Ashley A. Barbee, A MONTE CARLO EVALUATION OF THE APPLICATION OF VARIANCE PARTITIONING TO THE ASSESSMENT OF CONSTRUCT-RELATED VALIDITY (Under the direction of Dr. Mark C. Bowler) Department of Psychology

The present study presents a Monte Carlo evaluation of the application of variance partitioning to the assessment of the construct-related validity of assessment center (AC) post exercise dimension ratings (PEDRs). Data was produced by creating sixteen population models representing a variety of AC models by varying dimension factor loadings, exercise factor loadings, dimension intercorrelations, and exercise intercorrelations. Analyses demonstrated that variance partitioning differentiated among all sixteen varieties of AC models. Variance partitioning also detected other sources of variance including person effects, person by dimension effects, and person by exercise effects. These findings suggest that variance partitioning may be a more appropriate method for analyzing AC multitrait-multimethod (MTMM) data instead of the traditional confirmatory factor analysis (CFA) method.

A MONTE CARLO EVALUATION OF THE APPLICATION OF VARIANCE PARTITIONING TO THE ASSESSMENT OF CONSTRUCT-RELATED VALIDITY

by

Ashley A. Barbee

APPROVED BY:

DIRECTOR OF THESIS	______________________________ Mark C. Bowler, Ph.D.
COMMITTEE MEMBER	______________________________ Jennifer L. Bowler, Ph.D.
COMMITTEE MEMBER	______________________________ John G. Cope, Ph.D.
CHAIR OF THE DEPARTMENT OF PSYCHOLOGY	______________________________ Kathleen A. Row, Ph.D.
DEAN OF THE GRADUATE SCHOOL	______________________________ Paul J. Gemperline, Ph.D.

A MONTE CARLO EVALUATION OF THE APPLICATION OF VARIANCE PARTITIONING TO THE ASSESSMENT OF CONSTRUCT-RELATED VALIDITY

A Thesis

Presented to

the Faculty of the Department of Psychology

East Carolina University

In Partial Fulfillment

of the Requirements for the Degree

Master of Arts in Psychology

by

Ashley A. Barbee

April, 2012

Acknowledgements

First, I would like to thank my thesis committee chair, Dr. Mark Bowler, for his constant support, feedback, and words of encouragement. His knowledge, advice, and sense of humor have made this project an exciting learning experience. I would also like to thank my committee members, Dr. Jennifer Bowler and Dr. John Cope, for their invaluable time and feedback. Lastly, I would like to thank my mother, Julie Cole, for her patience and never-ending support throughout my entire educational career.

Table of Contents

List of Tables

CHAPTER 1: INTRODUCTION

Over the past 50 years, assessment centers (ACs) have emerged as one of the most popular tools for evaluating individual differences related to managerial performance (Chen, 2006; Joiner, 2002; Spychalski, Quinones, Gaugler, & Pohley, 1999). Designed for use with both employee selection and development, ACs are used to evaluate an individual's performance on a set of job-related dimensions via using multiple high-fidelity situational exercises. Despite their popularity, as well as their fundamental content validity (Binning & Barrett, 1989; Schmitt & Chan, 1998) and demonstrated criterion-related validity (Arthur, Day, McNelly, & Edens, 2003), the construct-related validity of AC post-exercise dimension ratings (PEDRs) continues to be problematic (cf. Lance, Foster, Gentry, & Thoresen, 2004; Lievens & Conway, 2001). Specifically, the prevailing view of the nature of AC PEDRs, the fundamental measuring block of ACs, is that they substantially reflect the effect of the exercises from which they are obtained rather than cross-exercise stability in candidate behavior on the dimensions being assessed (cf. Bycio, Alvarees, & Hahn, 1987; Fleenor, 1996; Lance, 2008; Lance, Lambert, Gewin, Lievens, & Conway, 2004; Lance et al., 2004; Lance, Newbolt, Gatewood, Foster, French, & Smith, 2000; Schneider & Schmidt, 1992). However, most studies that make this conclusion utilize analyses that are based on a confirmatory factor analysis (CFA) of a multitrait-multimethod (MTMM) matrix. Recent research has identified potential problems associated with this particular application of CFA (Lance, Woehr, & Meade, 2007; Lievens & Conway, 2001).

In their Monte Carlo study, Lance, Foster, and colleagues (2004) demonstrated that the application of CFA to the evaluation of MTMM data is problematic. Specifically, they noted that this analytical method frequently fails to generate an admissible solution. Moreover, when it does, even if the model is not the correct representation of the data (i.e., the sample data does not match the population data from which it was drawn), the fit statistics will indicate a good fit. Subsequently, they concluded that CFAs are problematic in that they can produce results that conflict with the true nature of the data. Furthermore, CFA examinations of AC PEDRs do not take into account the candidate being rated, the individuals conducting the rating, and various interactions (Bowler & Woehr, 2009). Thus, research into the internal structure of ACs should most likely not rely on CFAs to evaluate their construct-related validity of ACs.

Recently, Woehr, Putka, and Bowler (2011) have identified a novel method for assessing the construct-related validity of MTMM data. Specifically, the application of variance partitioning (i.e., G-theory), which involves modeling person, trait, and exercise effects, helps to circumvent the issues with CFA. Moreover, this method has direct analogues with the facets of construct-related originally identified by Campbell and Fiske (1959). This method will be utilized to assess the construct-related validity of AC PEDRs. The present study will evaluate a series of Monte Carlo simulations representing a variety AC models. Specifically, analyses will focus on whether variance partitioning differentiates among varieties of AC models and if variance partitioning diagnoses additional aspects of ACs beyond dimension and exercise effects.

Assessment Center Design

Assessment centers are used for both selection and career development purposes (Thornton, 1992). When utilized for selection, ACs are used to determine which applicants have the necessary knowledge, skills, and abilities (KSAs) to be successful in the relevant position. When utilized for career development, an AC is used to determine behaviors that a participant does well and which behaviors need improvement, with training being based on the later. Regardless, all assessment centers are comprised of a varying series of high-fidelity exercises, such as an in-basket, a leaderless group discussion, a one-on-one role play, and a case analysis.

Applicants' performances on these exercises are observed and rated independently by trained assessors. These exercises are intended to serve as stimulus materials for the measurement of the relevant job-related skills. Dimensions vary widely across ACs, from oral communication to decisiveness (Thornton, 1992). As noted by Arthur, Day, et al. (2003) there are over 138 dimensions in the AC literature with a typical AC being comprised of eleven dimensions and five exercises (Woehr & Arthur, 2003). Ideally, these dimensions are the foundation of AC functioning. Specifically, ACs are designed to evaluate the participant's standing on a dimension and are based on the assumption that his or her performance on the dimension will be stable across exercises (Lievens & Conway, 2001).

A typical assessment center involves several assessors, comprised of HR staff members and managers, and assessees who participate in the various exercises (Thornton & Mueller-Hanson, 2004). Prior to the assessment center, the assessors are thoroughly trained in the assessment process. During the AC, the assessors observe and rate participant behaviors in various situation exercises (Thornton & Mueller-Hanson, 2004). The assessors rate each assessee on the specified dimensions after each exercise is completed. Once the assessment center is complete, the assessors meet and discuss their reports on each participant. The assessors then rate each assessee on each performance dimension (e.g. on a five or seven point scale). If the purpose of the AC was promotion, the assessors discuss each assessee's probability of success if given the promotion, and the assessors make a recommendation of which candidate they believe is suitable for the position. Typically, each participant is also provided a report of their performance on each specified dimension.

