differentiated by dimension. It is notable that Sample 2 was obtained from an AC designed under the task-based framework (see Jackson, 2012), which is an AC framework developed to minimize the impact of dimensions and instead focus on exercise performance (Jackson et al., 2010; Thoresen & Thoresen, 2012).

Although the DWE structure received only mixed support we continued with our analyses examining the correlations between the DWE and the FFM and GMA variables. Regarding the first sample, despite evidence for dimensions within exercise in the factor analytic results, the DWE correlations with personality did not differ significantly from other dimensions within the same exercise. Moreover, the DWE factors that did significantly differ from other dimensions within the same exercise did not significantly differ from measures of that dimension within other exercises, the overall exercise score, or the overall dimension score. For instance, the correlation between the drive dimension when measured by the LGD and openness (see Table 11) significantly differed from the correlation between the relational skills dimension when measured by the LGD and openness, but did not significantly differ from other DWE correlations with openness, the correlation between the overall LGD score and openness, or the correlation between the overall drive score and openness. Moreover, this same trend can be found in the results from the second sample as well. Thus, contrary to our hypotheses, DWE

scores did not provide stronger evidence for the construct validity of ACs than what has already been provided by other score interpretations such as overall dimensions (Meriac et al., 2014).

In fact, the relationships between the DWE scores and the FFM and GMA variables were consistent with those found between overall dimension scores and the FFM and GMA variables (Meriac et al., 2014) and overall exercise scores and the FFM and GMA variables (Hoffman et al., Under Review). For instance, in both samples, GMA was generally the strongest correlate of the DWE scores, overall exercise scores, and overall dimension scores with magnitudes similar to those reported in previous meta-analytic reviews of AC construct validity (Hoffman et al., Under Review; Meriac et al., 2008; Meriac et al. 2014). Further, across both samples, extraversion and openness were generally the next strongest, positive correlates of the DWE scores, overall exercise scores, and overall dimension scores, again reflecting what has been found in large scale AC reviews. Conversely, the personality traits indicative of getting along (Hogan & Holland, 2003) were generally weakly and non-significantly related to all of the AC scores or significantly and negatively related to the AC scores. These results, along with those found in previous AC reviews (Hoffman et al., Under Review; Meriac et al., 2014), suggest a slight trend towards rating disagreeable managers as higher performers. Thus, it seems that regardless of the scoring method, ACs are prone to promote candidates who are typically more intelligent, less agreeable, and inclined towards more getting ahead type behaviors.

Implications

Continuing the discussion of the findings, several overarching implications for AC research and practice will be discussed. First, these findings help to further develop the theoretical role that TAT plays in helping to understand how exercises elicit dimensional behaviors. Second, they help to inform research on the interpretation of DWE scores in relation

to overall exercise and overall dimension scores, which has implications for how AC feedback is delivered.

First, after 60 years of research much is still not known about how AC exercises elicit dimension relevant behaviors (Brummel, Rupp, & Spain, 2009; Howard, 2009; Lievens, Tett, & Schleicher, 2009). That is, TAT argues that variance on AC dimensions within an exercise can be affected by the trait cues provided by an exercise and the situational strength of that exercise (Lievens et al., 2009). As of yet, research has been unable to empirically determine which trait cues are responsible for eliciting a given dimensional behavior within an exercise and, further, how the situational strength of an exercise affects which behaviors are elicited (Lievens et al., 2009). However, it is possible that the DWE model could allow research to empirically confirm the role of both these exercise characteristics.

Regarding trait cues, research could test whether manipulating certain cues within a given exercise leads to different, predictable dimension structures within that exercise. For example, an AC could be designed to contain two RP exercises with one designed to provide cues that elicit relational, drive, and administrative behaviors, whereas the other one could be designed to elicit only relational and administrative behaviors. Following this, CFA could be used to fit the DWE model to the data and determine if the within exercise dimensions predictably differed across the two RPs. It is likely we saw this very same effect in the IB exercise of the first sample. That is, because the IB exercise requires assessees to perform a limited set of administrative tasks and does not provide trait cues for interpersonal behaviors it is unlikely that interpersonal dimension variance could be partitioned from administrative or performance dimension variance, which is what we found with our results.

