University of Tennessee, Knoxville

TRACE: Tennessee Research and Creative Exchange

Doctoral Dissertations

Graduate School

12-2008

A Quantitative Review and Analysis of the Constructs Underlying Assessment Center Ratings: What are we Measuring?

John P. Meriac

University of Tennessee - Knoxville

To the Graduate Council:

I am submitting herewith a dissertation written by John P. Meriac entitled "A Quantitative Review and Analysis of the Constructs Underlying Assessment Center Ratings: What are we Measuring?." I have examined the final electronic copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Industrial and Organizational Psychology.

David J. Woehr, Major Professor

We have read this dissertation and recommend its acceptance:

Michael C. Rush, T. Russell Crook, Eric D. Sundstrom

Accepted for the Council:

Carolyn R. Hodges

Vice Provost and Dean of the Graduate School

(Original signatures are on file with official student records.)

To the Graduate Council:

I am submitting herewith a dissertation written by John P. Meriac entitled "A Quantitative Review and Analysis of the Constructs Underlying Assessment Center Ratings: What are we Measuring?" I have examined the final electronic copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Industrial and Organizational Psychology.

David J. Woehr, Major Professor

We have read this dissertation
and recommend its acceptance:

Michael C. Rush

T. Russell Crook

Eric D. Sundstrom

Accepted for the Council:

Carolyn R. Hodges, Vice Provost
and Dean of the Graduate School

(Original signatures are on file with official student records.)

A QUANTITATIVE REVIEW AND ANALYSIS OF THE CONSTRUCTS UNDERLYING ASSESSMENT CENTER RATINGS: WHAT ARE WE MEASURING?

A Dissertation Presented for the

Doctor of Philosophy

Degree

The University of Tennessee, Knoxville

John P. Meriac

December 2008

DEDICATION

This dissertation is dedicated to my loving wife Jamie. Your encouragement, patience, and support have made all of this possible. Becoming “Dr.” has been a fun ride, but without your reminders to slow down and relax when I need to, I would not have made it this far in one piece. To you I am eternally indebted.

ACKNOWLEDGEMENTS

I must of course thank my major professor Dave Woehr for his guidance not only through this dissertation but throughout my entire graduate career. I am forever grateful for your insight and assistance with the many projects we have completed. Being your student truly has been a great experience, and I aspire to be such a great advisor and role model for my own students. In addition, I would like to thank Mike Rush for showing me the importance of using theory to drive scientific discovery. Your perspective on science has made me a better researcher and I have enjoyed working with you. Also, I would like to thank Russell Crook and Eric Sundstrom for their insight and assistance throughout this project. It has been a joy to work with you and I hope we get the chance to work together again one day.

Of course I must again mention my wife Jamie for making the sacrifices she has for all these years. In addition to the time she has spent taking care of the both of us while I sit in front of my computer, she has endured as a patient partner. Without her support and encouragement this would not have been possible. In addition, my immediate family of my parents Ivan and Kathleen and my brother Paul Meriac, as well as my extended family of Pete, Linda, and Megan McKeon deserve many thanks. Your encouragement has been inspiring and helped me keep everything in perspective. Also, my faithful dog Tess has always made sure that I get plenty of exercise and take breaks away from writing and analyzing data to take a walk when I needed it.

I must also give credit to scholars who took the time to work with me in my earlier years. I would like to thank Lori Foster Thompson for taking on an abandoned undergraduate student and turning me on to I-O Psychology. My career as an I-O Psychologist truly started with your guidance. Also, I would like to thank Pete Villanova for providing me with additional direction.

You always took the time to push me to do my best, and you never let me take any shortcuts. I am truly grateful for all you have done for me.

As a graduate student, there have been many individuals who have helped me both as colleagues as well as great friends, and without you I never would have made it this far. I would like to thank Brian Hoffman, Mark Bowler, Carrie Blair, Joy Oliver and Taylor Poling for your tolerance on the projects we have worked on together, as well as your willingness to blow off steam when necessary. Whether we were coding studies for a meta-analysis, running endless analyses, living in a lab, or playing video games, hanging out at an impromptu happy hour, or eating Indian food, you guys were a huge part of my life for the last 4 years and it has been my honor to know you. Also, in no particular order I would like to thank Allen Gorman, Bill Walton, Cheryl Barksdale, Josh Bazzy, Shawn Bergman, Lisa Delise, Matt Fleisher, Elizabeth McGee, Abby Mello, Joana Pimentel, Josh Ray, Adam Smith, and Scott Turner for being such good friends and colleagues.

ABSTRACT

The overarching goal of this study was to clarify what constructs are being measured by assessment centers (ACs). ACs have been used and studied for years, yet have measurement problems that generally center on the use of information at the dimension-level. However, a necessary step in examining this issue has been neglected: a proper delineation of what constructs ACs actually measure. In an attempt to address this issue, this study's primary purpose was to explore the factor structure of AC dimensions. Several a priori models from both the AC and job performance literature were examined as frameworks for explicating the constructs representing dimensions. Data from two sources were used to address this question: Intercorrelations from primary studies were synthesized using meta-analysis (k = 57) and used as input for a series of confirmatory factor analysis models. In addition, the extent to which subject matter experts perceived these broader categories to operate as a summary framework was evaluated by asking experienced AC raters to categorize primary dimensions into the categories of each model.

The results showed that Arthur et al.'s (2003) framework provided a good fit to the data, offering additional evidence in support of this model. When compared against several alternative frameworks, Arthur et al.'s (2003) model also provided a better fit to the data than the alternatives. Hence, these seven categories provide a viable framework for explaining what constructs underlie AC dimension ratings. In addition, subject matter experts had the highest level of agreement when classifying primary dimensions into this framework. In addition, several hierarchical models were tested based on the a priori models examined in the study. Of these models, a hierarchical three-factor model fit the well, indicating that a set of higher-order summary categories may also explain variance in the seven factors of Arthur et al.'s (2003) framework.

Overall, this study provides some clarity on what constructs underlie AC dimension ratings. These findings are expected to make contributions for AC research and practice. Implications for these results, as well as limitations of the study and future directions for research are discussed.

TABLE OF CONTENTS

SECTION PAGE

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

In managerial selection and development, one of the most popular evaluation methods is the assessment center (AC; Howard, 1997). For administrative purposes, ACs have seen continued use due in large part to their high fidelity (Cascio & Aguinis, 2005), predictive validity evidence (Arthur, Day, McNelly, & Edens, 2003), and a lack of subgroup differences in AC ratings (c.f., Dean, Roth & Bobko, 2008; Thornton & Rupp, 2006). ACs have also seen an increased use in management and leadership development programs (Spychalski, Quiñones, Gaugler, & Pohley, 1997). Here, the AC serves as a rigorous training intervention, where feedback is provided to participants and serves as a baseline for identifying strengths and weaknesses in behavioral dimensions. In relation, ACs lead to more positive reactions from job applicants (Macan, Avedon, Paese, & Smith, 1994) and higher perceived objectivity compared with alternative assessment methods (e.g., multi-source feedback; Howard, 2006) as well as more positive feedback acceptance (Cascio & Aguinis, 2005). Given these positive characteristics, ACs are likely to see continued use in organizations.

A key characteristic of ACs is that they are typically designed to measure behavioral dimensions. As noted by Thornton and Rupp (2006), dimensions are the basic unit of analysis in ACs. Most AC studies regard dimensions as the 'constructs' that are measured via the AC method (Arthur, Day & Woehr, 2008), which are purported to be relevant to job performance. For example, commonly used dimensions might include problem solving, influencing others, and communication (e.g., Arthur et al., 2003). In typical ACs, behaviors that serve as indicators of these dimensions are observed as an assessee participates in an exercise; examples of exercises include simulations with confederates (i.e., role players), leaderless group discussions, and in-basket exercises (Spychalski et al., 1997). A rating on one of these dimensions would depend on the quality and quantity of which these behaviors are exhibited and observed. Further, dimensions are typically intended to be distinct from one another and observable in multiple exercises. The expected outcome of the AC design is that ratings on these dimensions will be consistent across exercises, where behaviors exhibited in the exercises will serve as a sample of the behaviors that comprise the overall dimensions (Arthur et al., 2008). The specific set of dimensions on which an AC provides ratings suggests what content it measures.

The measurement properties of AC ratings have been a topic of debate for decades. In short, the criterion-related validity of ACs is fairly well established; ACs have been demonstrated to be strong predictors of job performance (Arthur et al., 2003; Gaugler, Rosenthal, Thornton & Bentson, 1987; Schmidt & Hunter, 1998). However, evidence for the construct-related validity of AC ratings is more problematic. Central to this problem is how well ACs measure the dimensions (i.e., ‘constructs’) they purport to measure. Many studies have sought to examine the extent to which dimension ratings converge across exercises, predominantly through internal approaches (e.g., multi-trait multi-method, or MTMM designs; Campbell & Fiske, 1959; Thornton & Rupp, 2006). These studies have generally demonstrated that dimensions measured in ACs do not exhibit convergent and discriminant validity as expected (Lance, 2008; Lance, Lambert, Gewin, Lievens, & Conway, 2004). Rather, ratings are more consistent within exercises across dimensions than within dimensions across exercises. Based on these findings, it is suggested that AC ratings may not represent the dimensions they purport to, and the construct-related validity of AC dimensions is suspect.

Several explanations have been suggested for why this problem exists, including measurement / design issues (Woehr & Arthur, 2003; Woehr, Arthur & Meriac, 2007). For example, a question has been raised as to whether raters have the cognitive capacity to effectively distinguish among a large number of dimensions, and studies have shown that smaller dimension sets typically result in more reliable ratings (Gaugler & Thornton, 1989; Thornton & Rupp, 2006). As reported by Woehr and Arthur (2003), ACs measure on average 11 dimensions and in some cases as many as 25. Another potential reason for the lack of construct-related validity evidence hinges on the lack of properly developed constructs (Arthur et al., 2008). Specifically, the 'constructs' that are measured in ACs are not well developed or clearly defined. Subsequently, the existing research is fragmented with respect to the different dimensions that ACs measure, or what these dimensions represent.

The vast number of dimension names reported in the literature indicates that there is considerable variability in what ACs purport to measure. For example, in a review of the literature, Arthur et al. (2003) identified 168 dimension labels in the 34 articles they identified. Surely 168 or more dimensions are not necessary to explain what is measured in ACs. One approach toward rectifying this situation is to develop a unifying framework for classifying behaviors that would group these dimensions into broader categories.

Arthur et al. (2003) sought to develop such a framework, specifically for ACs. They conceptually grouped the primary dimension labels they identified into a set of seven broad categories. In their study, they demonstrated the usefulness of conceptualizing the criterion-related validity of ACs at the dimension-level, each of these broader categories significantly predicted job performance. Subsequent research has shown that this broader set of categories is useful for improving the validity of ACs. For instance, Bowler and Woehr (2006) have shown that collapsing primary dimensions into Arthur et al.'s (2003) taxonomy improves construct-related validity evidence associated with primary dimensions. Specifically, they demonstrated that dimension effects were relatively the same size as exercise effects and the strength of these effects varied depending on specific dimension-exercise combinations. Also, Meriac, Hoffman, Woehr and Fleisher (in press) demonstrated that these dimensions each explained incremental variance in job performance ratings above and beyond cognitive ability and personality variables. Thus, the common set of dimensions proposed by Arthur et al. (2003) has so far been demonstrated as a useful approach toward conceptualizing AC ratings, and appears to be a step in the right direction toward improving the measurement issues associated with ACs.

However, it is not clear whether Arthur et al.'s (2003) framework is the best way to conceptualize what it is that ACs measure. In their initial study, Arthur et al. (2003) conceptually grouped dimensions into categories, but the factor structure of their model has yet to be examined empirically (i.e., using factor analysis). In addition, although it serves as a more parsimonious model for conceptualizing AC dimensions than most ACs, other models with even fewer dimensions may further improve upon AC measurement issues.

The purpose of this study is to further explore what it is that ACs measure (i.e., what constructs), specifically with respect to a general model. Several models and frameworks have been proposed in the published AC literature, many of which are more parsimonious such that they contain fewer, broader dimensions. In addition, the job performance literature has proposed several general models of performance. As ACs are purported to measure constructs that are important for the prediction or development of job performance, models from this stream of research may also be relevant approaches toward conceptualizing what is measured in ACs.

In testing the different models that may be used to explicate the constructs underlying AC ratings, meta-analytic procedures were used. Data were gathered from studies reporting AC dimension interrelationships and used to examine which of several different dimension structures may serve as the most appropriate model for conceptualizing what it is that ACs measure. Confirmatory factor analysis (CFA) was used to evaluate the extent to which each of these models fits the data reported in the AC literature. In addition, the extent to which experienced AC raters can reliably classify primary AC dimensions into each of these dimension structures was explored. This effectively indexed the extent to which raters can conceptualize AC ratings via each of these alternative dimension structures. In summary, this study aimed to provide clarification on what constructs ACs measure.

CHAPTER 2

LITERATURE REVIEW

Assessment Center Dimensions and What they Represent

Dimensions are an important component of the AC method. As defined by Thornton and Rupp (2006, pp. 77-78), AC dimensions are “homogenous cluster[s] of observable behaviors” and they purportedly represent the “constructs” that ACs measure (Arthur et al., 2008). From a traditional psychometric perspective, dimensions have been viewed as latent factors, and the behaviors exhibited by participants in the exercises serve as observable manifest indicators. As noted by Hoeft and Schuler (2001), the original guiding principle behind AC development and use is at least in part to predict job-relevant criteria, as these same dimensions are presumed to underlie job performance. The reasoning behind this idea stems from claims that AC dimensions are in some way based on information from a job analysis.

Despite the vast amount of attention devoted to evaluating the measurement properties of ACs, relatively little effort has been directed toward understanding exactly how AC dimensions should be conceptualized and what constructs ACs are measuring. Most AC studies provide little or no information regarding how dimensions are developed or why they are important for performance. The process of developing dimensions in ACs rarely follows the same process as the development of constructs measured by paper-and-pencil methods (e.g., cognitive ability or personality variables). That is, typical construct development involves a rigorous, iterative process including careful definition and refinement of what it is that is measured. This simply has not been the case in AC research and practice (Arthur et al., 2008). Instead, AC designers have often casually labeled clusters of behaviors as dimensions. In most AC studies, there appears to be little or no a priori reasoning as to how the dimensions were chosen or conceptualized, nor the expected relationships among these dimensions. Subsequently, there is still a great deal of ambiguity regarding what constructs ACs measure.