Convergent and Discriminant Validity

In order for AC PEDRs to be considered to be construct-valid, a multitrait-multimethod (MTMM) matrix needs to demonstrate convergent and discriminant validity. Evaluation of MTMM matrices is nothing new as it was first discussed by Campbell and Fiske (1959). In ACs, MTMM matrices are often used to demonstrate the relationship between dimension and exercise factors. It is assumed that dimension factors will be observed in multiple exercises. Large trait factor loadings indicate support for convergent validity and large trait correlations indicate a lack of discriminant validity. Specifically, trait factor loadings are comprised of dimension scores across multiple exercises. A typical MTMM matrix is comprised of three types of correlations: (a) correlations among PEDRs sharing the same dimension, but different exercises, (b) correlations among PEDRs sharing the same exercise, but different dimensions, and (c) correlations among PEDRs that share neither dimensions nor exercises (Woehr et al., 2011). The purpose of a CFA is to evaluate the fit between the observed correlations among the PEDRs and a reproduced correlation matrix, and provide parameter estimates. However, recent meta-analyses have demonstrated a problem using CFA to determine model fit.

Meta-analyses of Model Fit using CFA

Lievens and Conway (2001). In their review of AC construct-related validity, Lievens and Conway (2001) reanalyzed 34 MTMM matrices from 24 assessment centers. Their analyses focused on fitting six different models to each of the 34 MTMM matrices using a CFA: (1) A correlated dimension model, (2) a correlated exercise model, (3) a correlated dimension-correlated exercise model, (4) a single dimension-correlated exercise model, (5) a direct product model, and (6) a correlated uniqueness model. A correlated dimension model reflects an AC with PEDRs that are solely the function of the dimensions being rated, whereas a correlated exercise model reflects an AC with PEDRs that are only a function of the exercises used to evaluate the dimensions. The correlated dimension-correlated exercise model places emphasis on both the dimensions being rated and exercise in which they are being rated. The single dimension-correlated exercise model reflects an AC with PEDRs that are comprised of a single dimension factor (e.g. a “g” factor or a “person” factor) as well as the exercises. The direct product model and the correlated uniqueness model are both statistical variations of the correlated dimension-correlated exercise model. In the direct product model the correlations between PEDRs are a multiplicative function between dimensions and exercises, and in the correlated uniqueness model the exercise effects are not explicitly differentiated from the correlations among the uniquenesses.

Results from the Lievens and Conway (2001) analyses indicated the dimension-only model and the exercise-only model showed poor fit. The percentage of matrices that demonstrated an acceptable fit for these two models were 3% and 29% respectively. The single dimension-correlated exercise model performed somewhat more favorably, producing an acceptable fit for 53% of the matrices. In contrast, the correlated dimension-correlated exercise model and the direct product model demonstrated acceptable fit with 85% and 81%, respectively. However, the correlated uniqueness model demonstrated the best fit, fitting 88% of the MTMM matrices. Based on these results, Lievens and Conway concluded that a model comprised of correlated dimensions and exercises modeled as correlated uniqueness was the most appropriate to use when analyzing AC MTMM matrices. Following this, Lievens and Conway estimated the impact dimension and exercise factors had on each model. Overall, they noted that the mean proportion of variance that was attributable to dimensions was .34 and the mean proportion of variance attributable to exercises was .34. Lievens and Conway also noted that these values varied greatly (from .17 to .62 for dimensions and .07 to .69 for exercises) with several models having high intercorrelated dimensions (e.g., .71). Nonetheless, they concluded that dimensions have more of an effect on AC ratings than was previously suspected but that their study fell short of demonstrating that dimensions actually do have a greater impact than exercises.

Lance, Foster, et al. (2004). Due to several problematic statistical issues with the Lievens and Conway (2001) results – specifically that the correlated uniqueness model used by Lievens and Conway can often lead to overestimated dimension effects – Lance, Foster, and colleagues (2004) replicated the study conducted with a different set of CFA models: (1) a correlated dimension, correlated exercise (CDCE) model, (2) a single dimension-correlated exercise model (1DCE), and (3) a correlated exercise model. Two additional models were also estimated in which "exercise effects were modeled as covariances among uniqueness for PEDRs measured in the same exercise" (pg. 379). These models were: (1) a model with correlated dimensions and correlated uniqueness (CDCU), and (2) a model with one dimension and correlated uniqueness (1DCU).

Results indicated that the CDCE model was a poor fit for all data sets used except for two. Both the correlated exercises model and the 1DCE model provided good fits, with the 1DCE model providing a better fit for the data overall (Lance, Foster, et al., 2004). Consistent with the Lievens and Conway (2001) findings, bias in dimension factor loadings for the CDCE and CDCU models were present when exercise factor correlations and exercise factor loadings were not zero. In addition, more bias in dimension effects were found in the 1DCU model compared to the 1DCE model (Lance, Foster, et al., 2004). The correlated uniqueness models yield upward-biased estimates of dimension effects and are potentially a good fit for evaluating AC PEDRs. Furthermore, Lance et al.'s study provided more evidence that exercise factors explain more variance in PEDRs than dimension factors. Based on this finding, Lance et al.

assert that ACs lack construct validity. Since AC PEDRs are dominated more by exercise effects, candidate traits (dimension factors) are not rated the same across exercises.

Instead of AC performance being represented as scores of participants' traits across various exercises, Lance, Foster, et al. (2004) proposed the idea that AC performance is comprised of two factors: (a) a general yet situation-specific performance representing performance on each exercise task and, (b) a stable overall performance factor that is consistent across exercises. Lance et al. also questioned the current method for providing performance feedback in assessment centers. Currently, feedback is based on dimension factors yet PEDRs are substantially influenced by exercise effects instead of dimension effects. Researchers instead proposed that developmental feedback with a focus on exercise performance may be more appropriate.

Bowler and Woehr (2006). With the conflicting viewpoints of Lievens & Conway (2001) and Lance, Foster, et al. (2004), Bowler and Woehr attempted to provide additional evidence on the impact of exercise and dimensions ratings. Specifically, they meta-analytically combined the previously analyzed MTMM matrices to form a single summary matrix. The single summary matrix was then evaluated on model fit and the impact of dimension and exercise effects. Bowler and Woehr also examined the impact of specific dimension and exercise factors on AC ratings as opposed to combining results across various dimension constructs and exercise types. Specifically, their analyses indicated that specific dimensions (e.g. communication) had higher construct-related validities than other dimensions. Similarly, specific exercises (e.g. in-basket) had higher construct-related validities than other exercises.

Twenty-four studies resulting in 35 MTMM matrices were used and analyzed for fit on six different models (Bowler & Woehr, 2006). The six models were (a) a model with six correlated dimensions, (b) a model with six correlated exercise factors, (c) a model with one dimension factor and 6 uncorrelated exercise factors, (d) a model with one dimension factor and six correlated exercise factors (1DCE), (e) a model with 6 correlated dimension factors and 6 uncorrelated exercise factors, and (f) six correlated dimension factors and six correlated exercise factors (CDCE). Results showed that 1DCE model and the CDCE model both failed to converge. These models were modified because of poor parameter estimates. After modification, both models converged to a proper solution.