As for situational strength, it has been defined as the "implicit or explicit cues provided by external entities regarding the desirability of potential behaviors" (Meyer, Dalal, & Hermida, 2010, p. 121). In the context of ACs, this is the idea that exercises (situations) differ in the extent to which there is uncertainty surrounding the required dimensional behaviors. As the situational strength of an exercise increases it begins this uncertainty decreases and the exercise begins to restrict the variety of potential dimensional behaviors that an assessee could exhibit (Lievens et al., 2008; Tett & Burnett, 2003; Tett, Simonet, Walser, & Brown, 2014). Empirically, as the situational strength of an exercise increases it should become harder to distinguish between different dimensions within a given exercise, which would lead to a general exercise factor. Following the same example from above, AC researchers could manipulate the clarity of relevant trait cues (Kleinmann, Kuptch, & Koller, 1996) for one RP and not the other and apply the DWE model to examine how the within exercise dimension structure differs across the two RPs.

So, by increasing an exercises' situational strength one should start to see more situational variance and less within exercise dimension variance, which would result in dimensions becoming less distinguishable within an exercise (Lievens et al., 2009). Correspondingly, manipulating an exercises' trait cues could potentially produce new kinds of within exercise dimension variance. For example, by adding interpersonal cues in an exercise that previously only provided administrative cues, the within exercise factor structure should change from a single administrative/performance factor to a two factor structure characterized by an administrative/performance factor and an interpersonal factor (Lievens et al., 2008; Tett & Burnett, 2003).

Beyond providing new ways to empirically examine AC exercises, the results of this study show that DWE scores can also be used to deliver more concise AC performance feedback.

That is, recent studies have argued that DWE scores should be incorporated into AC feedback (Hoffman & Meade, 2012). However, it was not known whether such feedback would lose its interpretive feedback. The results from both samples show that although DWEs are weakly correlated with the FFM and GMA variables, the correlations are similar to those seen with overall exercise and dimension scores. Thus, it is unlikely that the use of DWEs would result in the loss of interpretive value. However, it remains to be seen whether the information gains afforded by the DWE scores is worth the added complexity.

Limitations and Future Directions

As there are with any study, this study has a few notable limitations. First, because this study relied on AC samples obtained from two independent external sources, the ACs did not use the same exercises, nor did they measure the same narrow dimensions. As a result, the narrow dimensions that were recoded into the three broad AC dimensions differ by sample. Moreover, given that the Sample 2 CFA results differ from both the Sample 1 results and the results found in Hoffman and Meade (2012) it is unclear as to how our results generalize to the broader AC literature. Furthermore, both AC samples used personality scales that did not easily correspond to the FFM domain. Although prior research and the inventory manuals provided directions on how to recode the factors to fit within the FFM framework it is possible that different results would be obtained if a personality scale designed to measure the FFM was used.

Despite these limitations, the mixed-model approach to ACs is relatively new and as such it offers AC researchers a variety of substantive and methodological research directions. However, rather than discuss both the methodological and substantive research directions separately, we will discuss them together which allows us to show how different methodological approaches answer different substantive questions. First, research into the validity of

dimensional behaviors within an exercise can continue to use CFA; however it might also be interesting to include a measure of assessee's job performance to examine if the criterion-related validity of dimensions changes with exercises. Moreover, given previous research on the criterion-related validity of AC performance ratings, it is likely that the correlations between DWE factors and job performance will be larger than the correlations we found between DWE factors and the FFM. Additionally, AC research should begin to explore how differences in the design of ACs such as dimension-based ACs versus task-based ACs affect the subsequent factor structure of the AC performance ratings.