This ambiguity in what constructs are measured in ACs is problematic for several reasons. As reviewed by Arthur et al. (2003), the importance of constructs in psychology is paramount to the goal of describing, understanding, and predicting behavior (Binning & Barrett, 1989). In the AC literature, studies evaluating the construct-related validity of AC dimensions take these dimension names (i.e., the supposed 'constructs' that are evaluated) at face value. More specifically, the studies evaluating the construct-related validity of ACs treat dimensions as generic constructs with little or no examination of whether they are appropriately conceptualized.

It is important to explicitly make the distinction between dimensions and constructs. Dimensions represent clusters of observable behaviors. Take for example 'analysis' and 'judgment', two commonly evaluated dimensions. According to Thornton and Byham (1982, p. 139), analysis represents "identifying problems, securing relevant information, relating data from different sources, and identifying possible causes of problems" and judgment is "developing alternative courses of action and making decisions based on logical assumptions that reflect factual information". However, latent constructs may underlie these dimensions as the key variables of interest in ACs. For instance, Arthur et al. (2003) collapsed these dimensions into a broader category: 'problem solving', due to the conceptual similarity between these dimensions as well as several others. Rather than reporting information at the dimension-level, it may be more meaningful, both theoretically and practically, to evaluate how well these primary dimensions represent broader, underlying constructs. Further, evaluating constructs is important whether an AC is used for administrative purposes (i.e., predicting theoretically meaningful criteria) or developmental purposes (i.e., measuring important constructs for management or leadership development).

The conceptualization of AC dimensions in primary studies has been largely idiosyncratic in nature, such that individual ACs have employed substantially different dimensions. As a result, one of the primary problems faced in examining the construct-related validity of ACs is the overwhelming number of ‘dimensions’ found in the literature. Recent meta-analyses of the AC literature have reported identifying anywhere from 79 to 168 different dimension labels (Arthur et al. 2003; Bowler & Woehr, 2006; Woehr & Arthur, 2003). As noted by Woehr and Arthur (2003), out of the 48 distinct samples they identified in their review, on average 11 dimensions were measured; the standard deviation was 5 dimensions and ranged from as few as 3 to as many as 25. While human behavior is certainly complex, it seems unlikely that such a vast quantity of dimensions is required to explain work performance.

A Historical Perspective on Constructs and Dimensions

Interestingly, the earliest ACs put forth the idea that ‘constructs’ of sorts may underlie AC dimension ratings, but not in the way they are commonly treated today. Early ACs were developed by the Office of Strategic Services (OSS) for the purpose of selecting military personnel, with the results published in the book *Assessment of Men* (1948; Thornton & Byham, 1982). In these early studies, 11 dimensions were initially rated, but exploratory factor analysis yielded four components from these 11 variables. Even before ACs were used in industrial settings, this approach set the stage for the idea that constructs or factors underlying the dimensions may be a more appropriate way to explain what ACs are measuring than the primary dimension labels themselves.

This trend continued as ACs made their way into industrial settings, particularly through the famous Management Progress Study (MPS; Bray, 1964; Bray and Grant, 1966). The rating method and exercises utilized in the MPS have been, at least in some form, utilized throughout the last half century, and many ACs use practically the same approach as Bray and colleagues did when the study began in 1956. In the MPS, 25 primary dimension categories were assessed; however, the authors also factor-analyzed these dimensions to find a more parsimonious framework for explaining the results. Their solution(s) for college graduates and non-college graduates showed that 11 and 8 factors emerged, respectively. Specifically, Bray and Grant (1966, p. 9) state: "The factorial results also help to clarify the constructs used by the staff evaluators". In other words, they were aware that some constructs were being measured by ACs, but these were actually evaluated post-hoc and not equated with dimensions. This approach continued for several years, with later studies taking this same approach, oftentimes to find more parsimonious solutions. For example, Schmitt (1977) found that 3 factors emerged out of a set of 17 dimensions. In these studies, an exploratory approach was taken toward evaluating the constructs that were measured by ACs.

In this discussion of what it is that ACs actually measure, a major turning point took place in the early 1980's. The notion of an AC 'validity paradox' emerged, and has since been an important issue since it begs the question of whether ACs measure the constructs they are supposed to. Specifically, this paradox refers to the idea that ACs demonstrate evidence for content-related validity (Norton, 1977) and criterion-related validity (Arthur et al., 2003; Gaugler et al., 1989), yet they do not demonstrate evidence for construct-related validity (Sackett & Dreher, 1982). The reason this issue is viewed as a paradox is because based on Binning and Barrett's (1989) unitarian conceptualization of validity, if ACs exhibit two of these forms of validity evidence, then they should logically also exhibit the third form. Studies that have shown that this is not the case have focused on a specific unit of analysis: within-exercise dimension ratings (i.e., post-exercise dimension ratings; PEDRs).

This practice initially began when Archambeau (1979) put forth the idea that within-exercise performance may be meaningful, due to the observed high correlations among ratings within exercises across dimensions. This idea was further advocated by Sackett and Dreher (1982), and sparked a great deal of debate among AC researchers. This argument has since continued, and most recently, several researchers (e.g., Jackson et al., 2005; Lance, 2008; Lance et al., 2004) have advocated the use of exercise ratings (as opposed to dimension ratings), also known as task-based ACs. A central feature of these studies is that they all use PEDRs. These ratings are then analyzed using an MTMM design, which allows for the evaluation of convergent and discriminant validity. As AC ratings do not exhibit the expected pattern of results in this design (e.g., convergence of dimensions across exercises and low dimension intercorrelations within exercises), they are described as failing to exhibit construct-related validity evidence.

The 'constructs' that are being evaluated in these studies are the primary dimensions measured in the ACs. For the most part, PEDRs are an artifact of this research design. Actually, the primary dimension labels in these studies (e.g., Bray & Grant, 1966) were not initially intended to be measured this way (Howard, 2008). More specifically, the original AC design was never applied with the expectation that all of the dimensions assessed would emerge as meaningful factors if one were to conduct a factor analysis on AC ratings. Instead, as noted above, AC researchers long ago noted that broader factors are expected to emerge from the AC dimension ratings, which are more meaningful as the ‘constructs’ that ACs measure.

Progress toward a More Unified Framework

Several models and frameworks have been presented that may serve as a priori models for examining the latent constructs that underlie AC performance. Some potential models have been developed specifically for ACs, using conceptual groupings (e.g., Arthur et al., 2003) or from exploratory results of primary studies (e.g., Schmitt, 1977). In addition, models from the general job performance literature have been proposed that may be applied in AC settings (e.g., Borman and Brush, 1993). *The overarching goal of the present study is to explore what constructs underlie AC performance*. Toward this end, multiple research questions will be evaluated to gather evidence directed toward answering this broad question.

The first of these more specific questions will focus on evaluating Arthur et al.’s (2003) framework as one potential a priori model describing the constructs that are measured in ACs. Arthur et al.’s (2003) model presents one approach toward organizing the various AC dimensions reported in the literature, and thus far this model has shown impressive results as a useful summarizing framework (e.g., Bowler & Woehr, 2006; Meriac et al., in press). However, although Arthur et al. (2003) provided a conceptual grouping of these AC dimensions, they did not empirically evaluate the fit of their model. Subject matter experts conceptually grouped primary dimensions into these broader categories, but they did not examine the fit of their model based on empirical data (e.g., via confirmatory factor analysis; CFA). An additional step toward evaluating the usefulness of this framework is examining whether Arthur et al.'s (2003) model provides an acceptable fit to data from ACs. This is one goal of the present study. By using the same primary dimension labels that Arthur et al. (2003) used (i.e., Thornton & Byham, 1982), it will be possible to conduct an examination of how well this model fits the reported AC data.

Research Question 1: How well does the available AC data provide an empirical verification of Arthur et al.'s (2003) seven-dimension AC framework?

As the primary objective of the present study is to resume the initial line of inquiry initiated by Bray and colleagues and continue to explore what constructs underlie the primary dimensions assessed in ACs, this study will take a step back of sorts to examine the factors that emerge from primary AC dimensions. However, the initial approaches to evaluating what constructs were measured in ACs only took an exploratory perspective, utilizing exploratory factor analysis and naming the factors that emerged post-hoc. A confirmatory approach may also be useful, where existing a priori models may serve as a theoretical basis for explicating what constructs ACs measure. Further, several of the previous studies exploring this idea have only examined one model. A more rigorous approach would entail an examination of multiple competing models, to allow for a determination of whether one model is more appropriate than another (Bollen, 2000).

Toward this end, models from both the general job performance literature as well as the AC literature may serve as frameworks that can explain what constructs underlie the primary dimensions measured by ACs. Since the 1960's, much work has been done with the purpose of trying to better understand the job performance domain. Many general models of performance have been developed in both the job performance literature (e.g., Borman & Brush, 1993), as well as the AC literature (e.g., Arthur et al., 2003). Several of these models will be compared against one another, specifically with respect to which one best fits the available data. These models will be reviewed in the following section.

Research Question 2: Which of the alternative a priori models best fits the AC data?

Another related issue is how well raters can classify dimensions into these frameworks. One component of AC rating is the capacity for raters to effectively observe and record behaviors within exercises. A different approach toward evaluating how these dimension structures operate is to determine how well raters can place primary dimension labels into broader dimensions in each respective model. For instance, Arthur et al. (2003) found that subject matter experts (SMEs) were able to reliably classify AC dimension labels in primary studies into each of Thornton and Byham's (1982) list of 33 commonly used dimensions. However, they never directly assessed the extent to which SMEs were able to classify Thornton and Byham's list of dimension labels into their seven categories. The present study will evaluate how well this longer list of dimension labels can be reduced into a smaller set of seven. A similar approach will be taken toward evaluating the competing models in this study. Specifically, trained and experienced AC raters (SMEs) will be asked to categorize dimension labels from Thornton and Byham's (1982) list into the broader categories of the alternative models as well. In effect, this will index how well raters can reduce a larger set of dimensions into a more parsimonious smaller set, as well as which framework leads to the highest level of agreement.

Research Question 3: How reliably can AC raters classify primary dimension labels into the dimensions in each of the a priori models?

Statement of Purpose and Expected Outcomes

In summary, much debate has centered on AC validity evidence. However, little attention has been directed toward exploring what constructs underlie performance as measured in ACs. Various models have been presented, yet no large-scale studies have been conducted to empirically evaluate this structure. The most successful of these endeavors to date (at least in the AC domain) has been Arthur et al.'s (2003) framework. However, this model has not been scrutinized with respect to its fit with AC dimension ratings. Thus, the first goal of the present study is to determine how well it fits AC data reported in the literature. Next, this model as well as several alternative frameworks will be compared to evaluate which model best fits data reported in the AC literature. Finally, the extent to which AC raters can classify primary dimensions into these broader dimensions will be assessed. These three research questions are posed to help answer the overarching question: *What constructs do ACs measure*?

CHAPTER 3

ALTERNATIVE MODELS / FRAMEWORKS

Despite the recent focus on the results of MTMM-oriented studies, there has been no resolution on what constructs underlie AC performance. The studies that have employed MTMM analyses have almost exclusively treated primary dimensions as the ‘traits’ measured in ACs (i.e., constructs). Primary AC dimensions may not be the appropriate level to conceptualize the constructs measured in ACs, but rather these dimensions may serve as indicators of broader latent constructs. Several models have been proposed that may explicate the constructs measured by ACs, which all serve to reduce the complexity of what we are measuring (i.e., fewer dimensions). The models discussed in this section were examined as a priori frameworks for clarifying the constructs that underlie AC performance.

Comparisons Among Alternative Models

As reviewed by Viswesvaran and Ones (2000), early models of job performance (Fleishman, 1967; Guilford, 1954) described clusters of homogenous tasks that were applicable across jobs. Viswesvaran and Ones provided a framework to group models of job performance, which contain dimensions that are either applicable across jobs or specific to a given occupation, and are either stand-alone or part of a set. The present study seeks to examine models of dimensions that are part of a set and are generalizable across occupations. Toward this end, frameworks taken specifically from the AC literature, as well as general models of job performance were evaluated as alternative models for explicating the constructs that underlie AC performance.

In deciding specifically which models were empirically examined in this study, several frameworks / models were first evaluated with respect to their applicability across occupations; in other words, if models were too restrictive in that they only apply to a narrow range of jobs, they were not examined in this study. Next, models deemed relevant / applicable for the present study were conceptually compared with respect to the content of their dimensions. If the content of one model was subsumed by another equally parsimonious model, then the model that appears to encompass more of the content domain was included. In other words, each model was reviewed in terms of its theoretical and conceptual relevance to the AC content domain. Based on this review, five core models presented themselves as viable frameworks for explicating the constructs underlying AC dimension ratings. These models will be discussed in the following sections.

Thornton and Byham's (1982) List of Common AC Dimensions

Thornton and Byham (1982) presented a list of 33 common AC dimension labels, which served as one of the first approaches toward summarizing the information assessed in ACs (See Table 12, Appendix B). Specifically, they noted that a common set of dimensions could be used to compare data from different sources. This list of dimensions has been utilized by large-scale studies (e.g., Arthur et al., 2003) to provide a common grouping of primary AC dimensions. Arthur et al. (2003) found that the majority of AC dimensions listed in the primary studies they identified could be classified into one of these 33 categories. Thornton and Byham's (1982) list serves as a comprehensive taxonomy for comparing dimensions across studies and will be used as the starting point for integrating information across primary studies. Hence, primary dimensions that are categorized into the dimensions in this list will serve as indicators for the alternative models that will be tested in this study.

Arthur et al.'s (2003) Model of AC Performance

As noted above, Arthur et al.'s (2003) seven broad categories have been shown as a useful framework for organizing the vast quantity of dimensions reported in the AC literature. Specifically, the categories they proposed are: 1) problem solving, 2) tolerance for stress / uncertainty, 3) influencing others, 4) consideration / awareness of others, 5) communication, 6) organizing and planning, and 7) drive. Definitions of these categories are listed in Table 6 (Appendix A). Arthur et al.'s (2003) seven broad categories have thus far been the most promising approach toward unifying the fragmented AC literature.

Additional studies have supported the use of this framework in that its use may improve AC validity evidence. Meriac et al. (in press) showed that each of these seven categories explained a significant proportion of variance in job performance above and beyond cognitive ability and the big five personality variables. In comparison with results offered by Schmidt and Hunter (1998) regarding the incremental gain in using alternative predictors along with cognitive ability, conceptualizing AC ratings at the construct-level dramatically improved the incremental variance explained in job performance. Also, Bowler and Woehr (2006) demonstrated that when these categories are combined with different exercises, some of them (e.g., communication, influencing others, organizing and planning, and problem solving) were more construct valid than others (e.g., consideration / awareness of others and drive). In comparison with Lance et al.'s (2004) findings, the use of Arthur et al.'s (2003) framework provides somewhat more promising results from an internal (i.e., MTMM) approach. In summary, Arthur et al.'s (2003) model shows several advantages over using primary dimension labels. However, it is possible that these broader categories / constructs may be reduced further.