Based on fit values, all of the models fit the data well except for the dimension-only model (Bowler & Woehr, 2006). The modified CDCE model had the best fit value and was the best representation of the data. The CDCE model was then analyzed for the impact of dimension and exercise ratings. Results indicated that both dimension and exercise factors contribute to PEDRs. Across all models, Bowler and Woehr discovered that dimension factors accounted for less variance than exercise factors. However, the variance accounted for by dimension factors was higher than that found by Lance et al. (2004). Therefore, Bowler and Woehr asserted that both dimension and exercise factors contribute to AC PEDRs.

Overall, these three aforementioned reviews produced notably different results. Whereas results from the Lance, Foster, et al. (2004) analysis concluded that assessment center PEDRs are a function of exercises and not dimensions, Lievens and Conway (2001) and Bowler and Woehr (2006) produced results to the contrary. Specifically, both analyses concluded that exercise effects do not necessarily take precedence over dimension effects. Interestingly, the findings of all of these studies were based on the same statistical technique – a CFA of an MTMM matrix. Unfortunately, the suitability of this method has recently been called into question.

Generalizability Theory

Since previous research has demonstrated many problems with using confirmatory factor analysis to analyze AC MTMM data (e.g. Lance et al., 2007), Woehr, Putka, and Bowler (2011) discussed another method for modeling MTMM matrices, the Generalizability theory or variance partitioning. Three primary sources of variance are assessed using variance partitioning: the person being rated (p), the dimension being rated (t), and the exercise from which the rating was made (m). Written in variance partitioning shorthand, this concept can be described as p x t x m. Each person completes a number of trait measures and each trait measure is assessed using the same number of measurement methods. Subsequently, each person's observed score (X_{ptm}) on a given trait-method unit is a simple additive function:

$$X_{ptm} = \mu + v_p + v_t + v_m + v_{pt} + v_{pm} + v_{tm} + v_{ptm,r}, \quad (1)$$

where μ is the grand mean score across all persons, traits and methods; v_p is the person main effect and the expected value of a person's score across the population of traits and methods; v_t is the trait main effect and the expected value of the trait's effect across the population of persons and methods; v_m is the method main effect and the expected value of the method's effect across the population of persons and traits; v_{pt} is the person x trait interaction effect and reflects the differences in the ordering of the persons' expected scores (averaged over methods) across traits; v_{pm} is the person x method interaction effect and reflects differences in the ordering of the persons' expected scores (averaged over traits) across methods; v_{tm} is the trait x method interaction effect and reflects the differences in the ordering of traits' expected scores (averaged over persons) across methods; and lastly, $v_{ptm,r}$ is the remaining residual after accounting for all other effects in the model.

Assumptions

The assumptions of the above model are the same as the common random-effects ANOVA assumptions. All effects are assumed to be independent of each other with means of zero and variances of σ^2_p, σ^2_t, σ^2_m, σ^2_{pt}, σ^2_{pm}, σ^2_{tm}, and $\sigma^2_{ptm,r}$, respectively. These variance components are the focus of estimation in G-theory (Woehr et al., 2011). The expected total variance in scores across all p x t x m combinations in the population is expressed as:

$$\sigma^2_{\text{expected total}} = \sigma^2_p + \sigma^2_t + \sigma^2_m + \sigma^2_{pt} + \sigma^2_{pm} + \sigma^2_{tm} + \sigma^2_{ptm,r} \quad (2)$$

However, to model MTMM data, researchers are interested in expected observed variance in scores across persons, not expected total variance. The expected observed variance in scores across persons is expressed as:

$$\sigma^2_{\text{expected observed}} = \sigma^2_p + \sigma^2_{pt} + \sigma^2_{pm} + \sigma^2_{ptm,r} \quad (3)$$

Thus, the variance components involving solely traits and methods and the trait x method interaction effect are not included in calculating the expected observed variance since these components are constant across persons.

Relation to MTMM Modeling

Variance partitioning also relates to variance components in the MTMM matrix (Brennan, 2001). As noted by Woehr et al. (2011), relationships exist between variance components and the average correlations among trait-method units. Specifically,

Average monotrait-heteromethod (MTHM) $r = \sigma^2_p + \sigma^2_{pt}$, (4)

Average heterotrait-monomethod (HTMM) $r = \sigma^2_p + \sigma^2_{pm}$, and (5)

Average heterotrait-heteromethod (HTHM) $r = \sigma^2_p$. (6)

For example, σ^2_p can also be explained as shared variance, or covariance, among trait-method units that is not trait or method specific. In addition, σ^2_{pt} can also be explained as covariance among trait-method units that is specific to a given trait (Woehr et al., 2011). These formulas can be rearranged to generate standardized variance components as a function of average MTMM correlations:

$$\sigma^2_p = \text{Average HTHM } r, \quad (7)$$

$$\sigma^2_{pt} = \text{Average MTHM } r – \text{Average HTHM } r, \quad (8)$$

$$\sigma^2_{pm} = \text{Average HTMM } r – \text{Average HTHM } r, \quad (9)$$

$$\sigma^2_{ptm,r} = 1- \sigma^2_p + \sigma^2_{pt} + \sigma^2_{pm}. \quad (10)$$

Subsequently, the expected correlation between two different traits measured by a common method (e.g. HTMM) is general person effects. Method specific (σ^2_{pm}) and trait specific (σ^2_{pt}) effects can also be derived from expected heterotrait-heteromethod correlations.

Evidence for AC Construct-related Validity

In terms of AC construct-related validity, the person by dimension effect is most indicative of proper AC functioning. This effect represents variance in the PEDRs that is attributable to an individual's performance on a particular dimension irrespective of the exercise in which it was measured. Thus, this effect addresses the question of whether individual differences in dimensions are being measured. In contrast, the person by exercise effect is an important indicator of the situational specificity issue addressed by Lance, Newbolt, et al. (2000). This effect represents variance in the PEDRs that is attributable to an individual's performance in an exercise irrespective of the dimensions being measured. This addresses the question of whether individuals perform better in some situations than in others. Additionally, the dimension by exercise effect is an indicator of the dimension observability issue addressed by Lievens et al. (2006). This effect represents variance in the PEDRs that is attributable to the specific dimension being measured in a particular exercise. Thus, this effect addresses the question of whether some dimensions more are observable (i.e., easier to rate) in certain exercises than in others. In addition to the direct information provided by the second-order effects, their presence in the model helps to clarify the meaning of the first-order factors. The person effect addresses the question of whether a general performance factor impacts PEDRs irrespective of the person by dimension effect or the person by exercise effect. The dimension effect addresses the issue in which some dimensions are rated more leniently (or severely) than others. Finally, the exercise effect address the issue of whether PEDRs are higher or lower simply based on the exercise from which they are drawn (i.e., are some exercises more difficult than others?).

Woehr et al. (2011) further noted the direct relationship between these G-theory formulations and the traditional components of convergent and discriminant validity. Based on Campbell and Fiske's (1959) definitions, convergent validity is represented as consistency among different measures of a trait (i.e. σ^2_{pt}). This person x trait interaction represents common variance for measures that share common traits. Evidence for convergent validity is determined by the sum of σ^2_p and σ^2_{pt}, which is the proportion of expected observed variance in trait-method units attributable to (a) person main effects and (b) shared variance among persons specific to a given trait.