Alternatively, AC research can begin to explore the use of cross-classified models, which are a general class of mixed-effects linear models that allow researchers to estimate how assessee × dimension, assessee × exercise, and assessee × dimension × exercise variance affects AC performance ratings (Judd, Westfall, & Kenny, 2012; Raudenbush & Bryk, 2002). Putka and Hoffman (2013) provided an initial examination of these variance components using a random-effects linear model, but future research should include substantive predictors as well. That is, it is possible that future research could include predictors such as the complexity of an AC exercise or its situational strength to account for exercise variance and, further, research can examine how these exercise characteristics interact with assessee traits such GMA to predict assessee × exercise AC performance variance. Furthermore, it is possible for researchers to examine those variance components using AC ratings that have been averaged across assessors. That is, as long as each (or most) AC dimension is measured across each AC exercise with multiple items than it is possible to estimate and predict the components of reliable variance that Putka and Hoffman (2013) examined. Moreover, research could also use cross-classified models to test whether the

effects of an exercise characteristic such as clarity of the provided trait-cues varies across exercises.

REFERENCES

Arthur, W., Jr., 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. doi: 10.1111/j.1744-6570.2003.tb00146.x

Arthur, W. Jr., Day, E. A., & Woehr, D. J. (2008). Mend it, don't end it: An alternate view of assessment center construct-related validity evidence. *Industrial and Organizational Psychology: Perspectives on Science and Practice, 1,* 105-111. doi: 10.1111/j.1754-9434.2007.00019.x

Barrick, M. R., & Mount, M. K. (1991). The Big Five personality dimensions and job performance: A meta-analysis. *Personnel Psychology, 44,* 1-26. doi: 10.1111/j.1744-6570.1991.tb00688.x

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin, 107,* 238-246.

Blickle, G., Momm, T., Schneider, P. B., Gansen, D., & Kramer, J. (2009). Does acquisitive self-presentation in personality self-ratings enhance validity? Evidence from two experimental field studies. *International Journal of Selection and Assessment, 17,* 142-153. doi: 10.1111/j.1468-2389.2009.00458.x

Bollen, K. A. (1989). *Structural equations with latent variables.* New York: John Wiley & Sons.

Borman, W. C., & Brush, D. H. (1993). More progress toward a taxonomy of managerial performance requirements. *Human Performance, 6,* 1-21. doi: 10.1207/s15327043hup0601_1

Borman, W. C., & Motowidlo, S. J. (1997). Task performance and contextual performance: The meaning for personnel selection research. *Human Performance, 10,* 99-109.

Bowler, M. C., & Woehr, D. J. (2009). Assessment center construct-related validity: Stepping beyond the MTMM matrix. *Journal of Vocational Behavior, 75,* 173-182. doi: 10.1016/j.jvb.2009.03.008

Brummel, B. J., Rupp, D. E., & Spain, S. M. (2009). Constructing parallel simulation exercises for assessment centers and other forms of behavioral assessment. *Personnel Psychology, 62,* 137-170.

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

Chiaburu, D. S., Oh, I. S., Berry, C. M., Li, N., & Gardner, R. G. (2011). The five-factor model of personality traits and organizational citizenship behaviors: A meta-analysis. *Journal of Applied Psychology*, *96,* 1140. doi: 10.1037/a0024004

Christiansen, N. D., Hoffman, B. J., Lievens, F., & Speer, A. B. (2013). Assessment centers and the measurement of personality. In N. D. Christiansen & R. P. Tett (Eds.), *Handbook of Personality at Work* (pp. 477-497). New York, NY: Routledge.

DeYoung, C. G. (2006). Higher-order factors of the big five in a multi-informant sample. *Journal of Personality and Social Psychology, 91,* 1138-1151. doi: 10.1037/0022-3514.91.6.1138

Dierdorff, E. C., Rubin, R. S., & Bachrach, D. G. (2012). Role expectations as antecedents of citizenship and the moderating effects of work context. *Journal of Management, 38,* 573-598. doi: 10.1177/0149206309359199

Dierdorff, E. C., Rubin, R. S., & Morgeson, F. P. (2009). The milieu of managerial work: An integrative framework linking work context to role requirements. *Journal of Applied Psychology, 94,* 972-988. doi: 10.1037/a0015456

Digman, J. M. (1997). Higher-order factors of the big five. *Journal of Personality and Social Psychology, 73,* 1246-1256.