Alternative Models of AC Performance

One model that presents fewer groupings than Arthur et al.'s (2003) framework is a four-dimension framework developed by Borman and Brush (1993). Borman and Brush's (1993) taxonomy of managerial performance consists of 18 dimensions which cluster into four broad categories. These broad categories of managerial performance are: 1) interpersonal dealings and communication, 2) leadership and supervision, 3) technical activities and the "mechanics of management", and 4) useful personal behavior and skills. The content contained in each of these broad categories is listed in Table 7 (Appendix A). One of the reasons Borman and Brush (1993) developed this model was to serve as a unifying framework for allowing classification of behavior and comparison across performance taxonomies.

As is evident in the dimension descriptions, Borman and Brush's (1993) framework also has a great deal of overlap with the categories proposed by Arthur et al. (2003). Specifically, interpersonal dealings and communication contains elements of both communication and consideration / awareness of others. Arthur et al.'s (2003) influencing others category is similar to Borman and Brush's (1993) leadership and supervision. Borman and Brush's (1993) technical activities and the "mechanics of management" category contains elements of both organizing and planning as well as problem solving. Finally, Borman and Brush's (1993) useful personal behavior category has an overlap of content with both drive and stress tolerance. Hence, Borman and Brush’s (1993) model appears to conceptually subsume the content contained in Arthur et al.’s (2003) seven categories.

Given these similarities, the capacity for this model to serve as a broad model of performance that is designed to subsume all of the behaviors that managers are expected to exhibit, as well as its apparent capacity to generalize across administrative and developmental ACs, this model will be tested (i.e., compared against) Arthur et al.’s (2003) framework. A direct comparison between these two models will indicate whether the content measured by Arthur et al.’s (2003) categories can be more parsimoniously measured by four broader categories. Hence, Borman and Brush’s (1993) model may represent the constructs measured in ACs, and will be empirically evaluated as an alternative model of AC performance, in comparison with Arthur et al.’s (2003) framework. However, it is possible that AC dimensions may be further grouped into an even more parsimonious model.

Schmitt (1977) took an empirically-driven approach toward evaluating an AC factor structure (in a primary study), and conducted a principal components analysis to derive 3 broad factors out of a larger set of 17 dimensions. In defining the content of these factors, Schmitt simply listed the primary dimensions that loaded onto each one. These factors are: 1) “administrative skills”, which is comprised of inner work standards, organizing and planning, decision making, decisiveness, and written communication skills, 2) “interpersonal skills”, which contains tolerance of uncertainty, self-objectivity, behavior flexibility, and leadership skills, and 3) “activity or forcefulness”, which includes energy, resistance to stress, need advancement, forcefulness, reliance on others, and oral communication.

Descriptions of the behaviors included in this taxonomy are listed in Table 8 (Appendix A). Interpreting these three factors beyond a simple listing of the dimensions which load onto them is necessary for proper construct definition. Administrative skills appears to involve general problem solving skills, which includes making appropriate decisions based on information gathered and weighting and prioritizing information and tasks. Interpersonal skill appears to involve accomplishing work through interactions with others and considering the demands of interpersonal situations and acting accordingly. Activity or forcefulness appears to involve one's effort expended toward task accomplishment, ability to remain focused, and resourcefulness. Although they are very broad in nature, Schmitt's (1977) factors were clearly developed from AC content and appear to encompass many of the common dimensions in Thornton and Byham's (1982) list.

As reported by Schmitt (1977), these factors were very similar to a set of similar factors derived by Hinrichs (1969). Specifically, two of Hinrichs's (1969) factors were almost identical to those reported by Schmitt (1977; administrative skills and activity); however, Hinrichs (1969) did not measure as many primary dimensions as Schmitt (1977), and subsequently Hinrichs's (1969) factors will not be directly examined in the present study. Still, it is interesting to see that there is some convergence across these three-factor models.

In comparison with Borman and Brush's (1993) dimensions, it is evident that the content contained in Borman and Brush's interpersonal dealings and communication and leadership and supervision categories have much similarity with the information contained in Schmitt's interpersonal skills factor. Also, Borman and Brush's (1993) technical activities and the "mechanics of management" and useful personal behavior categories are quite similar to

Schmitt's (1977) administrative skills and activity / forcefulness factors, respectively. Given this similarity, it appears that Borman and Brush's (1993) categories may fit into Schmitt's (1977) factors. As with the comparison between Arthur et al.'s (2003) categories and Borman and Brush's (1993), the Borman and Brush categories will be further compared with Schmitt's (1977) factors to determine whether this more parsimonious model better fits the data.

With the further possibility of an even more parsimonious model still differentiating among dimensions, a two-factor model may fit the data better than Schmitt's (1977) three factors. Shore et al. (1990) proposed a two-factor model of AC performance which includes 1) performance-style and 2) interpersonal-style factors. Their study involved grouping 11 dimensions into these two broad categories. For more information on these factors, see Table 9 (Appendix A). The interpersonal-style factor subsumed four of these primary dimensions: amount of participation, impact, personal acceptability, and understanding of people. The performance-style factor was comprised of seven primary dimensions: originality, oral communication, recognizing priorities, need for structure, thoroughness, work quality, and work drive. In comparison with Schmitt's (1977) model, interpersonal skills and interpersonal-style are quite similar to one another. In other words, there is substantial conceptual overlap between these two factors. In addition, the material contained in Schmitt's (1977) activity / forcefulness and administrative skills factors both have similarities with Shore et al.'s (1990) performance-style factor; hence, they may be able to be collapsed into this broader factor.

By grouping primary dimensions into these broad categories, Shore et al. (1990) were able to demonstrate that the interpersonal-style factor was more strongly related to several relevant personality constructs, and the performance-style factor was more strongly related to cognitive ability, providing evidence for convergent and discriminant validity of these factors. However, in their study Shore et al. (1990) also did not examine these factors with respect to competing models, and although their dimensions showed promising results with respect to AC dimension construct-related validity, there is no existing evidence to support whether this is the most appropriate model in comparison with alternatives.

Finally, some evidence suggests that given the positive manifold across the categories contained in Arthur et al.'s (2003) framework, a unidimensional model may provide the best representation of AC performance. Certainly, previous research has frequently indicated relatively strong correlations among categories assessed in ACs. In addition, Lance et al. (2004) have argued that the structure of ACs is best described with multiple exercise factors and a single overall or general performance factor. This model is also consistent with the research suggesting a single general performance factor underlying performance ratings (Viswesvaran, Schmidt, & Ones, 2005). Further, it is worthwhile to note that a unidimensional model is consistent with the frequent use of an overall assessment rating (OAR) as a composite indicator of AC performance in both research and practice, where an OAR represents the aggregation of ratings on separate dimensions (and/or exercises) into a single overall score. Viswesvaran et al. (2005) factor-analyzed a hierarchical model of job performance and found that a general model best fit the data. Thus, the present study will also compare a one-factor model to the two-factor model proposed by Shore et al. (1990).

Given the various performance models discussed above, five key a priori models that may explicate the constructs that underlie the AC dimensions measured in primary studies will be examined (See Table 1). These models are nested in a sequence such that the more narrowly defined dimensions can be collapsed into the next broadest level of conceptualization across the five models (See Figure 1). By comparing these models directly, it will be possible to determine which model best fits the data. By conducting goodness-of-fit tests on these models and comparing the fit indices across models, the best fitting model can be determined.

Table 1. Summary of A Priori Models Included in the Study

Arthur et al. (2003)	Borman & Brush (1993)	Schmitt (1977)	Shore e al. (1992)	Viswesvaran et el. (2005)
Communication	Interpersonal Dealings and Communication	Administrative Skills	Interpersonal-Style	General AC Performance
Consideration and Awareness of Others	Leadership and Supervision	Interpersonal Skills	Performance-Style	
Influencing Others	Technical Activities and the "Mechanics of Management"	Activity / Forcefulness		
Organizing and Planning	Useful Personal Behavior			
Problem Solving				
Drive				
Tolerance for Stress				

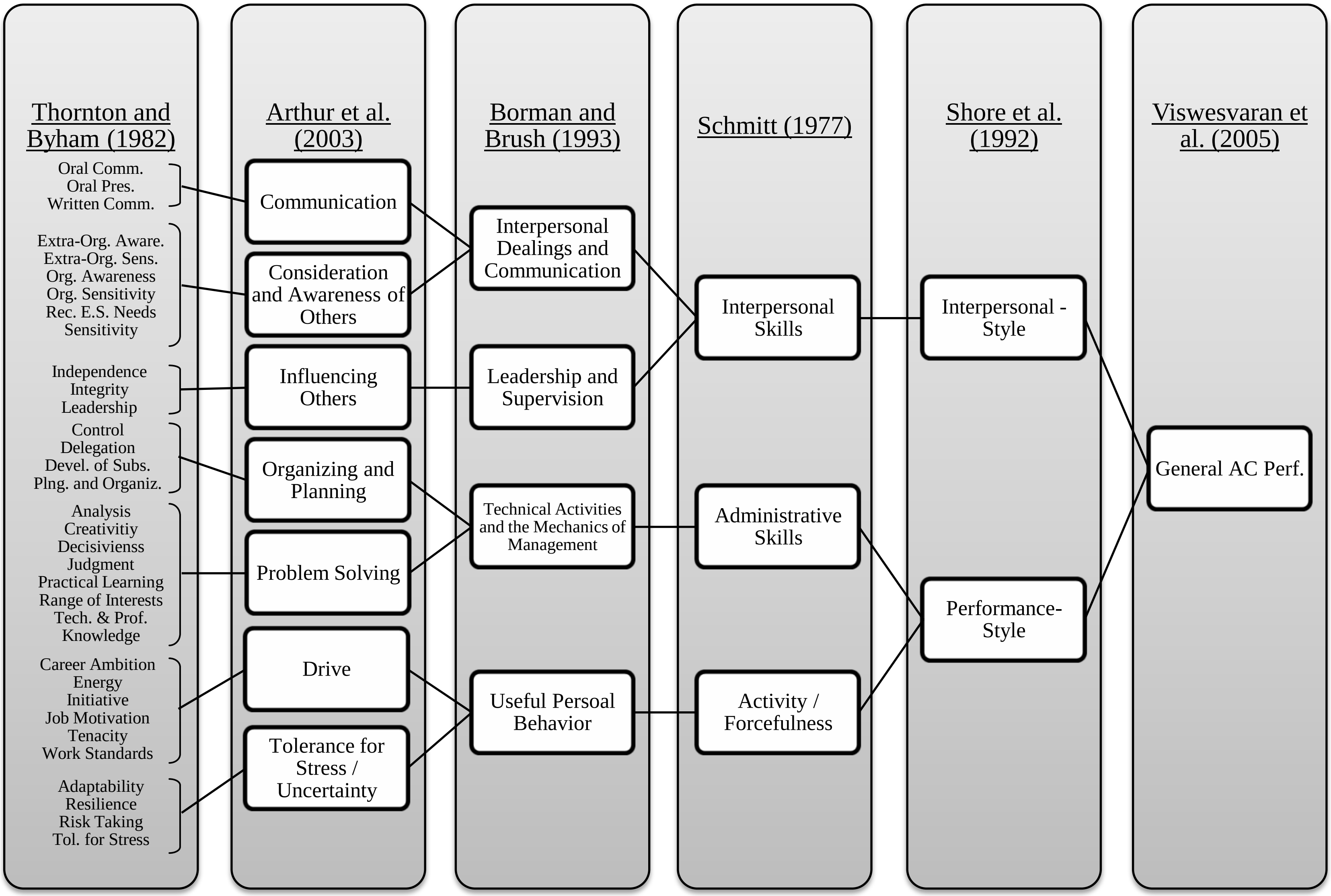


Figure 1. *Summary of Models to Examine and Proposed Nesting of Assessment Center Dimensions*

CHAPTER 4

METHOD

Step 1: Meta-Analytic Examination of AC Dimensions

The primary objective of the present study was to examine what constructs are being measured in ACs. To achieve this objective, the effect sizes of dimension intercorrelations reported in primary studies were synthesized using meta-analysis using the Hunter and Schmidt (2004) approach. In the search procedure, any primary studies that utilized the AC method were evaluated to make the results as generalizabe as possible. In searching for studies, the following databases were used: *PsycINFO*, *Web of Science*, and *Business Source Premier*. The following search terms were used in these databases to identify studies: assessment center, AC, dimension ratings. In addition, the reference lists of previous AC meta-analyses were examined, and authors of studies that only presented partial data were contacted to obtain the necessary information from their studies.

AC information has typically been conceptualized at three levels: 1) the overall assessment rating, or 'OAR', 2) post-consensus dimension ratings (PCDRs), and 3) within-exercise, or post-exercise dimension ratings (PEDRs). The present study evaluated results at the PCDR-level. This level of conceptualization was chosen for two key reasons: 1) this is the original intended level of analysis in typical ACs, and 2) PCDRs are fundamentally different from PEDRs, as PCDRs contain information that is sampled from multiple exercises and integrated via clinical judgment or some form of mechanical combination.

Criteria for Inclusion

Specific decision criteria were used to decide whether studies would be included in the meta-analysis. Primary studies must have: 1) provided dimension intercorrelations or values that could be converted into correlations, 2) these dimension intercorrelations must have been based on PCDRs (e.g., they could not be within-exercise dimension ratings), 3) dimension labels must have been provided and must have been able to be categorized into Thornton and Byham's (1982) list of common dimension labels, and 4) the studies must have reported the size of the sample on which dimension intercorrelations were computed. These inclusion criteria were chosen to remain consistent with the original ACs (e.g., Bray & Grant, 1966) as well as current research (Arthur et al., 2003).

From the primary studies, correlations (or values that can be converted to correlations) were recorded, along with the study's sample size. To provide a common basis for primary dimension labels, Thornton and Byham's (1982) list of common AC dimension labels was used as a framework for classifying primary study dimensions (See Table 10, Appendix B). These primary dimension labels have been utilized in previous studies involving conceptual groupings of AC dimensions (e.g., Arthur et al., 2003). As is common in ACs, slight variations on dimension labels often emerge (e.g., 'stress tolerance' versus 'tolerance for stress'). Hence, dimensions were coded into a common framework to develop the correlation matrix among all available AC dimension labels.