According to Campbell and Fiske (1959), three conditions must be met in order to determine discriminant validity. In Condition 1, person x trait interactions should be greater than zero. Specifically, multiple exercises are intended to determine a person's score on the specified dimensions. In Condition 2, person x trait interactions should be larger than person x method interactions. Ideally, PEDRs are determined by a person's dimension scores across exercises (person x trait interaction). The third condition cannot be easily tested and is assumed to be met under the G-theory model (Woehr et al., 2011).

Woehr and colleagues (2011) presented the flexibility of variance partitioning when analyzing MTMM matrices. This flexibility is useful in analyzing designs that involve multiple measurement entities such as using several exercises in ACs. Because of the convergence problems associated with MTMM CFA models, Woehr et al. proposed the need to refocus attention on variance partitioning models when analyzing MTMM data.

Utilization of Variance Partitioning

As previously noted, recent Monte Carlo research has identified potential problems associated with applying CFA models to MTMM data (Lance, Woehr, & Meade, 2007). In particular, Lance et al. (2007) presented a Monte Carlo investigation in which they simulated data representing three potential "true" models of AC PEDRs. The models utilized (1) a correlated dimensions and correlated exercises model (CDCE), (2) a single dimension, correlated exercises model (1DCE), and (3) an uncorrelated dimensions, correlated exercises, plus *g* model (UDCE+*g*). The CDCE model most closely matches the original conceptualization of the MTMM matrix (Kenny & Kashy, 1992). In this model, AC PEDRs are a function of both the dimensions being measured and the exercises used to measure the dimensions. The 1DCE model is based upon the findings of numerous empirical studies (e.g., Lance, Foster, et al., 2004; Lance et al., 2000; Schneider & Schmidt, 1992). In this model, AC PEDRs are primarily a function of the exercises utilized by the AC and a single general performance dimension. In contrast, the UDCE+*g* model holds AC PEDRs as a function of the dimensions, exercises, and a general performance factor. Thus, this model incorporates the basic features of both the CDCE model and the 1DCE model.

For each of these three population models, Lance et al. (2007) generated 500 sample MTMM matrices. These 1500 sample matrices were then separately analyzed via CFA with each population model being applied to each sample matrix. Thus, Lance et al. were able to directly evaluate whether a CFA could produce convergent and admissible solutions when the appropriate population model was applied to the respective sample matrix (e.g., does the CDCE model fit the CDCE data?). Moreover, they were able to evaluate what would happen when a particular population model was applied to a differing sample matrix (e.g., does the 1DCE model fit the CDCE data?).

In their simulations, when the CDCE model was applied to the CDCE data (i.e., the appropriate model for the data), the CFA converged to an admissible solution for only 61% of the matrices (Lance et al., 2007). However, when the 1DCE model was applied to the CDCE data (i.e., an improper model for the data), the CFA converged to an admissible solution for 100% of the matrices. Thus, it is highly likely that AC data conforming to the CDCE model would be misidentified as conforming to the 1DCE model due to a lack of admissible model convergence. Similarly, when the UDCE+*g* model was applied to the UDCE+*g* data (i.e., the proper model for the data), the CFA converged to an admissible solution for only 52% of the matrices. However, when the 1DCE model was applied to the UDCE+*g* data (i.e., an improper model for the data), the CFA converged to an admissible solution for 99% of the matrices. As with the CDCE data, there is a strong possibility that AC data that truly conforms to the UDCE+*g* model would be misidentified as conforming to the 1DCE model. In stark contrast, when the 1DCE model was applied to the 1DCE data (i.e., the proper model for the data), the CFA converged to an admissible solution for 100% of the matrices. Moreover, what is even more striking is that, for all models that reached an admissible solution, regardless of whether or not the fitted model matched the population model from which data was based, traditional model fit statistics (e.g., RMSEA, CFI, NNFI, ECVI, etc.) indicated that the model constituted a good fit for the data. Based on these results, Lance et al. concluded that CFA of an MTMM matrix is a problematic analytical technique due to its propensity to produce results that conflict with the true nature of the data.

In addition to the difficulties with model identification, Lance et al. (2007) utilized a relatively simple model of AC PEDRs. Their results were based on a simulated AC that measured five performance dimensions with three exercises. This approaches the minimum number required for a CFA of MTMM matrix (Marsh & Grayson, 1995). More complex models (those with more dimensions and exercises) are even more likely to experience estimation and convergence issues (Kenny & Kashy, 1992; Marsh, 1989), which is problematic because the vast majority of operational ACs include a relatively large number of dimensions and exercises. In their review of AC design features relevant to construct-related validity, Woehr and Arthur (2003) indicated than AC feature an average of almost 11 dimensions (*M* = 10.60, *SD* = 5.11) and 5 exercises (*M* = 4.79, *SD* = 1.47) per AC. Thus, if a more realistic number of dimensions and exercises were to be used, the models tested by Lance et al. would most likely have displayed far greater estimation problems. In fact, these issues are reflected in previous AC studies. As discussed previously, Lance, Lambert, et al. (2004) applied both the CDCE model and the 1DCE model to 39 unique AC MTMM matrices drawn from the literature. They found that the CDCE and 1DCE models converged to a proper solution in only 5% and 56% of the cases, respectively.

Beyond the serious analytic issues demonstrated by Lance et al. (2007), conceptual problems with the MTMM model of AC PEDRs also exist. Typical CFA examinations of AC PEDRs are limited in that they focus on the variance associated with AC dimensions and exercises while ignoring potential variation attributable to other sources such as the individual being rated (i.e., a person effect), the individuals assigning the ratings (i.e., a rater effect), and the numerous interactions. For example, the extent to which individuals demonstrate different patterns of performance across dimensions (i.e., a person by dimension interaction) is not included in the MTMM model. Similarly, the extent to which some dimensions might be more effectively assessed in particular exercises rather than in others (i.e., a dimension by exercise interaction) is not included. Thus, even to the extent that CFA approaches are able to provide appropriate estimates of dimension and exercise effects, these estimates are likely misleading in that they represent an overly simplistic model of AC PEDRs.

Purpose of Present Paper

The primary goal in the present paper will be to extend the Lance et al. (2007) simulations to evaluate the appropriateness of utilizing variance partitioning (Brennan, 1994; Cronbach, Glesser, Nanda, & Rajaratanm, 1972; Shavelson & Webb, 1991) for assessing the construct-related validity of AC PEDRs. Despite being recommended as an alternative method for dealing with AC data (Brannick, Michaels, & Baker, 1989), to date only three studies have utilized this approach (i.e., Arthur, Woehr, & Maldegen, 2000; Jackson, Stillman, & Atkins, 2005; Woehr et al., 2011). The present study will focus on whether variance partitioning gives accurate results. In addition, the present study will focus on diagnosing additional variance components such as the person effect and interaction effects. Specifically, (1) does variance partitioning differentiate between different models of AC functioning, and (2) is variance partitioning able to determine different aspects of AC functioning beyond dimension and exercise effects?