Gaugler, B.B., Rosenthal, D.B., Thornton, G.C., III, & Bentson, C. (1987). Meta-analysis of assessment center validity. *Journal of Applied Psychology, 72*, 493-511. doi:10.1037/0021-9010.72.3.493

Gough, H. G., & Bradley, P. (1996). *California psychological inventory manual.* Palo Alto, CA: Counseling Psychologist's Press.

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.

Hoffman, B. J., & Meade, A. (2012). Alternate approaches to understanding the psychometric properties of assessment centers: An analysis of the structure and equivalence of exercise ratings. *International Journal of Selection and Assessment, 20,* 82-97. doi: 10.1111/j.1468-2389.2012.00581.x

Hoffman, B. J. (2012). Exercises, dimensions, and the great battle of Lilliput: A call for multifaceted interpretation of assessment center performance. In D. J. R. Jackson, C. E. Lance, & B. J. Hoffman (Eds.), *The Psychology of Assessment Centers* (pp. 281-306). New York: Routledge.

Hoffman, B. J., Melchers, K. G., Blair, C. A., Kleinmann, M., Ladd, R. T. (2011). Exercises and dimensions are the currency of assessment centers. *Personnel Psychology, 64,* 351-395. doi: 10.1111/j.1744-6570.2011.01213.x

Hoffman, B. J., Monahan, E. L., Kennedy, C., LoPilato, A. C., Sutton, A., & Lance, C. E. (under review). A review of the content, criterion-related, and construct-related validity of assessment center exercises.

Hoffman, B. J., & Woehr, D. J. (2009). Disentangling the meaning of multisource performance rating source and dimension factors. *Personnel Psychology, 62,* 735-765. doi: 10.1111/j.1744-6570.2009.01156.x

Hogan, R., & Blickle, G. (2013). Socioanalytic Theory. In N. D. Christiansen & R. P. Tett (Eds.), *Handbook of Personality at Work* (pp. 53-70). New York, NY: Routledge.

Hogan, J., & Holland, B. (2003). Using theory to evaluate personality and job-performance relations: A socioanalytic perspective. *Journal of Applied Psychology, 88,* 100-112. doi: 10.1037/0021-9010.88.1.100

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. doi: 10.1111/j.1754-9434.2007.00018.x

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Convential criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6,* 1-55. doi: 10.1080/10705519909540118

Jackson, D. J. R., Lance, C. E., & Hoffman, B. J. (2012). *The psychology of assessment centers*. New York, NY: Routledge.

Jackson, D. J. R., Stillman, J. A., & Englert, P. (2010). Task-based assessment centers: Empirical support for a systems model. *International Journal of Selection and Assessment, 18,* 141-154. doi: 10.1111/j.1468-2389.2010.00496.x

Judd, C. M., Westfall, J., & Kenny D. A. (2012). Treating stimuli as a random factor in social psychology: A new and comprehensive solution to a pervasive but largely ignored problem. *Journal of Personality and Social Psychology, 103,* 54-69.

Kleinmann, M., Kuptsch, C., & Koller, O. (1996). Transparency: A necessary requirement for the construct validity of assessment centres. *Applied Psychology: An International Review, 45,* 67-84.

Lance, C. E. (2008). Why assessment centers do not work the way they're supposed to. *Industrial and Organizational Psychology: Perspectives on Science and Practice, 1,* 87-100. doi:10.1111/j.1754-9434.2007.00017.x

Lance, C. E. (2012) Research into task-based assessment centers. In D. J. R. Jackson, C. E. Lance, & B. J. Hoffman (Eds.), *The psychology of assessment centers* (pp. 218-233). New York, NY: Routledge.

Lance, C. E., Baranik, L. E., Lau, A. R., & Scharlau, E.A. (2009). If it ain't trait it must be method: (Mis)application of the multitrait-multimethod methodology in organizational research. In C. E. Lance & R. J. Vandenberg (Eds.), *Statistical and methodological myths and urban legends: Doctrine, verity, and fable in organizational and social science* (pp. 337-360). New York, NY: Routledge.