Primary Study Characteristics

The initial search resulted in a total of 574 studies that were further reviewed for inclusion. Each study was evaluated based on the criteria for inclusion, and the authors of these studies were contacted to request additional information if only partial information was presented (e.g., correlations between AC dimensions and external variables but not dimension intercorrelations). After evaluating whether studies contained relevant information, a total of 42 studies were identified that could be included in the analyses (See Appendix C). As several of these studies contained multiple samples (i.e., study 1 + study 2), 57 individual samples were identified. These studies evaluated an average of 13.34 dimensions (*Mdn* = 12) measured in an average of 5.53 exercises (*Mdn* = 5) (See Table 2). The mean number of participants in each study was 378.72 (*Mdn* = 156, *Total N* = 21,587).

Agreement Among Dimension Classifications and Final Data Set

Each data point was coded by at least two independent researchers. All dimension intercorrelations reported in primary studies were coded into a data file with their corresponding correlation coefficient. Definitions of dimensions were reviewed, and dimensions were then coded by the researchers into Thornton and Byham's (1982) dimension labels. The researchers initially agreed on 98% of the dimension classifications, and the remaining discrepancies were resolved by discussing each decision prior to running any analyses. In instances where the researchers could not come to agreement on how it should be classified, the dimension was excluded from the analyses. When recoded by the researchers, studies in the final data set reported an average of 9.51 dimensions (*Mdn* = 10).

Analyses

Based on the content sorting of Thornton and Byham's (1982) dimension labels into the AC models described earlier, the meta-analysis procedures developed by Hunter and Schmidt (2004) were employed, and sample-weighted mean correlations were computed using the SAS

Table 2. Summary of Primary Study Characteristics

	Number of Exercises	Number of Dimensions (Pre)	Number of Dimensions (Post)
Mean	5.53	13.25	9.51
SD	1.76	5.98	3.25
Median	5	12	10

Note. Pre = summary information as reported directly in the primary study; Post = summary information after being categorized into Thornton and Byahm's (1982) dimensions.

PROC MEANS syntax developed by Arthur, Bennett, and Huffcutt (2001). Although this meta-analytic procedure typically involves a correction for multiple statistical artifacts such as sampling error, measurement error and range restriction, in the present study corrections were only made for sampling error (i.e., sample size).

The meta-analytically derived correlation coefficients were used to construct a correlation matrix among the AC dimensions mentioned above (See Table 12, Appendix E). This matrix essentially represents the correlations among all of the recoded dimension labels. In total, this matrix is composed of 136 meta-analytic correlation coefficients. These values were derived to serve as input for the CFA analyses.

When two primary study dimensions were identified as representing the same dimension in Thornton and Byham's (1982) list, the square root of the average correlation among these two labels was used to compute a reliability estimate. Essentially this is a version of an alternate-forms reliability index (i.e., leadership = .80). These values are reported on the diagonal in Table 12 (Appendix E). Although these values were not used to correct for attenuation in the meta-analysis procedure, they can be used to make corrections in the CFA (confirmatory factor analysis) model. In addition, these values may be beneficial in resolving model identification issues if they arise.

As described by Viswesvaran and Ones (1995), the combination of meta-analysis with covariance structure analysis offers the opportunity to conduct a CFA or evaluate a structural equation model (SEM) with the data. Once meta-analytic estimates were derived, CFA was used to examine evaluate the fit of each a priori model to the data by using the meta-analytic correlation matrix as input. These dimension intercorrelations served as indicators for the a priori models described above, and allowed for a comparison to provide answers to research questions 1 and 2. These analyses were conducted using LISREL 8.70 (Jöreskog & Sörbom, 2004), and models were compared using several model-data fit indices (Jöreskog, 1993). As recommended by Viswesvaran and Ones (1995), the harmonic mean (918) of the sample sizes for the individual mean correlations was used as the sample size for the subsequent analyses.

Models were first evaluated to determine whether they converged to an admissible solution. Obtaining a proper solution is a key requirement for evaluating model fit; specifically, a lack of convergence or convergence to an improper solution often indicates that the model in question is inconsistent with the data or that model identification problems are present (Marsh, 1989). Next, overall model fit was examined by comparing the relative fit across models. Seven goodness-of-fit indices were examined: the chi square (χ^2) model fit test statistic, Steiger's (1990) root mean square error of approximation (RMSEA), Bentler and Bonnett's (1980) normed fit index (NFI), Bentler and Bonnett's (1980) non-normed fit index (NNFI), the comparative fit index (CFI; Bentler, 1990), James, Mulaik and Brett's (1982) parsimonious normed fit index (PNFI), and Browne and Cudeck's (1989) expected cross validation index (ECVI).

The χ^2 test for goodness of fit is the most conservative of the chosen fit indices, and is essentially a test of perfect fit. As this value has a known distribution, significance tests can be conducted. However, this value is sensitive to sample size, and is rarely used in isolation, since it will often reject anything other than perfect fit (Brown, 2006). RMSEA (Steiger, 1990) provides a test that makes an adjustment for model complexity (i.e., impacted by degrees of freedom). Values of 0.05 or less indicate a close fit to the data, and values above 0.08 are out of acceptable range (Browne & Cudeck, 1993), where smaller values indicate closer fit. Bentler and Bonnet's (1980) NFI is an incremental fit index, such that it compares the fit of the proposed model to a baseline model (i.e., the independence model); however, the NFI has been shown to underestimate fit in small samples (Byrne, 2001). The CFI makes an adjustment to the NFI based on sample size. It is an incremental measure of fit relative to a null model; CFI values can range from 0.0 to 1.0, where values of 0.90 or greater indicate an acceptable level of fit. Although the NNFI's values can fall outside of the range of 0 to 1.0, values are typically interpreted in the same manner as the CFI and NFI (Brown, 2006), where larger values indicate better fit, and values above .90 are generally acceptable. James et al.'s (1982) PNFI represents another parsimony-adjusted approach toward model fit, where model complexity is taken into account by adjusting the NFI by a parsimony ratio. In general, larger values represent better fit. Finally, the ECVI indexes whether the sample would cross-validate to a sample of similar size (Browne & Cudeck, 1989). The ECVI takes into account both model fit and the number of parameters used. Although the ECVI can take on any value, they can be compared in size where smaller values indicate better fit, as well as examined using confidence intervals (Byrne, 2001). Models were compared by evaluating this set of model-data fit indices to determine which model best explains the structure of AC dimensions. Hence, research questions 1 and 2 were answered in this manner.

Step 2: Grouping of Lower-Order Dimensions into Higher-Order Dimensions

The third objective in this study involved the classification of AC dimensions into existing frameworks (i.e., the a priori models). This analysis revealed information regarding how reliably AC dimensions could be grouped into these categories by subject matter experts (SMEs).

This process involved classifying Thornton and Byham’s (1982) list of AC dimensions into the taxonomies reviewed above (See Appendix G). Some of these models were derived in primary studies that examined the factor structure of AC dimensions within single samples, often from an exploratory perspective (e.g., Schmitt, 1977), where others were derived in large-scale studies using content sorting procedures (e.g., Arthur et al., 2003). The models included in the present study were selected on their capacity to serve as potential frameworks for grouping primary (i.e., lower-order) dimensions into higher-order constructs that emerge from these primary dimensions.

Dimensions from Thornton and Byham’s (1982) list were classified into the models listed in Table 1 (with the exception of a single-factor model) by eight SMEs. SMEs were selected based on their experience with AC research and practice. SMEs had all been formally trained in at least a four-day frame-of-reference training session. SMEs had an average of 3 years of assessment experience (min = 2, max = 5) and had worked with ACs in both selection as well as developmental contexts. Each SME was instructed to group the dimensions presented in Thornton and Byham’s (1982) list into the dimensions of the various performance models.

The rating forms and instructions given to raters are presented in Appendix G. The sorting process involved providing each SME with a form containing labels and boxes for each dimension of the higher-order model, and they were instructed to sort the common dimensions from Thornton and Byham’s list into these categories. In case they perceived that a dimension did not belong to any of the broader categories, the forms contained an ‘unclassifiable’ box where they could sort such dimensions. Each of the SMEs completed the rating task for all five models. The level of agreement among SMEs was assessed by evaluating the relative frequency in which they placed Thornton and Byham’s (1982) dimensions into the broader categories in each a priori model.

CHAPTER 5

RESULTS

Meta-Analysis Results

The meta-analytically-derived dimension intercorrelations are presented in Table 11 (Appendix D). This 17 x 17 matrix of dimension intercorrelations was used as input for the analyses conducted to answer the research questions. Specifically, CFA was applied and the fit of each model to the data was examined. In essence, intercorrelations were present for over half of Thornton and Byham's (1982) 33 dimensions. Some of the values were unable to be computed because they were simply not reported in any of the studies identified. The missing values are represented by an X in Table 12 (Appendix E). Although several of Thornton and Byham's (1982) dimensions were not included in the CFA analyses, the final matrix represents enough of the dimension intercorrelations to evaluate the a priori models. Each of the analyses and the results are discussed in the following sections.

Research Question 1: Empirical verification of Arthur et al.'s (2003) model

The first objective of this study was to evaluate whether Arthur et al.'s (2003) framework fit the data as an acceptable model for explicating the constructs underlying AC dimension ratings. To address this research question, the CFA model shown in Figure 2 was tested (Appendix F). Each of the Thornton and Byham (1982) dimensions loaded onto one of Arthur et al.'s (2003) seven AC categories as proposed in their initial study. To allow the model to be fully identified, one of the loading weights from a manifest indicator to a latent factor had to be constrained to a set value. This was necessary since only one manifest indicator loaded onto the latent factor. More specifically, only one indicator (leadership) was present for Arthur et al.'s (2003) influencing others factor. The path coefficient from the exogenous latent variable (i.e., influencing others) to its single manifest indicator (i.e., leadership) was constrained to .80 and the loading of the disturbance term on the manifest indicator to .20. This value was obtained by using a parallel forms type of reliability based on the average intercorrelation of values that were coded as 'leadership'. Otherwise, the remaining six of the seven dimensions posited by Arthur et al. (2003) had multiple indicators and were freely estimated. The only other constraints placed on the model were fixing one of the loadings for each latent variable to 1.0 to allow the model to converge (i.e., reference indicators).

Arthur et al.'s (2003) model fit the data very well. Overall, the fit indices were all within acceptable ranges (χ^2 = 316.29, *df* = 99; RMSEA = 0.049; CFI = 0.96; NNFI = 0.95; See Table 2). Based on the results of this CFA model, these results suggest that Arthur et al.'s (2003) framework provides a good representation of the relationship among AC dimensions. In other words, these seven categories explain covariance among the lower-order dimensions very well. Further, Arthur et al.'s (2003) framework serves as an acceptable baseline model to compare the alternative a priori models against.

Research Question 2: Comparison of alternative models

The second objective of this study was to examine whether Arthur et al.'s (2003) model is the best-fitting model, or whether one of the alternative models better fits the data. To address this research question, each of the alternative a priori CFA models presented in Figure 1 were tested as a first-order factor model (i.e., the 17 indicators loaded directly onto the latent factors). These CFA models and their factor loadings are shown in Figures 3 – 6 (Appendix F).

First, all of the alternative models fit the data well in absolute terms (See Table 3). Specifically, the χ^2, RMSEA, CFI, and NNFI all fell within acceptable ranges, based on rules of thumb. In evaluating each model in isolation, the reported fit indices would indicate that each one provides an acceptable fit to the data. However, in evaluating what constructs underlie AC dimension ratings, a comparison of alternative models provides a more thorough explanation of observed covariance in AC ratings (Bollen, 2000). Arthur et al.'s (2003) model had the best χ^2 (316.29, df = 99), its RMSEA (0.049), CFI (0.96) and NFI (0.95) were the same as the next-best fitting models, and its NNFI did not improve when the number of factors was reduced. In comparing the models to each other, the ECVI was best for the seven-factor model (ECVI = 0.46), and worst for the two-factor (ECVI = 0.53) and one-factor (ECVI = 0.53) models. However, the PNFI, which provides an adjustment for model parsimony, indicated that the one-factor model fit the best. Further, the biggest increase in the PNFI was evident in comparing the seven-factor (PNFI = 0.69) and four-factor models (0.79). In further reducing the number of factors, the PNFI changed 0.01 between the three-factor, two-factor, and one-factor models, respectively.

Taken together, it appears as though model fit is not substantially improved by a further reduction in the number of dimensions. Specifically, Arthur et al.'s (2003) model had the best χ^2, ECVI, and NFI values. Further, the RMSEA, NNFI, and CFI values were the same as the next best values for the alternative a priori models. Only the PNFI improved as the number of dimensions was reduced. However, this should be expected as PNFI makes an adjustment for

Table 3. Model-Data Fit Indices for First-Order CFA Models

Model	χ^2	*Df*	χ^2 / *df*	RMSEA	ECVI	90% CI	CFI	NFI	NNFI	PNFI
7-Factor	316.29	99	3.19	0.049	0.46	0.41; 0.53	0.96	0.95	0.95	0.69
4-Factor	374.60	114	3.29	0.050	0.49	0.43; 0.56	0.96	0.94	0.95	0.79
3-Factor	374.61	116	3.23	0.049	0.49	0.43; 0.56	0.96	0.94	0.95	0.80
2-Factor	411.98	118	3.49	0.052	0.53	0.46; 0.60	0.95	0.94	0.95	0.81
1-Factor	415.59	119	3.49	0.052	0.53	0.46; 0.60	0.95	0.94	0.95	0.82

Note. Each of the models tested represent the sequence in Figure 1: 7-Factor represents Arthur et al.'s (2003) model, 4-Factor represents Borman and Brush's (1993) model, 3-Factor represents Schmitt's (1977) model, 2-Factor represents Shore et al.'s (1992) model, and 1-Factor represents a unidimensional model (i.e., Viswesvaran et al., 2005).

model parsimony. In short, Arthur et al.'s (2003) model appears to provide the best fit to the data and may best represent the constructs underlying AC dimension ratings.

Post-Hoc Analyses / Hierarchical Models

Since all of the a priori models fit very well when tested as first-order models, a post hoc-analysis was conducted to determine whether one of several hierarchical models fit the data better than the first-order seven-factor model. In other words, might the addition of a higher-order set of latent factors improve the fit of the model (i.e., explain variance in the seven factors)? To answer this question, each of the alternative models were examined as a set of higher-order factors that might explain variance in Arthur et al.'s (2003) seven factors. These models are presented in Appendix I. Results of these analyses are presented in Table 4.

In testing these models, an adjustment had to be made to the four-factor Borman and Brush (1993) higher-order factor model. Since the sole indicator for the higher-order 'leadership and supervision' factor was Arthur et al.'s (2003) 'leadership' factor, this factor served as a lower-order factor that did not correlate with the other three higher-order factors. In order to construct a fully-identified model, the model presented in Figure 7 (Appendix I) was evaluated.