CHAPTER II: METHOD

Population Models

In order to evaluate the appropriateness of applying variance partitioning to AC PEDRs, sixteen population models were generated to serve as the basis of the simulations. In keeping with the Lance et al. (2007) study, the models represented an AC with five dimensions that are each measured in three exercises (i.e., a fully crossed design). However, unlike the Lance et al. simulations, the population models were chosen to represent a greater range of possible AC functioning than the models that are predominant in the literature (and often based on the fallacious CFA of an MTMM matrix).

Population models were based on varying four parameters: (a) dimension factor loadings, (b) exercise factor loadings, (c) dimension intercorrelations, and (d) exercise intercorrelations. Fifty percent of the dimension factor loadings had a loading of .70 and the other fifty percent had loadings of .35. These two variations represented optimal AC functioning (high dimension factor ratings) and poor AC functioning (low dimension factor ratings). Additionally, half of the exercise factor loadings had loadings of .70 and the other half had loadings of .35. These two variations also represented optimal AC functioning (low exercise factor loadings) and poor AC functioning (high exercise factor loadings). The dimension and exercise intercorrelations were also altered. Ideally, these factors should have relatively low intercorrelations. However, there are several instances of surprisingly high dimension factor intercorrelations in the AC literature (e.g., Bowler & Woehr, 2006). Subsequently, fifty percent of the dimension factor intercorrelations had loadings of .70 and the other fifty percent had loadings of .35. Furthermore, half of the exercise intercorrelations had factor loadings of .70 and the other half had loadings of .35.

Within these sixteen population models, one model was considered ideal. Based on the theoretical design of ACs (Bray & Grant, 1966; Thornton, 1992; Thornton & Mueller-Hanson, 2004), this model was comprised of high dimension factor loadings (M_a = .70), low exercise factor loadings (M_b = .35), low dimension intercorrelations (M_c = .35), and low exercise intercorrelations (M_d = .35). In contrast, although several models represent poor functioning, one was considered the worst case model. This model was comprised of low dimension factor loadings (M_a = .35), high exercise factor loadings (M_b = .70), high dimension intercorrelations (M_c = .70), and high exercise intercorrelations (M_d = .70). Specifically, the dimensions had very little effect on PEDRs and the dimension and exercises highly correlate with one another making it very difficult to discriminate between dimensions and exercises. Every possible combination was created among the four parameters. Table 1 provides the initial framework used to generate the sixteen models.

Table 1

Population Factor Loadings for Dimension and Exercise Factors and Intercorrelations

Facet	Population Factor Loadings			
M_a	.70	.70	.70	.70
M_b	.35	.35	.35	.35
M_c	.35	.70	.70	.35
M_d	.35	.35	.70	.70
M_a	.35	.35	.35	.35
M_b	.70	.70	.70	.70
M_c	.35	.70	.70	.35
M_d	.35	.35	.70	.70
M_a	.35	.35	.35	.35
M_b	.35	.35	.35	.35
M_c	.35	.70	.70	.35
M_d	.35	.35	.70	.70
M_a	.70	.70	.70	.70
M_b	.70	.70	.70	.70
M_c	.35	.70	.70	.35
M_d	.35	.35	.70	.70

Note. a = dimension factor loading; b = exercise factor loading; c = dimension intercorrelation; d = exercise intercorrelation.

Data Analyses

For each of the sixteen population models noted above, 500 sample data sets were generated and each data set had a sample size of 200. These data sets were created using SPSS software via the utilization of a Cholesky matrix. Each data set was then analyzed via the MIVQUE0 method via the SAS VARCOMP procedure. The MIVQUE0 method makes no assumptions regarding the normality of the data and can be utilized for analyzing unbalanced (i.e., ACs that do not fully cross dimensions and exercises) designs (Hartley, Rao, & Lamotte, 1978). Additionally, MIVQUE0 is one of the most efficient computational methods available (Bell, 1985; Brennan, 2001). This study included analyses on 8,000 individual data sets – thus, analytical efficiency was of paramount concern.

CHAPTER 3: RESULTS

The results for the sixteen population models are displayed in Tables 2 to 5. Expected values were calculated using the following formulas:

$$\text{Average MTHM } r = a^2 + (a^2 \times d), \quad (11)$$

$$\text{Average HTMM } r = b^2 + (a^2 \times c), \text{ and} \quad (12)$$

$$\text{Average HTHM } r = (a^2 \times c) + (b^2 \times d), \quad (13)$$

where a is the dimension loading, b is the exercise loading, c is the dimension intercorrelations, and d is the exercise intercorrelations. As predicted, the Monte Carlo mean estimates for the sixteen models are almost identical to the expected values. The very minor differences are due to the small sample size (*n* = 200) that was utilized as being indicative of typical AC research. However, even with the small sample size all sixteen models had very small standard deviations indicating little variance among the samples. Thus, overall, the variance partitioning of all sixteen models appears to have provided accurate results.

As previously discussed, variance partitioning can indicate values of convergence, discrimination, and method variance using variance components of person effects, person by dimension effects, and person by exercise effects (Woehr et al., 2011). Values for convergence, two conditions of discrimination, and method variance for each population model were calculated using the following formulas derived by Woehr and colleagues:

$$C1 = \sigma^2_{p} + \sigma^2_{pt}, \quad (14)$$

$$D1 = \sigma^2_{pt}, \quad (15)$$

$$D2 = \sigma^2_{pt} - \sigma^2_{pm}, \text{ and} \quad (16)$$

$$MV = \sigma^2_{pm} \quad (17)$$

and are also displayed in Tables 2 to 5. For ideal AC models, values of C1, D1, and D2 should be high and method variance should be low.

As shown in Table 2, for population models with low dimension and low exercise intercorrelations, the first model in which both dimension and exercise loadings are .70, the person effect, person by dimension effect, and person by exercise effect are all high with variance components of .342, .322, and .317 respectively. The second model in which dimension loadings are .70 and exercise loadings are .35 represents the overall ideal model. The person effect and person by dimension effect are relatively high with variance components of .215 and .320 indicating convergence. Additionally, the person by exercise component is very low (.076). Moreover, variance components for convergence (.535) and discrimination (D1 = .320 and D2 = .244) are high and the variance component for method variance (.076) is very low. This is what would be expected in an appropriately functioning AC. The third model in which dimension loadings are .35 and exercise loadings are .70 indicates a bad model. The person effect is relatively large at .215; however, the person by dimension effect is very low at .082 indicating low convergence which is seen in the value of C1. The person by exercise effect is also very high (.321) thus producing a large amount of method variance. The final model in which both dimension and exercise loadings are .35, the person effect, person by dimension effect, and the person by exercise effects all fall below .077 with the error variance component very high at .753.