Lance, C. E., Foster, M. R., Gentry, W. A., & Thoresen, J. D. (2004). Assessor cognitive processes in an operational assessment center. *Journal of Applied Psychology, 89,* 22-35. doi: 10.1037/0021-9010.89.1.22

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., 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 dimensions ratings. *Journal of Applied Psychology, 89,* 377-385. doi: 10.1037/0021-9010.89.2.377

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.

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

Lievens, F., & Christiansen, C. (2012). Core debates in assessment center research: Dimensions 'versus' tasks. In D. Jackson, C. Lance, & B. Hoffman (Eds.) *The psychology of assessment centers* (pp. 68-94). New York, NY: Routledge.

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. doi: 10.1037/0021-9010.91.2.247

Lievens, F., De Koster, L., & Schollaert, E. (2008). Current theory and practice of assessment centres: The importance of trait activation. In S. Cartwright & C.L. Cooper (Eds.) *The oxford handbook of personnel psychology* (pp. 215-233). New York, NY: Oxford University Press.

Lievens, F., Tett, R. P., & Schleicher, D. J. (2009). Assessment centers at the crossroads: Toward a reconceptualization of assessment center exercises. In J. J. Martocchio & H. Liao (Eds.), *Research in Personnel and Human Resources Management* (pp. 99-152). Bingley: JAI Press.

Marsh, H. W. (1994). Confirmatory factor analysis models of factorial invariance: A multifaceted approach. *Structural Equation Modeling: A Multidisciplinary Journal, 1,* 5-34. doi: 0.1080/10705519409539960

Meng, X., Rosenthal, R., Rubin, D. B. (1992). Comparing correlated correlation coefficients. *Psychological Bulletin, 111,* 172-175. doi: 10.1037/0033-2909.111.1.172

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. doi: 10.1037/0021-9010.93.5.1042

Meyer, R. D., Dalal, R. S., & Hermida, R. (2010). A review and synthesis of situational strength in the organizational science. *Journal of Management, 36,* 121-140.

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

Monahan, E. L., Hoffman, B. J., Lance, C. E., Jackson, D. R. J., & Foster, M. R. (2013). Now you see them, now you do not: The influence of indicator-factor ratio on support for assessment center dimensions. *Personnel Psychology, 66,* 1009-1047. doi: 10.1111/peps.12049

Mumford, M. A., Zaccaro, S. J., Harding, F. D., Jacobs, T. O., & Fleishman, E. A. (2000). Leadership skills for a changing world: Solving complex social problems. *The Leadership Quarterly, 11,* 11-35. doi: http://dx.doi.org/10.1016/S1048-9843(99)00041-7

Muthén, L. K., & Muthén, B. O. (2012). *Mplus user's guide* (7th ed.). Los Angeles, CA: Muthén & Muthén.

Organ, D. W. (1988). *Organizational citizenship behavior: The good soldier syndrome.* Lexington, MA: Lexington Books.

Psytech International. (1991). *General, critical, & graduate test battery: The technical manual.* Pulloxhill, Bedfordshire, UK: Published Psychometrics Ltd.

Psytech International. (2002). *15FQ + fifteen factor questionnaire technical manual.* Pulloxhill, Bedfordshire, UK: Published Psychometrics Ltd.

Putka, D. J., & Hoffman, B. J. (2013). Clarifying the contribution of assessee-, dimension-, exercise-, and assessor-related effects to reliable and unreliable variance in assessment center ratings. *Journal of Applied Psychology, 98,* 114-133. doi: 10.1037/a0030887

Raudenbush, S. W. & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods.* Thousand Oaks, CA: Sage.