Of these models, the hierarchical three-factor Schmitt (1977) model generally fit the data as well as the Arthur et al. (2003) seven-factor first-order model. Each of the remaining models fit worse than these two. More specifically, the first-order Arthur et al. (2003) seven-factor model had a smaller χ^2 value when compared with all of the others models. However, this value was not significantly different from the three-factor hierarchical model ($\Delta\chi^2 = 21.05$, $\Delta df = 12$, $p = .05$). In addition, the hierarchical three-factor model had a smaller RMSEA (0.047) as well as a smaller χ^2 / *df* ratio (3.07) than the seven-factor first-order model (3.19). Further, the ECVI, CFI,

Table 4. Model-Data Fit Indices for Second-Order / Hierarchical CFA Models

Model	χ^2	*Df*	χ^2 / *df*	RMSEA	ECVI	90% CI	CFI	NFI	NNFI	PNFI
7-Factor†	316.29	99	3.19	0.049	0.46	0.41; 0.53	0.96	0.95	0.95	0.69
4-Factor	529.13	111	4.77	0.064	0.67	0.59; 0.75	0.93	0.91	0.91	0.75
3-Factor	337.34	110	3.07	0.047	0.46	0.41; 0.53	0.96	0.95	0.95	0.77
2-Factor	371.97	112	3.32	0.050	0.50	0.40; 0.56	0.96	0.94	0.95	0.78
1-Factor	372.99	113	3.30	0.050	0.49	0.43; 0.56	0.96	0.94	0.95	0.78

Note. †7-Factor represents Arthur et al.’s (2003) model as tested above, presented here for ease of comparison. Each of the hierarchical models tested represent a set of higher order factors the explain variance in Arthur et al.’s (2003) 7-factor model.

NFI, and NNFI were all the same for these two models. In general, the hierarchical three-factor model appears to fit slightly better, if not the same as the seven-factor first-order model. Hence, this model might serve as a useful framework for conceptualizing the constructs underlying AC dimensions.

Research Question 3: Classification of dimensions by SMEs

The third objective of this study was to provide additional evidence for how well experienced raters can sort primary dimensions into broader categories (i.e., the a priori models). To address this question the relative frequency in which the SMEs categorized each of Thornton and Byham's (1982) dimensions into the broader categories of each of the a priori models was evaluated. Here, the cutoff agreement level was 75% (i.e., if the 8 SMEs categorized the dimension into the same category at 75% or greater). Results are presented in Table 3. For the relative frequency of classifications, see Tables 13 – 16 (Appendix H).

Arthur et al.'s (2003) seven-factor model provided the greatest number of agreed-upon dimension classifications. Specifically, SMEs categorized dimensions into the same category for 23 out of the 33 dimensions. For the four-factor and two-factor models, SMEs categorized 21 out of the 33 dimensions into the same category, and for the three-factor model only 14 of the 33 dimensions were categorized into the same category by SMEs. Hence, these results suggest that Arthur et al.'s (2003) model facilitated the greatest level of agreement among SMEs in how they perceived that dimensions grouped together. Overall, it appears that Arthur et al.'s (2003) dimensions provided SMEs with a more useful framework for categorizing lower-order dimensions.

Table 5. SME Agreement for Dimension Classifications

Model	Number of Dimensions at 75% Agreement (%)
Arthur et al. (2003) 7-Factor	23 (70%)
Borman and Brush (1993) 4-Factor	21 (64%)
Schmitt (1977) 3-Factor	14 (42%)
Shore et al. (1992) 2-Factor	21 (64%)

At a 75% level of agreement, the results provide additional support for the usefulness of the seven-factor Arthur et al. (2003) model, as it allows for the highest level of agreement in the classification of lower-order dimensions. Hence, these results demonstrate additional evidence for the usefulness of Arthur et al.'s (2003) seven dimensions in terms of how useful they are for raters, such that SMEs can most effectively categorize primary dimensions into this framework.

Summary of Results

In summary, the results support the usefulness of Arthur et al.'s (2003) framework as a set of constructs that underlie AC dimension ratings. Empirically, this model fit the data well based on the model-data fit indices. Also, in comparison with the alternative models, Arthur et al.'s (2003) model fit the data the best. In addition, supplemental analyses show that a higher-order model with three broader factors fit the data as well (if not slightly better) than Arthur et al.'s (2003) first-order seven-factor model. Further, SMEs were able to best classify primary dimensions into these seven dimensions in comparison with the alternative models. Overall, these results provide additional evidence for how these broader categories may represent the constructs underlying AC dimension ratings.

CHAPTER 6

DISCUSSION

Summary of Findings

The AC method has been popular for years, and for several reasons. Large-scale studies have demonstrated that ACs predict work performance (Arthur et al., 2003; Gaugler et al., 1983), and do so above and beyond commonly used paper-and-pencil predictors (Meriac et al., in press). In addition, applicants have shown more positive reactions to ACs in comparison with other common predictors (Macan et al., 1994). Recently however, the AC method has been criticized for several reasons, most of which center on construct-related validity evidence for dimensions (Lance et al., 2004).

Until the early 1980's, AC researchers conducted several studies with the purpose of examining the constructs underlying dimension ratings, largely taking an exploratory approach (i.e., exploratory factor analysis). These studies were carried out to determine if broader categories or "constructs" emerged from dimension ratings, and several studies (e.g., Schmitt, 1977) revealed that broader latent categories did explain variance in dimension ratings. This line of inquiry shifted as the MTMM approach was applied to within-exercise dimension ratings (Sackett & Dreher, 1982), with a focus on whether dimension or exercise effects were stronger in PEDRs. For the most part, these studies concluded that observed variance is attributable to exercise effects more than dimension effects, calling into question the construct-related validity of AC dimension ratings.

However, despite over 20 years of examining AC dimensions using MTMM analyses, no consensus has been reached regarding AC construct-related validity issues. Some researchers have recently called into question the use of the MTMM analytic design's applicability toward AC research altogether, as AC exercises were never designed to 'equally' measure all dimensions, nor are dimensions the same as 'traits' (Howard, 2008; Lance et al., 2007). Furthermore, a logical prerequisite has been neglected in MTMM-oriented studies, in that AC dimensions have typically been taken at face value, without a proper development or exploration of underlying constructs (Arthur et al., 2008). More specifically, the constructs that should exhibit validity evidence have never been clearly articulated.

This study's core purpose was to provide some clarification on what constructs underlie AC dimension ratings. Toward this end, an approach similar to that taken by early AC researchers was used, where broader latent categories were examined as possible constructs underlying dimension ratings. However, these earlier studies were almost exclusively conducted on primary samples, and typically took an exploratory approach. This study took a large-scale focus by using meta-analysis, and integrated several a priori models in a confirmatory approach (i.e., using CFA). Several models from both the AC literature as well as the general job performance literature emerged as possible frameworks for explaining the constructs that underlie AC ratings. An examination of several alternative a priori models in this context represents the first attempt to integrate multiple theoretical models with respect to AC dimensions. Further, SMEs were asked to group primary AC dimensions into each of the a priori models based on dimension definitions. The classification of dimensions by SMEs into higher-order categories sheds light on how assessors think about the constructs that are measured in

ACs (i.e., schemas of how dimensions group together), and how they might treat these categories if they were used to group primary dimensions. These two approaches provide different types of information, yet they contribute to our understanding of what constructs underlie ACs.

The CFA results revealed that Arthur et al.'s (2003) seven-category framework fit the empirical data very well. In addition, a comparison of these models revealed that Arthur et al.'s (2003) framework provided the best fit to the empirical data when compared with the alternative models. These results compliment the findings of previous studies and suggest that this model may serve as a good representation of the constructs underlying AC dimension ratings. Both Arthur et al. (2003) and Meriac et al. (in press) used Arthur et al.'s (2003) framework to examine the predictive validity of AC dimensions (i.e., constructs). The results of both studies showed that these categories are strong predictors of work performance. Bowler and Woehr (2006) evaluated the construct-related validity of AC dimensions by using six of these categories in a meta-analytic MTMM design, and found that, in general, this model improves upon the common problem where larger 'exercise effects' emerge in comparison with 'dimension effects' in primary studies. Specifically, recent studies (e.g., Lance et al., 2004) have shown that exercise effects are much larger than dimension effects, but Bowler and Woehr demonstrated that these effects are roughly the same size. Hence, these results help clarify what constructs underlie AC dimension ratings by providing additional support for Arthur et al.'s (2003) seven dimensions.

Further, this study provides additional evidence for how raters can use these constructs. Specifically, the results demonstrate that AC raters can more easily group primary dimensions into Arthur et al.'s (2003) seven categories than the categories of the other alternative frameworks. This study differs from previous research in that although Arthur et al. (2003)

grouped Thornton and Byham's (1982) dimensions into higher-level categories, they did not directly evaluate the level of agreement among SMEs. This study provides additional information about how experienced AC raters view primary dimensions as belonging to broader categories. As many AC studies contain potentially redundant dimension labels, the capacity for raters to utilize a summary framework for reporting this information is important. In addition, the extent to which raters agreed upon dimension classifications suggests how raters perceive these dimensions as grouping together, such that raters may be to some extent predisposed to use this framework for the classification of dimensions. This is important for underscoring the usefulness of this framework for existing ACs.

In addition to demonstrating additional support for Arthur et al.'s (2003) model, this study also tested a set of hierarchical models to determine whether one of these additional models provided a better fit to the meta-analytic data. These post-hoc analyses revealed that a set of three higher-order factors, based on Schmitt's (1977) framework, serves as a more parsimonious set of latent variables for explaining variance in AC dimension ratings. This model may serve as a viable alternative framework for offering feedback to assessees or conveying information to managers. Although Arthur et al.'s (2003) model fit the best as a first-order model, these three higher-order latent factors may serve as an alternative approach for providing feedback when summary information at an even more general level is helpful. This set of higher-order factors fit better than a 4, 2, or 1-factor higher-order model, indicating that although a more parsimonious set of categories did have good model-data fit, a 2 or 1-factor model is simply too general to be useful in the AC context. Hence, although Arthur et al.'s (2003) model did have the best fit as a first-order model, a set of summary dimensions may also be useful.

Implications

Foremost, these results provide some clarification on the constructs underlying AC ratings, as they offer evidence for the notion that a set of latent variables explains variance in AC dimensions. In line with previous research (Arthur et al., 2003; Bowler & Woehr, 2006; Meriac et al., in press), this study provides additional evidence for Arthur et al.'s (2003) dimensions as a set of constructs that underlie AC dimension ratings. Hence, these results offer important empirical support for the factor structure of these seven categories. Given the lack of development of constructs in AC research, this is valuable information that may foster a greater understanding of what exactly is being measured in ACs.

These findings are important for several reasons. In particular, constructs are the foundation of psychological science (Arthur et al., 2008; Landy, 1986), and theory is typically discussed at this level. According to Binning and Barrett (1989) and reviewed by Arthur and Villado (2008), validity itself ultimately represents a series of inferences regarding the linkages between constructs in different domains (e.g., cognitive ability and job performance). Ignoring the constructs that are operating in ACs (or taking them at face value) has been a problem with ACs for years (Arthur et al., 2008). This study provides important information and is intended to address this issue. Studies that have evaluated the 'validity' of ACs have often treated primary dimensions as generic 'constructs' or made validity inferences about a method. By considering the constructs that operate in ACs, inferences made regarding their validity are more appropriate as the constructs that are operating will have been more clearly delineated to begin with. Further, taking a construct-centered approach, it may be possible to better develop AC theory and gain a clearer understanding of how these AC constructs relate to other commonly-used (and more rigorously-developed) predictors (e.g., personality and mental ability variables).

In addition to improving upon validity evidence for the constructs that ACs measure, discussing information at the construct-level is important for making comparisons with other predictors. A frequent topic in personnel selection research is how the use of some predictors results in disparate impact for applicant subgroups. For example, cognitive ability is generally regarded as the strongest predictor of work performance (Schmidt & Hunter, 1998). However, results of several studies have demonstrated that it results in disparate impact for minority groups. With few exceptions (e.g., Dean et al., 2008), ACs are regarded as strong predictors of performance, yet they do not result in disparate impact. The construct-method confusion in the AC literature (Arthur & Villado, 2008) has resulted in a comparison between disparate impact caused by the use of a *construct* (cognitive ability) and a *method* (ACs). A more meaningful discussion of how predictors operate should not confuse construct and method, or at least separate these two sources of variance. Hence, by taking a construct-centered approach to AC ratings, researchers could explore subgroup differences on scores of ‘organizing and planning’ or ‘influencing others’ rather than ‘AC ratings’, or perhaps whether different methods of measuring these and other constructs results in disparate impact or not.

These results also highlight the usefulness of Arthur et al.’s (2003) model as a categorizing framework for information typically reported at the primary dimension-level. Specifically, AC practitioners could take primary dimension ratings and use this framework for grouping existing dimensions for the purpose of providing feedback as well as a mechanism for conveying information to managers and other organizational decision makers. Hence, these dimensions are useful for many ACs in their existing form, and most ACs may not necessarily need to be re-designed to utilize the findings of this study. As demonstrated by Bowler and Woehr (2006), utilizing this framework in such a manner has shown improvement in the convergent and discriminant validity of AC ratings.

AC designers could however use this framework as a starting point to design new ACs by ensuring that all of these constructs are in some way measured, at least if they wish to tap the full content domain of common ACs. This would essentially entail designing new ACs around the measurement of constructs. Additionally, ACs that do not encompass this full content domain could be re-designed so that these constructs are adequately represented in the behavioral exercises. Taking this approach may improve upon the measurement properties of ACs by at least ensuring that some constructs are measured before making further inferences regarding their validity (e.g., criterion-related validity).

In addition, a set of three higher-order factors may serve as a viable means for conveying AC ratings to managers (i.e., an abbreviated summary). Rather than presenting managers who will be making selection or promotion decisions with a long list of several dimensions, information could be provided in more parsimonious manner based around Schmitt's (1977) three factors. Managers may be able to more effectively process information when presented with a more parsimonious set of broader dimensions. This could be particularly helpful when comparing the performance of multiple assessees on multiple dimensions. It is important to underscore that especially with the higher-order three-factor model, the results of this study indicate that primary dimensions should only be grouped into these categories, not measured directly through three factors, as Arthur et al.'s (2003) model fit better than a three-factor model

(Schmitt, 1977) when these were tested as first-order models. The three-factor model only fit better when tested as a hierarchical model. This is not to say that if raters are trained to think of dimensions as belonging to higher-order categories this may improve convergent and discriminant validity.