Table 2

Simulation Results for Population Matrices Demonstrating Low Dimension Intercorreations (.35) and Low Exercises Intercorrelations (.35)

Dimension	Exercise	Facet	Expected Values	Population Matrix	Monte Carlo Estimates[a] *M*	*SD*
.70	.70	p	.343	.343	.342	.028
		pd	.319	.319	.322	.028
		px	.319	.319	.317	.035
		pdx,r	.020	.020	.019	.002
		C1	.662	.662	.664	
		D1	.319	.319	.322	
		D2	.000	.000	.005	
		MV	.319	.319	.317	
.70	.35	p	.214	.214	.215	.024
		pd	.319	.319	.320	.038
		px	.080	.080	.076	.009
		pdx,r	.388	.388	.385	.039
		C1	.533	.533	.535	
		D1	.319	.319	.320	
		D2	.239	.239	.244	
		MV	.080	.080	.076	
.35	.70	p	.214	.214	.215	.024
		pd	.080	.080	.082	.008
		px	.319	.319	.321	.027
		pdx,r	.388	.388	.386	.033
		C1	.294	.294	.296	
		D1	.080	.080	.082	
		D2	-.239	-.239	-.239	
		MV	.319	.319	.321	
.35	.35	p	.086	.086	.086	.008
		pd	.080	.080	.077	.008
		px	.080	.080	.077	.007
		pdx,r	.755	.755	.753	.073
		C1	.165	.165	.163	
		D1	.080	.080	.077	
		D2	.000	.000	.000	
		MV	.080	.080	.077	

Note. p = person; pd = person x dimension; px = person x exercise; pdx,r = error. [a]All Monte Carlo estimates are from *k* = 500 samples with *n* = 200 and analyzed via the MIVQUE(0) model in SAS 9.3.

In Table 3 for population models with low dimension intercorrelations and high exercise intercorrelations, the first model indicates a high person effect, high person by dimension effect, and high person by exercise effect with variance components of .513, .322, and .143 respectively. The second model with high dimension loadings and low exercise loadings indicates an acceptable AC functioning model. The person effect and person by dimension effects are high with variance components of .258 and .316 respectively, indicating convergence. Moreover, further evidence of convergence is seen in C1 with a large variance component of .574. Additionally, the person by exercise effect is very low (.040) indicating a very small amount of method variance. The third model with low dimension loadings and high exercise loadings demonstrates a model with low convergence. The person effect is high (.383); however the person by dimension effect is very low (.083). Additionally, the person by exercise effect is somewhat large (.150). Additionally, D1 and D2 variance components demonstrate almost nonsignificant evidence of discrimination. The fourth model with both low dimension and exercise loadings demonstrates a horrible model with a relatively large person effect (.132) and very large error (.756). The person by dimension effect (.080) was very small indicating hardly any convergence.

Table 3

Simulation Results for Population Matrices Demonstrating Low Dimension Intercorreations (.35) and High Exercises Intercorrelations (.70)

Dimension	Exercise	Facet	Expected Values	Population Matrix	Monte Carlo Estimates[a] *M*	*SD*
.70	.70	p	.515	.515	.513	.057
		pd	.319	.319	.322	.027
		px	.147	.147	.143	.015
		pdx,r	.020	.020	.022	.002
		C1	.833	.833	.836	
		D1	.319	.319	.322	
		D2	.172	.172	.179	
		MV	.147	.147	.143	
.70	.35	p	.257	.257	.258	.028
		pd	.319	.319	.316	.028
		px	.037	.037	.040	.003
		pdx,r	.388	.388	.387	.032
		C1	.576	.576	.574	
		D1	.319	.319	.316	
		D2	.282	.282	.276	
		MV	.037	.037	.040	
.35	.70	p	.386	.386	.383	.034
		pd	.080	.080	.083	.008
		px	.147	.147	.150	.013
		pdx,r	.388	.388	.388	.037
		C1	.466	.466	.466	
		D1	.080	.080	.083	
		D2	-.067	-.067	-.067	
		MV	.147	.147	.150	
.35	.35	p	.129	.129	.132	.014
		pd	.080	.080	.080	.009
		px	.037	.037	.037	.004
		pdx,r	.755	.755	.756	.079
		C1	.208	.208	.211	
		D1	.080	.080	.080	
		D2	.043	.043	.043	
		MV	.037	.037	.037	

Note. p = person; pd = person x dimension; px = person x exercise; pdx,r = error. [a]All Monte Carlo estimates are from *k* = 500 samples with *n* = 200 and analyzed via the MIVQUE(0) model in SAS 9.3.

As shown in Table 4, with high dimension intercorrelations and low exercise intercorrelations, the first model in which both dimension and exercise loadings are high, the person effect, person by dimension effect, and person by exercise effects are high, with variance components of .512, .144, and .320, respectively. The second model with high dimension loadings and low exercise loadings indicates an acceptable AC model. The person by dimension effect is relatively large (.145) indicating convergence and the person by exercise effect is very small (.080) indicating low method variance. There is also a very large person effect of .384. Further evidence of convergence is demonstrated by the C1 variance component. Evidence of discrimination is problematic since D2 is very low with a value of .065. The third model with low dimension loadings and high exercise loadings indicates a poor AC functioning model. The person by dimension effect is very small (.035) indicating very poor convergence. Moreover, the C1 variance component is relatively low and evidence of discrimination is nonexistent. Additionally, the person by exercise effect is very large (.316) indicating large method variance. The final model with both low dimension and exercise loadings indicates a relatively large person effect and very small person by dimension effects and person by exercise effects with variance components of .131, .038, and .084 respectively. This results in a very large error variance component of .754.

Table 4

Simulation Results for Population Matrices Demonstrating High Dimension Intercorreations (.70) and Low Exercises Intercorrelations (.35)

Dimension	Exercise	Facet	Expected Values	Population Matrix	Monte Carlo Estimates[a]	
					M	*SD*
.70	.70	p	.515	.515	.512	.047
		pd	.147	.147	.144	.013
		px	.319	.319	.320	.034
		pdx,r	.020	.020	.020	.002
		C1	.662	.662	.655	
		D1	.147	.147	.144	
		D2	-.172	-.172	-.177	
		MV	.319	.319	.320	
.70	.35	p	.386	.386	.384	.034
		pd	.147	.147	.145	.012
		px	.080	.080	.080	.007
		pdx,r	.388	.388	.391	.042
		C1	.533	.533	.528	
		D1	.147	.147	.145	
		D2	.067	.067	.065	
		MV	.080	.080	.080	
.35	.70	p	.257	.257	.256	.028
		pd	.037	.037	.035	.003
		px	.319	.319	.316	.036
		pdx,r	.388	.388	.390	.037
		C1	.294	.294	.291	
		D1	.037	.037	.035	
		D2	-.282	-.282	-.281	
		MV	.319	.319	.316	
.35	.35	p	.129	.129	.131	.012
		pd	.037	.037	.038	.004
		px	.080	.080	.084	.009
		pdx,r	.755	.755	.754	.090
		C1	.165	.165	.169	
		D1	.037	.037	.038	
		D2	-.043	-.043	-.045	
		MV	.080	.080	.084	

Note. p = person; pd = person x dimension; px = person x exercise; pdx,r = error. [a]All Monte Carlo estimates are from k = 500 samples with n = 200 and analyzed via the MIVQUE(0) model in SAS 9.3.

In Table 5, which shows population models with high dimension and high exercise intercorrelations, the first model with both high dimension and exercise loadings indicates a very high person effect, and relatively large person by dimension effects and person by exercise effects, with variance components of .689, .147, and .147, respectively. Furthermore, the model demonstrates very high convergence (C1 = .836), but low discrimination. Additionally, the model demonstrates relatively large method variance (MV = .147). The second model with high dimension loadings and low exercise loadings indicates an acceptable AC model. The person effect is very high (.429), but there is also a relatively large person by dimension effect (.145) which indicates some convergence. Further evidence of convergence is seen in C1 with a variance component of .576. Additionally, the person by exercise effect is very small with a variance component of .033, indicating very low method variance. The third model with low dimension loadings and high exercise loadings is considered the overall worst case model out of all sixteen models. The person effect is very large (.429) with a relatively large person by exercise effect (.144). Additionally, the person by dimension effect is very small (.033), indicating little convergence. Moreover, very little discrimination can be seen in D1 and D2 as well as a relatively large amount of method variance. The last model with both low dimension and exercise loadings indicates a relatively high person effect (.174) and a very high error variance component of .753. Both the person by dimension effect and the person by exercise effect are very small with variance components of .034 and .036, respectively indicating little convergence, very little evidence of discrimination, and little method variance.