Rupp, D. E., Gibbons, A. M., Baldwin, A. M., Snyder, L. A., Spain, S. M., Woo, S. E., Brummel, B. J., Sims, C. S., & Kim, M. (2006). An initial validation of developmental assessment centers as accurate assessments and effective training interventions. *The Psychologist-Manager Journal, 9,* 171-200. doi: 10.1207/s15503461tpmj0902_7

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. doi: 10.1111/j.1754-9434.2007.00021.x

Sackett, P. R., & Dreher, G.F. (1982). Constructs and assessment center dimensions: Some troubling empirical findings. *Journal of Applied Psychology, 67,* 401-410. doi: 10.1037/0021-9010.67.4.401

Steiger, J. H. (1990). Structural model evaluation and modification: An interval estimation approach. *Multivariate Behavioral Research, 25,* 173-180. doi: 10.1207/s15327906mbr2502_4

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

Tett, R. P., & Guterman, H. A. (2000). Situation trait relevance, trait expression, and cross-situational consistency: Testing a principle of trait activation. *Journal of Research in Personality, 34,* 397-423. doi: http://dx.doi.org/10.1006/jrpe.2000.2292

Tett, R. P., Simonet, D. V., Walser, B., & Brown, C. (2013). Trait activation theory: Applications, developments, and implications for person-workplace fit. In N. D. Christiansen & R. P. Tett (Eds.), *Handbook of Personality at Work* (pp. 71-100). New York, NY: Routledge.

Thoresen, C. J. & Thoresen, J. D. (2012). How to design and implement a task-based assessment center. In D. J. R. Jackson, C. E. Lance, & B. J. Hoffman (Eds.), *The Psychology of Assessment Centers* (pp. 190-217). New York: Routledge.

Thornton, G. C., III, & Mueller-Hanson, R. A. (2004). *Developing organizational simulations: A guide for practitioners and students.* Mahwah, NJ: Lawrence Erlbaum Associates.

Viswesvaran, C., Schmidt, F. L., & Ones, D. S. (2005). Is there a general factor in ratings of job performance? A meta-analytic framework for disentangling substantive and error

influences. *Journal of Applied Psychology, 90,* 108-131. doi: 10.1037/0021-9010.90.1.108

Watson, G. , & Glaser, E. (1980). *Manual for the Watson Glaser critical thinking appraisal.* New York, NY: Psychological Corporation.

Weiss, H. M. & Adler, S. (1984). Personality and organizational behavior. *Research in Organizational Behavior, 6,* 1-50.

Woehr, D. J., Meriac, J. P., & Bowler, M. C. (2012). Methods and data analysis for assessment centers. In D. Jackson, C. Lance, & B. Hoffman (Eds.) *The psychology of assessment centers* (pp. 45-67). New York, NY: Routledge.

APPENDIX A

SAMPLE 1 AND 2 DIMENSION LABELS

Sample 1 Dimension Labels

Dimension Label	Corresponding Seven-Factor Model Label	Corresponding Three-Factor Model Label
Analyze Issues	Problem Solving	Administrative
Sound Judgment	Problem Solving	Administrative
Think Strategically	Problem Solving	Administrative
Establish Plans	Planning and Organizing	Administrative
Manage Execution	Planning and Organizing	Administrative
Lead Courageously	Influencing Others	Relational
Influence Others	Influencing Others	Relational
Coaching and Development	Planning and Organizing	Administrative
Foster Teamwork	Communication	Relational
Champion Change	Drive	Drive
Build Relationships	Consideration/Awareness of Others	Relational
Manage Disagreements	Communication	Relational
Fostering Open Communication	Communication	Relational
Show Drive and Development	Drive	Drive
Customer Focus	Problem Solving	Administrative
Demonstrate Adaptability	Stress Tolerance	Relational

Sample 2 Dimension Labels

Dimension Label	Corresponding Seven-Factor Model Label	Corresponding Three-Factor Model Label
Business Acumen	Problem Solving	Administrative
Developing Direct Reports	Planning and Organizing	Administrative
Informing	Planning and Organizing	Administrative
Managing and Measuring Work	Problem Solving	Administrative
Listening	Consideration/Awareness of Others	Relational
Confronting Direct Reports	Influencing Others	Relational
Drive for Results	Drive	Drive
Customer Focus	Drive	Drive
Managing Others	Planning and Organizing	Administrative
Open Minded	Problem Solving	Administrative
Conflict Manager	Consideration/Awareness of Others	Relational
Cool Transactor	Stress Tolerance	Relational
Agile Communicator	Consideration/Awareness of Others	Relational
Visioning	Influencing Others	Relational
Critical Thinking	Problem Solving	Administrative