This study is the first to provide an empirical verification of how well Arthur et al.'s (2003) framework fits the reported AC data. In addition, this study is the first to compare how well Arthur et al.'s (2003) model fits in comparison with alternative frameworks. Further, until now no study has provided a direct evaluation of how well trained and experienced raters can utilize each of these frameworks for categorizing primary dimensions. Overall, the results of this study have several implications for AC research and practice. Providing clarity on the constructs operating in ACs may help guide AC theory by fostering a discussion of ratings at a proper level of conceptualization.

Limitations

Despite the implications of these findings, there are several limitations that must be discussed. One shortcoming of the present study is that all of Thornton and Byham's (1982) dimension intercorrelations could not be included in the CFA analyses. More specifically, from the MA results, 17 of the 33 dimensions were able to be used in the analyses, yet 16 could not. If more dimension intercorrelations had been available, the results may have provided a more rigorous examination of how these models compare. As additional AC data become available, it is possible that several more dimension intercorrelations can be reported and included. Still, with multiple manifest indicators for all but one of Arthur et al.'s (2003) dimensions, this study provided a satisfactory test of this model and helps increase our understanding of how these constructs operate.

An additional concern became apparent as the dimensions were coded into Thornton and Byham's (1982) list: Many of the dimensions that they listed were simply not measured (or at least reported) by ACs at all. Thornton and Byham's (1982) list of commonly used AC dimensions might be improved by including more interpersonal and influence-oriented dimensions. A refinement of this list may be helpful for AC researchers and practitioners by more clearly specifying the dimensions that should be grouped into broader categories (i.e., constructs). Even if researchers' and practitioners' primary objectives are to group dimensions into Arthur et al.'s (2003) categories, increased specificity in Thornton and Byham's (1982) list of primary dimensions would aid in the task of grouping primary dimensions. In addition, it is possible that a modified list of commonly used AC dimensions may impact the fit of these models when different dimensions are used as the manifest indicators. However, in line with previous research (i.e., Arthur et al., 2003), the present study utilized the same classification taxonomy as Arthur et al., therefore providing results that are more easily comparable with their findings.

An additional concern arose in the evaluation of how well SMEs were able to group primary dimensions into the a priori models. Specifically, raters had the lowest level of agreement when using the three-factor (Schmitt, 1977) framework. When categorizing dimensions into this three-factor framework, this model resulted in the fewest number of dimensions classified into the same categories by SMEs. Hence, raters might not be able to effectively utilize this model as well for grouping dimensions into broader categories (i.e., they may not use this AC performance schema). Considering all of the information together, there is still some uncertainty as to how well a hierarchical model may operate for AC raters, even though a model with three higher-order factors fit the data as well as the seven-factor model. Hence, the usefulness of a three-factor hierarchical model as a higher-order summary framework must be evaluated further if AC administrators would like to use this summary framework to convey information to others (i.e., managers or assessees). The discrepancy between raters' agreement in their classifications and the results of the CFA necessitate further evaluation.

Although the raters that were asked to classify primary dimensions into the broader categories of the five frameworks were trained and experienced with both administrative and developmental ACs, they were all trained using the same rating approach. Specifically, they were trained in the common process where assessees are observed while they participate in exercises, behaviors are recorded and ratings are made on dimensions within exercises, and then these ratings are compiled in a consensus / discussion meeting where final dimension ratings are decided. Two variations on the rating approach are the original AT&T method (Howard, 2008) where ratings are only assigned after behavior is observed across all exercises, and the task-based approach (Jackson et al., 2005), where raters are trained with the intention of rating performance on exercises rather than dimensions. In addition, many rating procedures utilize behavioral checklists or behavioral observation scales (Spychalski et al., 1997), which is yet another variation of behavior-recording component of the AC rating process. Further, the extent to which raters are trained to utilize different primary dimension frameworks poses yet another consideration for how the raters who participated in the present study may be different from raters from other backgrounds in their dimension classifications. Hence, generalizability may be a concern with raters' dimension classifications. Future studies should include more diverse samples with raters from different backgrounds and evaluate whether any differences exist in how they classify primary dimensions.

Future Directions

The models tested in this study were general in nature (i.e., they are expected to apply across job settings). In addition, the use of meta-analytic data to empirically examine these models represents a more general examination of how they operate, as opposed to how they work in specific job contexts. Future studies should further evaluate how well these seven constructs operate in a general capacity. The data cumulated in the present study did not allow for an examination of moderator variables, but future studies may be able to test how these constructs operate *across* job settings. It is possible that some jobs that have a great deal more influence and interpersonal components to them (i.e., managerial work) may find more value in the more interpersonally-oriented constructs (i.e., influencing others). On the other hand, lower-level supervisory jobs may find more importance in constructs such as planning and organizing. Measurement invariance studies could help determine whether this same factor structure is applicable across different job settings (Vandenburg & Lance, 2000). In addition, the application of modern analytic techniques for determining relative importance such as dominance analysis (Budescu, 1993; Johnson & LeBreton, 2004) on different samples may reveal if some of these constructs are more important for different job types than others.

AC research and practice may also benefit from a closer examination of how these constructs improve AC psychometric issues when ACs are explicitly designed to measure them. Bowler and Woehr (2006) have already demonstrated that the use of these categories improves upon the convergent and discriminant validity evidence in ACs. However, these constructs have only been employed in a post-hoc context; more specifically, primary dimension labels have been collapsed into them, rather than using them directly. Although the results of the present study provide support for categorizing AC dimensions in their existing form into broader categories, using Arthur et al.'s (2003) categories as a starting point for new ACs may further allow for an evaluation of the usefulness of these constructs. Rather than designing an new AC with a long list of dimensions, researchers and practitioners could simply use Arthur et al.'s (2003) seven categories as the focal constructs of interest that are measured by the AC. As demonstrated by Woehr and Arthur (2003), ACs that measure fewer dimensions typically demonstrate greater convergent validity evidence. Also, reducing the number of dimensions measured in an AC can improve discriminant validity, as raters can more easily distinguish among fewer (i.e., broader) dimensions (Gaugler & Thornton, 1989). ACs that are designed to explicitly measure these constructs, as well as rater training that centers around this framework, may improve upon the measurement properties exhibited by ACs.

However, Arthur et al.'s (2003) dimensions may be best represented as the models in this study were constructed: A set of latent factors that explain variance in a set of observable dimension ratings. In other words, practitioners may still see value in the use of more narrowly-defined dimension labels for making ratings and classifying observed behaviors. Viewing primary dimensions as belonging to higher-order categories may be possible for providing more specific feedback as well as a more parsimonious set of constructs for other purposes (e.g., administrative). Such approaches have been utilized for years in psychological research and practice. For example, the five-factor model of personality constructs operate such that they are five broad constructs, yet they have facet-level information that provides more specific information about an individual's personality (Barrick & Mount, 1991). These AC dimensions may operate in a similar manner. Future studies should make a direct comparison between these two approaches and evaluate the effectiveness of designing ACs around each approach.

Another potential avenue for future research is an examination of how these dimensions operate in feedback contexts. Previously, Arthur et al.'s (2003) framework has only been evaluated in selection or administrative-oriented contexts (i.e., how well they predict job performance; Arthur et al., 2003; Meriac et al., in press). A broad four-factor model has been developed specifically for developmental ACs (Gibbons et al., 2006). A comparison of this model with Arthur et al.'s (2003) dimensions for feedback reactions and developability may be helpful to determine how well these constructs generalize to this context. It is possible that providing feedback using these broader categories may aid both assessees and feedback administrators by grounding performance feedback in a construct-oriented nature (as opposed to a loose configuration of several performance dimensions).

Regarding the cognitive representation of these constructs, future studies should further evaluate the AC performance schemas of raters. More specifically, analytical approaches such as multidimensional scaling (MDS) may provide additional evidence of how raters perceive primary dimensions to group together based on their similarity, and what constructs emerge based on similarity ratings. It is possible that an evaluation of rater schemas may reveal additional information about the constructs underlying AC dimension ratings. For instance, similarities or differences between the CFA results and MDS results may reveal ways in which these constructs could be further developed or modified to increase their convergence and potentially improve upon the psychometric properties of AC ratings.

This study has made an attempt at uncovering the constructs that underlie dimension ratings by integrating and examining several a priori models. It is apparent that Arthur et al.'s (2003) seven factor-model seems to best explain covariance among observed dimension ratings. However, much work remains if these dimensions are to approximate the empirical rigor of more mainstream psychological constructs (e.g., personality and cognitive ability variables). An additional step toward doing this is by taking an external approach toward construct validation (i.e., further developing the nomological network of these constructs). Meriac et al. (in press) compared these constructs with the big five personality factors and general mental ability, and found that these seven constructs shared a modest proportion of variance with these variables. However, there are other popular constructs that may relate to these AC dimensions in different ways. For example, more interpersonally-oriented constructs, such as respondents' ratings on transactional and transformational leadership (Bass, 1985) or social intelligence (Zacarro, 2002) may relate differently to these dimensions and further clarify the nomological network of these AC constructs. Also, these constructs have thus far only been examined as predictors of general job performance ratings (e.g., Arthur et al., 2003; Meriac et al., in press). It would be informative to examine how well these constructs predict multidimensional criteria, as well as extra-role performance (i.e., organizational citizenship behavior). A more thorough examination of these constructs' nomological network with additional individual difference constructs as well as additional criteria will further help clarify how these constructs operate.

In addition, using the MTMM design (Campbell & Fiske, 1959) in a more appropriate manner may actually improve upon our understanding of the constructs measured in ACs, however not in the way that it has been recently implemented in the literature. Rather than treating AC exercises as 'methods', it would be more appropriate to treat the AC itself as a method, and design other methods (e.g., situational judgment tests, situational interviews) to measures these same constructs (Rupp, Thornton & Gibbons, 2008). In this context, convergent and discriminant validity could be evaluated by comparing these seven constructs measured by ACs as well as other methods (Arthur & Villado, 2008). As this study has provided additional evidence that ACs measure constructs, future studies should continue this line of inquiry to help gather additional validity evidence.

Summary and Conclusions

In general, this study provides some clarity on the constructs that underlie AC dimension ratings. A rigorous examination was taken that incorporated multiple a priori models, spanning both the AC literature as well as the general job performance literature. In addition, two data sources were used to address the research questions. As reviewed by Arthur et al. (2008), when compared to more mainstream predictors in Psychological research, ACs have been deficient with respect to describing the constructs that are measured. Here this issue was addressed by integrating existing models from both the AC and general job performance literature. Overall, Arthur et al.'s (2003) framework appears to provide the best fit to the data, as well as the greatest ease in classifying dimensions for experienced AC raters. The results of this study provide additional evidence that these dimensions may be treated as relevant constructs for what is measured in ACs. Based on these results, it is possible to improve our understanding and use of the AC method by taking a construct-centered approach toward what they actually measure.

REFERENCES

Arthur, W. Jr., Bennett, W. Jr., & Huffcutt, A. I. (2001). *Conducting meta-analysis using SAS.* Mahwah, NJ: LEA.

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.

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.

Arthur, W. Jr. & Villado, A. J. (2008). The importance of distinguishing between constructs and methods when comparing predictors in personnel selection research and practice. *Journal of Applied Psychology, 93,* 435-442.

Barrick, M. R., & Mount M. K. (1991). The Big Five Personality Dimensions and Job Performance: A Meta-Analysis. *Personnel Psychology, 44,* 1-26.

Bass, B.M. (1985). *Leadership and performance beyond expectations.* New York: Free Press.

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

Bentler, P. M. & Bonnett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin, 88,* 588-606.

Bollen, K. A. (2000). Modeling strategies: In search of the holy grail. *Structural Equation Modeling, 7,* 74-81.

Borman, W. C., & Brush, B. D. (1993). More progress toward a taxonomy of managerial performance requirements. *Human Performance, 6,* 1-21.

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.

Brown, T. A. (2006). *Confirmatory Factor Analysis for Applied Research*. New York, NY: Guilford Press.

Browne, M. W. & Cudeck, R. (1989). Single-sample cross-validation índices for covariacne structures. *Multivariate Behavioral Research, 24,* 445-455.

Budescu, D. V. (1993). Dominance analysis: A new approach to the problem of relative importance of predictors in multiple regression. *Psychological Bulletin, 114,* 542-551.

Byrne, B. M. (2001). *Structural equation modeling with AMOS.* Mahwah, NJ: Lawrence Erlbaum and Associates.

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

Cascio, W. F., & Aguinis, H. (2005). *Applied psychology in human resource management* (6th ed.). New Jersey: Prentice Hall.

Dean, M. A., Roth, P. L., & Bobko, P. (2008). Ethnic and gender subgroup differences in assessment center ratings: A meta-analysis. *Journal of Applied Psychology, 93,* 685-691.

Fleishman, E. A. (1967). Performance assessment based on an empirically derived task taxonomy. *Human Factors, 9,* 349-366.

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

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

Gibbons, A. M., Rupp, D. E., Snyder, L. A., Hollub, A. S., and Woo, S. E. (2006). A preliminary investigation of developable dimensions. *The Psychologist-Manager Journal, 9*, 99-123.

Guilford, J. P. (1954). *Psychometric Methods* (2nd ed.). New York: McGraw-Hill.

Hinrichs, J. R. (1969). Comparison of "real life" assessments of management potential with situational exercises, paper-and-pencil ability tests, and personality inventories. *Journal of Applied Psychology, 53*, 425-432.

Hoeft, S., & Schuler, H. (2001). The conceptual basis of assessment center ratings. *International Journal of Selection and Assessment*, *9,* 114-123.

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

Howard, A. (2006). A commentary by Ann Howard. *The Psychologist-Manager Journal, 9,* 201-205.

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

Huck, J. R. (1973). Assessment centers: A review of the external and internal validities. *Personnel Psychology, 26,* 191-212.

Huck, J. R. & Bray, D. W. (1976). Management assessment center evaluations and subsequent job performance of black and white females. *Personnel Psychology, 29,* 13-30.

Hunter, J. E., & Schmidt, F. L. (2004). *Methods of meta-analysis: Correcting error and bias in research findings* (2nd ed.). Thousand Oaks, CA: Sage Publications.

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

James, L. R., Mulaik, S. A., & Brett, J. M. (1982). Causal analysis: Assumptions, models, and data. Beverly Hills, CA: Sage.

Johnson, J.W. & LeBreton, J. M. (2004). History and use of relative importance indices in organizational research. *Organizational Research Methods*, *7*, 238-257.

Jöreskog, K. (1993). Testing structural equation models. In K. A. Bollen & J. S. Long (Eds.), *Testing Structural Equation Models* (pp. 294-316). Newbury Park, CA: Sage Publications.

Jöreskog, K., & Sörbom, D. (2004). LISREL 8.70. Chicago: Scientific Software International Inc.

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., 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., 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.