Table 5

Simulation Results for Population Matrices Demonstrating High Dimension Intercorreations (.70) and High Exercises Intercorrelations (.70)

Dimension	Exercise	Facet	Expected Values	Population Matrix	Monte Carlo Estimates[a] *M*	*SD*
.70	.70	p	.686	.686	.689	.079
		pd	.147	.147	.147	.014
		px	.147	.147	.147	.012
		pdx,r	.020	.020	.021	.002
		C1	.833	.833	.836	
		D1	.147	.147	.147	
		D2	.000	.000	.000	
		MV	.147	.147	.147	
.70	.35	p	.429	.429	.431	.036
		pd	.147	.147	.145	.015
		px	.037	.037	.033	.004
		pdx,r	.388	.388	.391	.036
		C1	.576	.576	.576	
		D1	.147	.147	.145	
		D2	.110	.110	.112	
		MV	.037	.037	.033	
.35	.70	p	.429	.429	.429	.038
		pd	.037	.037	.033	.004
		px	.147	.147	.144	.017
		pdx,r	.388	.388	.386	.042
		C1	.466	.466	.463	
		D1	.037	.037	.033	
		D2	-.110	-.110	-.111	
		MV	.147	.147	.144	
.35	.35	p	.172	.172	.174	.018
		pd	.037	.037	.034	.003
		px	.037	.037	.036	.004
		pdx,r	.755	.755	.753	.071
		C1	.208	.208	.208	
		D1	.037	.037	.034	
		D2	.000	.000	-.002	
		MV	.037	.037	.036	

Note. p = person; pd = person x dimension; px = person x exercise; pdx,r = error. [a]All Monte Carlo estimates are from k = 500 samples with n = 200 and analyzed via the MIVQUE(0) model in SAS 9.3.

In summary, across all sixteen models variance partitioning produced adequate variance components regardless of the type of model. All models produced variance components very similar, if not identical, to the expected values derived from MTMM matrices with slight differences due to sample size. Moreover, variance partitioning was able to successfully differentiate among person effects, person by dimension effects, person by exercise effects, and error for all 16 models. Variance partitioning correctly distinguished among the sources of variance in PEDRs that would be expected based upon the population models from which they were drawn. Specifically, population models that represented adequate AC functioning produced variance components that demonstrated adequate levels of construct-related validity. Furthermore, convergence, discrimination, and method variance criterions provided additional evidence for construct-related validity.

CHAPTER IV: DISCUSSION

The primary goal of the present study was to determine whether variance partitioning differentiates among varieties of AC models and diagnoses additional aspects of ACs beyond basic dimension and exercise effects. The results of this study provide support for the use of variance partitioning. This method clearly differentiated among sixteen varieties of AC models ranging from an ideal model to a worst case model. In addition to being able to distinguish among models, variance partitioning was also able to detect person effects, person by dimension effects, and person by exercise effects. When the ideal model was analyzed, as expected, the person by dimension effect was the primary source of variance, indicating convergent validity. When the worst case model was analyzed, error was the primary source of variance followed by the person by exercise effect, as expected. Moreover, dimension factor intercorrelations made a considerable impact on the person effect, with greater intercorrelations yielding a more substantial person effect.

Another benefit of variance partitioning is that it can disregard the issues that arise with CFA model fit and admissible solutions. Variance partitioning only focuses on the produced variance components; however, CFA involves assessing the model's fit and whether a CFA can produce an admissible solution for the data. Lance et al. (2007) portrays the problem with CFA producing appropriate model fit. As previously mentioned, traditional fit statistics in Lance et al.'s research indicated that their various models constituted a good fit for the data regardless of whether or not the fitted model matched the population model from which the data was based. Variance partitioning is not plagued by such serious analytic issues. Not only does variance partitioning separate explained variance into more variance components compared to CFA, but it also can demonstrate a model's convergence without using model fit statistics.

Implications

The results of Lance et al. (2007) highlight the necessity of finding a new method of evaluating the construct-related validity of ACs. For almost 30 years, a substantial amount of AC research has been based upon the application of CFA techniques to the analysis of AC MTMM matrices (e.g. Lance, Foster, et al., 2004; Lance, Lambert, et al., 2004; Lance et al., 2000; Schneider & Schmidt, 1992). As previously discussed, the CFA approach to analyzing MTMM matrices is problematic in numerous ways (Lance et al., 2004). The convergence problems with CFA have led researchers to believe ACs do not function as intended and therefore have little construct validity (Lance et al., 2007; Lievens and Conway, 2001). Because of the flaws associated with CFA, it is critical for AC researchers to begin utilizing alternative methods in evaluating AC functioning. The present study provides support for such a method to be an effective way to analyze AC MTMM matrices.

This new method also determines additional variance components (e.g. person effects) to assess the effects of individual AC raters as well as the individuals being assessed. Even more critical are the interactions between facets. The ability to differentiate between a person effect, a dimension effect, and a person by dimension interaction provides substantially more information than a dimension effect alone. Variance partitioning distinguishes among several sources of variance whereas CFA only has the ability to distinguish between dimension and exercise effects which may be an overly simplistic view of AC functioning. Similarly, numerous researchers have claimed that exercise effects represent valid performance-related information. This may indeed be the case; however, previous research has provided little information regarding these effects. Are these exercise effects indicative of general exercise difficulty or of something more along the lines of situational specific? Although some research has attempted to address this issue (e.g. Lance et al., 2000), most has simply focused on “exercise effects.” The traditional dichotomy between dimensions and exercises is incapable of answering questions such as whether exercise effects are due to general exercise difficulty or due to situational specificity. However, variance partitioning can provide answers to these questions.

Convergence, discrimination, and method variance can also be easily assessed in variance partitioning by using the standardized methodology described by Woehr et al. (2011). Ideally, the convergence criterion and the two conditions of discrimination should have large variance components whereas the method variance criterion should have a small variance component. As previously mentioned, ACs should ideally demonstrate variance attributable to the same dimension measured across several exercises (convergence) and not variance attributable to within exercises (method variance). It would be impossible to assess all possible population model combinations; however, inferences can be made based on the correlation matrix of Woehr et al.’s convergence, discrimination, and method variance criterions (2011). As seen in Table 6, C1 has a significant positive relationship with both conditions of discrimination which is to be expected. More importantly, C1, D1, and D2 are all negatively related to method variance. Thus, these three criterions can be manipulated without increasing the level of method variance. Specifically, one can increase the convergence criterion and subsequently decrease method variance.