Landy, F. J., Shankster-Cawley, L., & Moran, S. K. (1995). Advancing personnel selection and placement methods. In Howard, A. (Ed.), *The Changing Nature of Work* (pp. 252 – 289). San Francisco: Jossey-Bass.

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

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

Macan, T. H., Avedon, M. J., Paese, M., & Smith, D. E. (1994). The effects of applicants' reactions to cognitive ability tests and an assessment center. *Personnel Psychology, 47,* 715-738.

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

Meriac, J. P., Hoffman, B. H., Woehr, D. J. & Fleisher, M. S. (in press). 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*.

Norton, S. D. (1977). The empirical and content validity of assessment centers vs. traditional methods for predicting managerial success. *Academy of Management Review, 2,* 442-453.

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

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

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

Shore, T. H., Thornton, G. C., & Shore, L. M. (1990). Construct validity of two categories of assessment center dimension ratings. *Personnel Psychology, 43*, 101-116.

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

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

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

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

Vandenburg, R. J. & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods, 3,* 4-69.

Viswesvaran, C. & Ones, D. S. (1995). Theory testing: Combining psychometric meta-analysis and structural equation modeling. *Personnel Psychology, 48,* 865-885.

Viswesvaran, C. & Ones, D. S. (2000). Perspectives on models of job performance. *International Journal of Selection and Assessment, 8,* 216-226.

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.

Woehr, D. J., & Arthur, W. Jr. (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., Arthur, W. Jr., & Meriac, J. P. (2007). Method Factors Instead of Error Variance: Meta-Analysis of Assessment Center Construct Validity. In Schuler, H. (Ed.) *Assessment Center zur Potenzialanalyse.* Hogrefe-Verlag.

Zacarro, S. J. (2002). Organizational leadership and social intelligence. In R. E. Riggio, S. E. Murphy, & F. J. Pirozzolo (Eds.), *Multiple Intelligences and leadership.* Mahwah, NJ, Lawrence Erlbaum Associates.

APPENDICES

Appendix A

A Priori Models Included and Dimension Definitions

Table 6. Arthur et al.'s (2003) Taxonomy

Label	Definition
1. Communication	conveying oral and written information and responding to questions and challenges
2. Consideration / Awareness of Others	considering the feelings and needs of others as well as being aware of the impact of decisions relevant to other constituents both inside and outside the organization
3. Drive	originating and maintaining a high activity level, setting high performance standards and persisting in their achievement, and expressing the desire to advance to higher job levels
4. Influencing Others	persuading others to do something or adopt a point of view in order to produce desired results and takes action in which the dominant influence is one's own convictions rather than the influence of others' opinions
5. Organizing and Planning	the extent to which an individual systematically arranges his/her own work and resources as well as that of others for efficient task accomplishment; and the extent to which an individual anticipates and prepares for the future
6. Problem Solving	gathering information, understanding relevant technical and professional information, generating viable options, ideas and solutions, selecting supportable courses of action for problems and situations, using available resources in new ways, and generating and recognizing imaginative solutions
7. Stress Tolerance	the extent to which an individual maintains effectiveness in diverse situations under varying degrees of pressure, opposition, and disappointment

Table 7. Borman and Brush's (1993) Taxonomy

Label	Definition
1. Interpersonal Dealings and Communication	Communicating effectively and keeping others informed; Representing the organization to customers and the public; Maintaining good working relationships; Selling / Influencing
2. Leadership and Supervision	Guiding, directing, and motivating subordinates and providing feedback; Training, coaching, and developing subordinates; Coordinating subordinates and others resources to get the job done
3. Technical Activities and the "Mechanics of Management"	Planning and organizing; Technical proficiency; Administration and paperwork; Decision making / problem solving; Staffing; Monitoring and controlling resources; Delegating; Collecting and interpreting data
4. Useful Personal Behavior and Skills	Persisting to reach goals; Handling crises and stress; Organizational commitment

Table 8. Schmitt's (1977) Taxonomy

Label	Definition
1. Administrative Skills	inner work standards, organizing and planning, decision making, decisiveness, and written communication skills
2. Interpersonal Skills	tolerance of uncertainty, self-objectivity, behavior flexibility, and leadership skills
3. Activity / Forcefulness	energy, resistance to stress, need advancement, forcefulness, reliance on others, and oral communication

Table 9. Shore et al.’s (1990) Taxonomy

Label	Definition
1. Interpersonal-Style	an assessee’s amount of participation, the impact they had on outcomes, personal acceptability, and understanding of people
2. Performance-Style	originality, oral communication, recognizing priorities, need for structure, thoroughness, work quality and work drive

Appendix B

Thornton and Byham's (1982) List of Common AC Dimensions

Table 10. Thornton and Byham’s (1982) List of Common AC Dimensions

1. Oral communication	Effective expression in individual or group situations (includes gestures and nonverbal communications)
2. Oral presentation	Effective expression when presenting ideas or tasks to an individual or to a group when given time for preparation (includes gestures and nonverbal communication)
3. Written communication	Clear expression of ideas in writing and use of good grammatical form
4. Planning and Organizing	Establishing a course of action for self and/or others to accomplish a specific goal; planning proper assignments of personnel and appropriate allocation of resources
5. Delegation	Utilizing subordinates effectively; allocating decision making and other responsibilities to the appropriate subordinates
6. Control	Establishing procedures to monitor and/or regulate processes, tasks, or activities of subordinates and job activities and responsibilities; taking action to monitor the results of delegated assignments or projects
7. Development of Subordinates	Developing the skills and competencies of subordinates through training and development activities related to current and future jobs
8. Organizational Sensitivity	Action that indicates an awareness of the impact and the implications of decisions on other components of the organization
9. Extraorganizational Sensitivity	Action that indicates an awareness of the impact and implications of decisions relevant to societal and governmental factors
10. Extraorganizational Awareness	Use of knowledge of changing societal and governmental pressures outside the organization to identify potential problems and opportunities
11. Organizational Awareness	Use of knowledge of changing situations and pressures inside the organization to identify potential organizational problems and opportunities
12. Sensitivity	Actions that indicate a consideration for the feelings and needs of others
13. Leadership	Utilization of appropriate interpersonal styles and methods in guiding individuals (subordinates, peers, superiors) or groups toward task accomplishment
14. Recognition of Employee Safety Needs	Awareness of conditions that affect employees’ safety needs and taking action to resolve inadequacies and discrepancies

15. Analysis	Identifying problems, securing relevant information, relating data from different sources, and identifying possible causes of problems
16. Judgment	Developing alternative courses of action and making decisions based on logical assumptions that reflect factual information
17. Creativity	Generating and/or recognizing imaginative solutions and innovations in work-related situations
18. Risk-Taking	Taking or initiating action that involves a deliberate gamble in order to achieve a recognized benefit or advantage
19. Decisiveness	Readiness to make decisions, render judgments, take action or commit oneself
20. Technical and Professional Knowledge	Level of understanding of relevant technical and professional information
21. Energy	Maintaining a high activity level
22. Range of Interests	Breadth and diversity of general business related knowledge – well informed
23. Initiative	Active attempts to influence events to achieve goals; self-starting rather than passive acceptance. Taking action to achieve goals beyond those called for; originating action.
24. Tolerance for Stress	Stability of performance under pressure and/or opposition
25. Adaptability	Maintaining effectiveness in varying environments, with various tasks, responsibilities or people
26. Independence	Taking action in which the dominant influence is one’s own convictions rather than the influence of others’ opinions
27. Tenacity	Staying with a position or plan of action until the desired objective is achieved or is no longer reasonably attainable
28. Job Motivation	The extent to which activities and responsibilities available in the job overlap with activities and responsibilities that result in personal satisfaction
29. Career Ambition	The expressed desire to advance to higher job levels with active efforts toward self-development and advancement
30. Integrity	Maintaining social, ethical, and organizational norms in job-related activities
31. Work Standards	Setting high goals or standards of performance for self, subordinates, others and organization. Dissatisfied with average performance
32. Resilience	Handling disappointment and/or rejection while maintaining effectiveness
33. Practical Learning	Assimilating and applying new, job-related information, taking into consideration rate and complexity

Appendix C

Primary Studies Included in the Meta-Analysis

Abraham, J. D., Morrison, J. D., & Burnett, D. D. (2006). Feedback seeking among developmental assessment center participants. *Journal of Business and Psychology, 20,* 383-394.

Ackerman, L.D. (1996). Investigation of the external construct validity of the assessment center method. *Dissertation Abstracts International: Section B: The Sciences and Engineering, 57,* 2194.

Archambeau, D. J. (1979). Relationships among skill ratings assigned in an assessment center. *Journal of Assessment Center Technology, 2,* 7-19.

Archambeau, D.J. (1982). The predictability of job performance via the assessment center method: Implications for career planning and development. Unpublished Doctoral Dissertation. University of South Florida.

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

Bart, L.A. (1987). The role of the assessee in developmental assessment centers: A field experiment. Unpublished doctoral dissertation. Ohio State University.

Beardsley, S.A. (1986). Unanticipated outcomes experienced by assessors in a management skills assessment center: An exploratory analysis. Unpublished doctoral dissertation. University of California, Berkeley.

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

Clapham, M.M. (1998). A comparison of assessor and self dimension ratings in an advanced management assessment center. *Journal of Occupational and Organizational Psychology*, *71,* 193-203.

Coles, J.S. (1990). The relationship between personal characteristics, locus-of-control and administrative assessment center performance. University of Wisconsin, Madison.

Denton, D.W. (1997). A structural equations analysis of the relationship between personality and performance in a supervisory assessment center. Unpublished doctoral dissertation. The University of Tennessee.

Fishback, P.D. (1992). The relationship between selected demographic factors and personal characteristics and skill dimension scores of participants in an assessment center. Unpublished doctoral dissertation. Virginia Commonwealth University.

Foster, M.R. (1995). The use of biographical information to determine skill levels as measured in an assessment center. Unpublished doctoral dissertation. University of Georgia.

Goffin, R.D.; Rothstein, M.G.; Johnston, N.G. (1996). Personality testing and the assessment center: Incremental validity for managerial selection. *Journal of Applied Psychology, 81*, 746-756.

Goldstein, H.W.; Yusko, K.P.; Nicolopoulos, V. (2001). Exploring Black-White subgroup differences of managerial competencies. *Personnel Psychology, 54,* 783-807.

Gomez, J. J. (1985). The criterion-related validity of an assessment center for the selection of school-level administrators. Unpublished doctoral dissertation, University of Miami, Coral Gables.

Heon, S.M. (1990). Self-monitoring and cognitive complexity as related to assessment center performance. Unpublished Doctoral Dissertation.

Hoffman, B.J. (2006). *Disentangling the meaning of multisource feedback: An examination of the nomological network surrounding source and dimension factors.* Unpublished doctoral dissertation. The University of Tennessee.

Hoffman, B.J. & Kudisch, J. (2002). Expanding the nomological network of assessment center dimensions using external measures of cognitive ability and personality. Paper presented at the 29th annual International Congress on the Assessment Center Method, Minneapolis, MN.

Howard, A. (1983). Measuring management abilities and motivation. *New Directions for Testing and Measurement, 17,* 31-44.

Huck, J.R. (1974). Determinants of assessment center ratings for White and Black females and the relationship of these dimensions to subsequent performance effectiveness. Unpublished doctoral dissertation. Wayne State University.

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

Krause, D.E., Kersting, M., Heggestad, E.D., & Thornton, G.C. (2006). Incremental validity of Assessment Center rating over cognitive ability tests: A study at the executive management level. *International Journal of Selection and Assessment, 14*, 360-371.

Kudisch, J.D. (1997). Factors related to participant's acceptance of developmental assessment center feedback. Unpublished doctoral dissertation. The University of Tennessee.

Lynch, F.P. (1995). Construct validity of assessment center dimension ratings: Comparing assessor ratings and coworker ratings. Unpublished doctoral dissertation. California School of Professional Psychology, San Diego.

Maciejczyk, M. (1992). The relationship of self-ratings to assessor ratings of assessment center performance. Unpublished doctoral dissertation. Iowa State University.

Massie, E.W. (1992). An analysis of the predictive validity of the NASSP Assessment Center in the Maryland Assessment Center Program. Unpublished doctoral dissertation. University of Maryland.

Mitchel, J.O. (1974). A longitudinal study of the predictive validity of an assessment center. Unpublished doctoral dissertation. Bowling Green State University.

Moses, J.L. (1972). Assessment center performance and management progress. *Studies in Personnel Psychology, 4,* 7-12.

Moses, J.L. (1973). The development of an assessment center for the early identification of supervisory potential. *Personnel Psychology, 26,* 569-580.

Ogletree, C.D. (1996). An investigation of tacit ability in assessment center performance. Unpublished doctoral dissertation. University of Georgia.

Rode, J.C., Arthaud-Day, M.L., Mooney, C.H., Near, J.P., Baldwin, T.T., Bommer, W.H. & Rubin, R.S. (2005). Life Satisfaction and Student Performance. *Academy of Management Learning & Education, 4, 4,* 421–433.

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

Russell, C.J. (1983). An analysis of a decision process in an assessment center. Unpublished doctoral dissertation. University of Iowa.

Schmitt, N. & Cohen, S. A. (1990). Criterion-related validity of the national association of secondary school principals' assessment center. Unpublished technical report.

Shore, L.M., Tetrick, L.E. & Shore, T.H. (1998). A comparison of self-, peer, and assessor evaluations of managerial potential. *Journal of Social Behavior and Personality, 13,* 85-101.

Steilberg, R.C. (1996). Construct validation of a managerial assessment center: A nomological network spanning time and method. Unpublished doctoral dissertation. Georgia Institute of Technology.

Thomson, H.A. (1970). Internal and external validation of an industrial assessment center program. Unpublished doctoral dissertation. Case Western Reserve University.

Thornton, G.C., Tziner, A., Dahan, M., Clevenger, J.P. & Meir, E. (1997). Construct validity of assessment center judgments: Analyses of the behavioral reporting method. *Journal of Social Behavior and Personality, 12,* 109-128.

Tyler, C.L. (2003). The predictive validity and adverse impact of an assessment center, a multirater appraisal system and top-down appraisal for selecting retail store managers. Unpublished doctoral dissertation. Florida Atlantic University.

Williams, R.W. (1998). Using personality traits to predict the criterion space. Unpublished doctoral dissertation. The Union Institute and University.

Yurkon, A. C. (1998). An examination of the criterion-related validity of a developmental assessment center. Unpublished master's thesis, University of North Texas.