Table 6
Correlational Relationship Between Convergence, Discrimination, and Method Variance

	1	2	3	4
Convergence	1.00			
Discrimination 1	.82**	1.00		
Discrimination 2	.70**	.90**	1.00	
Method Variance	-.09*	-.19*	-.61**	1.00

Notes. *p < .05; **p < .01

Future Research

Future research should determine cut off values for the various variance components similar to those established for alpha and the .05 criterion for *p* values. Currently, researchers determine the appropriateness of variance component values by comparing them to each other. For example, a person by dimension effect adequately demonstrates construct-related validity when it is larger than the person by exercise effect. There are currently no established variance component values to determine construct-related validity thresholds. Such a value would signify the threshold that differentiates adequate construct-related validity from poor construct-related validity. Future research should establish such threshold or cut off values to ensure that all researchers have the same interpretation of the different variance component values.

Additionally, future research should also investigate the use of variance partitioning in other areas that use MTMM data. For example, 360 degree feedback currently uses the traditional CFA approach to analyze MTMM data. Since CFA clearly has problems analyzing MTMM data, it may be useful if 360 degree feedback data was analyzed using variance partitioning instead of the traditional CFA method. Researchers in the school psychology field also currently use CFAs to analyze the appropriateness of scales. Specifically, Hill and Hughes (2007) examined the convergent and discriminant validity of the Strengths and Difficulties

Questionnaire (SDQ) to determine its factor structure. Each of the constructs was examined by students, teachers, and parents. A CFA was used to analyze the MTMM data to determine the convergent and discriminant validity of the SDQ. Since the present study presented problems using the CFA method, Hill and Hughes could potentially have incorrect results. Additionally, variance partitioning enables researchers to disregard various goodness-of-fit indices to determine the if the model demonstrated an adequate fit for the data, since the use of variance partitioning only focuses on the produced variance components. Moreover, variance partitioning has the ability to differentiate among numerous facets of variance. Other research areas that use MTMM data can benefit from the additional variance components identified in variance partitioning.

Study Limitations

The primary limitation of the present study is the use of Monte Carlo data to determine the usefulness of variance partitioning. The data analyzed was not produced from real participants in ACs, but rather from sixteen population models with predetermined parameters. However, the analysis techniques used in this study are considered some of the most efficient methods available to researchers. Furthermore, an additional limitation involves the complexity of the models evaluated. As previously mentioned, it is impossible to evaluate all possible combinations of the population models. In order to align with the typical AC framework, sixteen models were used to represent AC functioning. Numerous other combinations of population factor loadings and intercorrelations could have been utilized to generate countless other population models. Nonetheless, the sixteen population models that were analyzed represent the fundamental theoretical spectrum of AC functioning and many of the possible variants that are found in the literature.

Conclusion

Variance partitioning appears to be a fruitful approach for analyzing the construct-related validity of AC PEDRs. Unlike the traditional CFA method for analyzing MTMM matrices, variance partitioning correctly recognizes the sources of variance in PEDRs that is expected based on the corresponding population models. Furthermore, variance partitioning correctly identifies other sources of variance such as person effects and interactions between person, dimension, and exercise effects that the traditional CFA method is incapable of analyzing. Variance partitioning also allows the researcher to increase convergence and discrimination criterions while subsequently decreasing method variance, further demonstrating strong evidence of construct-related validity. Researchers should use this method when analyzing future MTMM matrices.

References

Arthur, W., Day, E., McNelly, T., & Edens, P. (2003). A meta-analysis of the criterion-related validity of assessment center dimensions. *Personnel Psychology, 56,* 125-154.

Arthur, W., Woehr, D. J., & Maldegen, R. (2000). Convergent and discriminant validity of assessment center dimensions: A conceptual and empirical re-examination of the assessment center construct-related validity paradox. *Journal of Management, 26,* 813-835.

Bell, J. F. (1985). Generalizability theory: The software problem. *Journal of Educational Behavior and Statistics*, *10*, 19-29.

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

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

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

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

Brennan, R. L. (2001). *Generalizability theory.* New York: Springer-Verlag.

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

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

Chen, H. C. (2006). Assessment center: A critical mechanism for assessing HRD effectiveness and accountability. *Advances in Developing Human Resources, 8*, 247-264.

Cronbach, L. J., Glesser, G. C., Nanda, H., & Rajaratnam, N. (1972). *The Dependability of Behavioral Measurements: Theory of generalizability for scores and profiles.* New York: John Wiley & Sons, Inc.

Fleenor, J. W. (1996). Constructs and developmental assessment centers: Further troubling empirical findings. *Journal of Business and Psychology, 3,* 319-335.

Hartley, H. O., Rao, J. N. K., & LaMotte, L. (1978). A Simple Synthesis-Based Method of Variance Component Estimation. *Biometrics, 34*, 233-244.

Hill, C. R., & Hughes, J. N. (2007). An examination of the convergent and discriminant validity of the strengths and difficulties questionnaire. *School Psychology Quarterly, 22,* 380-406.

Jackson, D. J., Stillman, J. A., & Atkins, S. G. (2005). Rating task versus dimensions in assessment centers: A psychometric comparison. *Human Performance, 18,* 213-241.

Joiner, D. A. (2002). Assessment centers: whats new? *Public Personnel Management, 31,* 179-185.

Kenny, D. A., & Kashy, D. A. (1992). Analysis of the multitrait-multimethod matrix by confirmatory factor analysis. *Psychological Bulletin, 112,* 165-172.

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

Lance, C. E., 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.

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 postexercise dimension ratings. *Journal of Applied Psychology, 89,* 377-385.

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

Lance, C. E., Woehr, D. J., & Meade, A. W. (2007). Case study: A Monte Carlo investigation of assessment center construct validity models. *Organizational Research Methods, 10,* 430-448.

Lievens, F., Chasteen, C. S., Day, E. A.,& Christiansen, N. D. (2006). Large-scale investigation of the role of trait activation theory for understanding assessment center convergent and discriminant validity. Journal of Applied Psychology, 91, 247-258.

Lievens, F., & Conway, J. M. (2001). Dimension and exercise variance in assessment center scores: A large-scale evaluation of multitrait-multimethod studies. *Journal of Applied Psychology, 86,* 1202-1222.

Marsh, H. W. (1989). Confirmatory factor analysis of multitrait-multimethod data: Many problems and a few solutions. *Applied Psychological Measurement, 13,* 335-361.

Marsh, H. W., & Grayson, D. (1995). Latent variable models of multitrait-multimethod data. In R. H. Hoyle (Ed), *Structural equation modeling: Concepts, issues, and applications* (pp. 177-198). Thousand Oaks, CA, US: Sage Publications, Inc.

Schmitt, N. & Chan, D. (1988). Personnel selection: A theoretical approach. Thousand Oaks, CA: Sage.

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

Shavelson, R. J., & Webb, N. M. (1991). *Generalizability theory: A primer.* Thousand Oaks, CA, US: Sage Publications, Inc.

Spychalski, A. C., Quinones, M. A., Gaugler, B. B., & Pohley, K. (1999). A survey of assessment center practices in organizations in the United States. *Personnel Psychology*, *50*, 71-90.

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

Thornton, G.C. & Mueller-Hanson, R. A. (2004). *Developing organizational simulations.* Philadelphia, PA: Lawrence Erlbaum Associates.

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

Woehr, D. J., Putka, D. J., & Bowler, M. C. (2011). An examination of g-theory methods for modeling multitrait-multimethod data: Clarifying links to construct validity and confirmatory factor analysis. *Organizational Research Methods.*