Appendix D

Final Dimension Intercorrelations, Total Sample Size (*N*), and Number of Studies (*k*) on Which Correlations are Based

Table 11. Dimension Intercorrelations, Sample Size (*N*), and Number of Studies (*k*)

	1.	2.	3.	4.	5.	6.	7.	8.
1. Oral Communication	.58	13749 28	1042 7	10369 39	1025 6	2898 18	2357 13	4673 8
2. Written Communication	.41	-	3464 9	7641 25	700 4	3697 11	823 3	4673 8
3. Organizational Awareness	.43	.25	-	1042 7	1126 5	4354 13	172 1	172 1
4. Sensitivity	.45	.37	.46	.35	1025 6	2904 18	2357 13	4608 8
5. Career Ambition	.45	.30	.31	.26	.47	1629 8	172 1	266 2
6. Energy	.56	.28	.43	.43	.48	-	1440 8	263 2
7. Initiative	.45	.18	.61	.33	.64	.52	-	172 1
8. Job Motivation	.49	.37	.68	.44	.64	.53	.64	-
9. Leadership	.54	.32	.50	.46	.33	.64	.57	.41
10. Delegation	.31	.43	.61	.21	.62	.56	.59	.61
11. Planning and Organizing	.46	.42	.40	.38	.35	.50	.44	.43
12. Analysis	.41	.36	.40	.41	.34	.43	.34	.47
13. Creativity	.46	.28	.64	.41	.50	.50	.38	.57
14. Decisiveness	.49	.32	.21	.36	.33	.34	.51	.44
15. Judgment	.45	.38	.44	.43	.36	.44	.44	.44
16. Adaptability	.45	.23	.63	.52	.34	.51	.40	.34
17. Stress Tolerance	.55	.36	.39	.44	.32	.50	.38	.54

Table 11 (continued)

	9.	10.	11.	12.	13.	14.	15.	16.	17.
1.	16261 40	702 3	15941 35	12562 22	2578 12	5616 18	15282 35	3267 20	6135 18
2.	15958 28	601 2	16495 30	14146 19	1095 4	7154 15	16107 29	5194 19	5910 16
3.	4354 13	172 1	4187 12	3664 8	172 1	3664 8	4354 13	4354 13	2163 12
4.	10693 40	702 3	10374 36	6129 19	2669 13	5126 16	9464 34	3311 20	6226 19
5.	1924 10	172 1	2073 11	1421 7	675 4	1421 7	2073 11	1723 9	1872 10
6.	6301 25	172 1	5266 17	3664 8	1894 11	3804 9	5173 19	4997 17	2757 16
7.	2357 13	273 2	1500 6	1089 6	1933 8	484 4	1229 7	963 4	906 3
8.	4460 9	172 1	4767 9	4385 8	263 2	3987 7	4460 9	804 4	3922 7
9.	.64	702 3	19167 40	15495 27	2669 13	8947 24	19239 44	6873 28	7292 24
10.	.46	-	702 3	273 2	172 1	273 2	702 3	172 1	273 2
11.	.51	.45	.60	15636 26	1886 7	8784 23	19309 41	1606 27	7183 23
12.	.47	.60	.59	-	805 2	8882 24	15602 28	4816 16	5866 14
13.	.49	.57	.45	.28	-	172 1	1541 7	1450 6	1541 7
14.	.40	.63	.56	.48	.67	-	8947 24	4305 15	5233 13
15.	.52	.48	.65	.61	.53	.57	.68	6873 28	7441 25
16.	.56	.45	.36	.38	.39	.28	.41	.44	3535 18
17.	.49	.42	.46	.47	.32	.50	.49	.50	.63

Appendix E

Missing Meta-Analytic Correlations

Table 12. Missing Meta-Analytic Dimension Intercorrelations

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1. Oral communication	.	.	.	X	X	.	.	.	.	.	.	.	.	.
2. Oral presentation		.	.	X	X	.	.	X	.	.	.	.	.	X
3. Written communication			X	X	X	.	.	.	.	.	.	.	.	.
4. Extra-org. awareness				X	X	X	X	X	X	X	X	X	X	X
5. Extra-org. sensitivity					X	X	X	X	X	X	X	X	X	X
6. Organizational awareness						X	X	X	.	.	.	.	.	.
7. Organizational sensitivity							X	X	.	.	.	.	X	X
8. Rec. of safety needs								X	X	X	X	X	X	X
9. Sensitivity									.	.	.	.	.	.
10. Career ambition										.	.	.	.	.
11. Energy											X	.	.	.
12. Initiative												X	.	.
13. Job motivation													X	.
14. Tenacity														X
15. Work standards														
16. Independence														
17. Integrity														
18. Leadership														
19. Control														
20. Delegation														
21. Develop. of subordinates														
22. Planning and organization														
23. Analysis														
24. Creativity														
25. Decisiveness														
26. Judgment														
27. Practical learning														
28. Range of interests														
29. Tech. and prof. know.														
30. Adaptability														
31. Resilience														
32. Risk taking														
33. Tolerance for stress														

Table 12 (continued)

	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33
1.	.	.	X	.	.	.	.	.	.	.	.	.	X	.	.	.	X	X	.
2.	.	.	X	.	.	X	.	.	.	X	.	.	X	X	X	.	X	.	.
3.	.	.	X	.	.	.	.	.	.	.	.	.	X	.	.	.	X	X	.
4.	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
5.	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
6.	.	.	X	.	.	.	X	.	.	.	.	.	X	.	X	.	.	.	.
7.	.	X	X	.	.	X	.	.	.	.	X	.	X	.	X	.	X	X	.
8.	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
9.	.	.	X	.	.	.	.	.	.	.	.	.	.	.	.	.	X	X	.
10.	.	.	X	.	X	.	X	.	.	.	.	.	X	.	X	.	X	.	.
11.	.	.	X	.	.	.	X	.	.	.	.	.	X	.	X	.	X	.	.
12.	.	.	X	.	.	.	.	.	.	.	.	.	X	.	X	.	X	X	.
13.	.	.	X	.	X	.	X	.	.	.	.	.	X	.	.	.	X	.	.
14.	.	X	X	.	X	.	X	.	.	.	.	.	X	X	X	.	X	.	.
15.	.	.	X	.	.	X	.	.	.	.	.	.	.	.	X	.	X	.	.
16.		X	X	.	X	.	.	.	.	X	.	.	X	X	X	.	X	.	.
17.			X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
18.				.	.	.	.	.	.	.	.	.	X	.	X	.	X	.	.
19.					.	X	.	.	.	.	.	.	X	X	.	.	X	X	.
20.						X	X	.	.	.	.	.	X	X	X	.	X	X	.
21.							X	.	.	X	.	.	.	X	X	.	X	X	X
22.								.	.	.	.	.	.	.	.	.	X	.	.
23.									X	.	.	.	X	.	.	.	X	.	.
24.										X	.	.	X	.	X	.	X	X	.
25.											X	.	X	.	.	.	X	.	.
26.												.	X	.	X	.	X	.	.
27.													X	X	X	X	X	X	X
28.														X	X	.	X	X	.
29.															X	X	X	X	X
30.																.	X	.	.
31.																	X	X	X
32.																		X	X
33.																			.

Note. Intercorrelations that were not available are marked with an X. Rows and columns shaded in dark gray represent excluded variables.

Appendix F

A Priori Models Examined in CFA Analyses

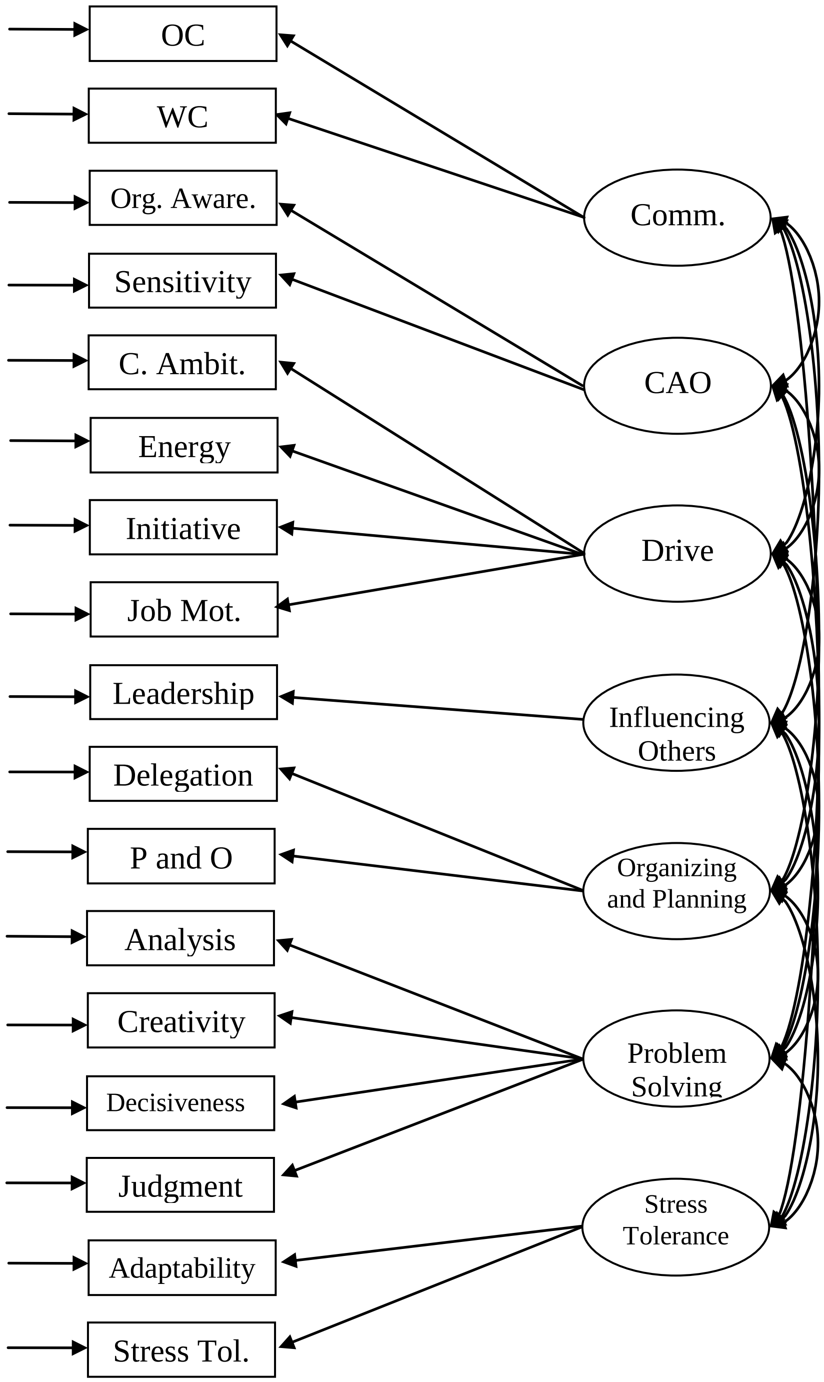


Figure 2. *CFA Model for Arthur et al.'s (2003) Seven-Factor Framework*

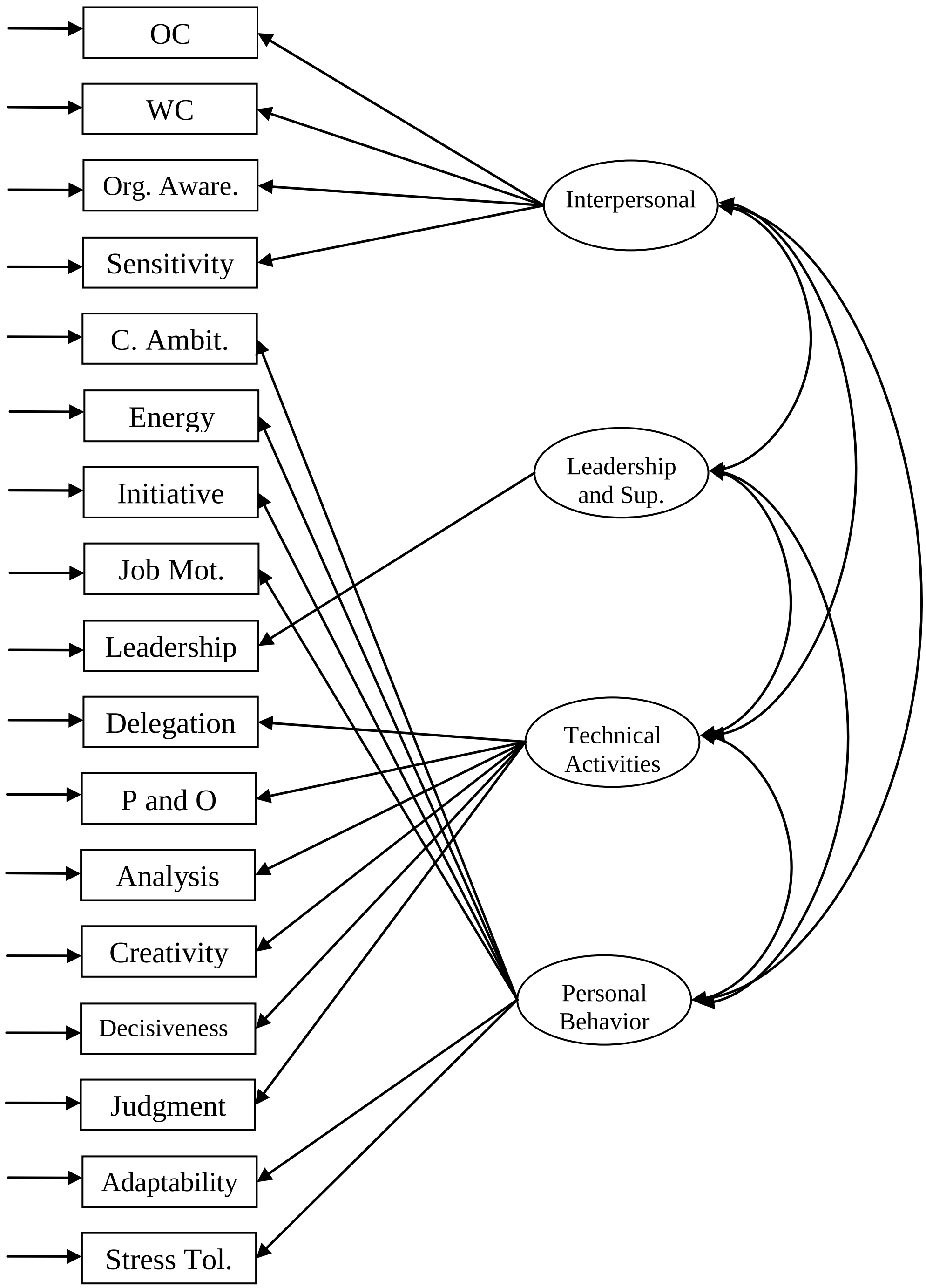


Figure 3. *CFA Model for Borman and Brush's (1993) Four-Factor Framework*

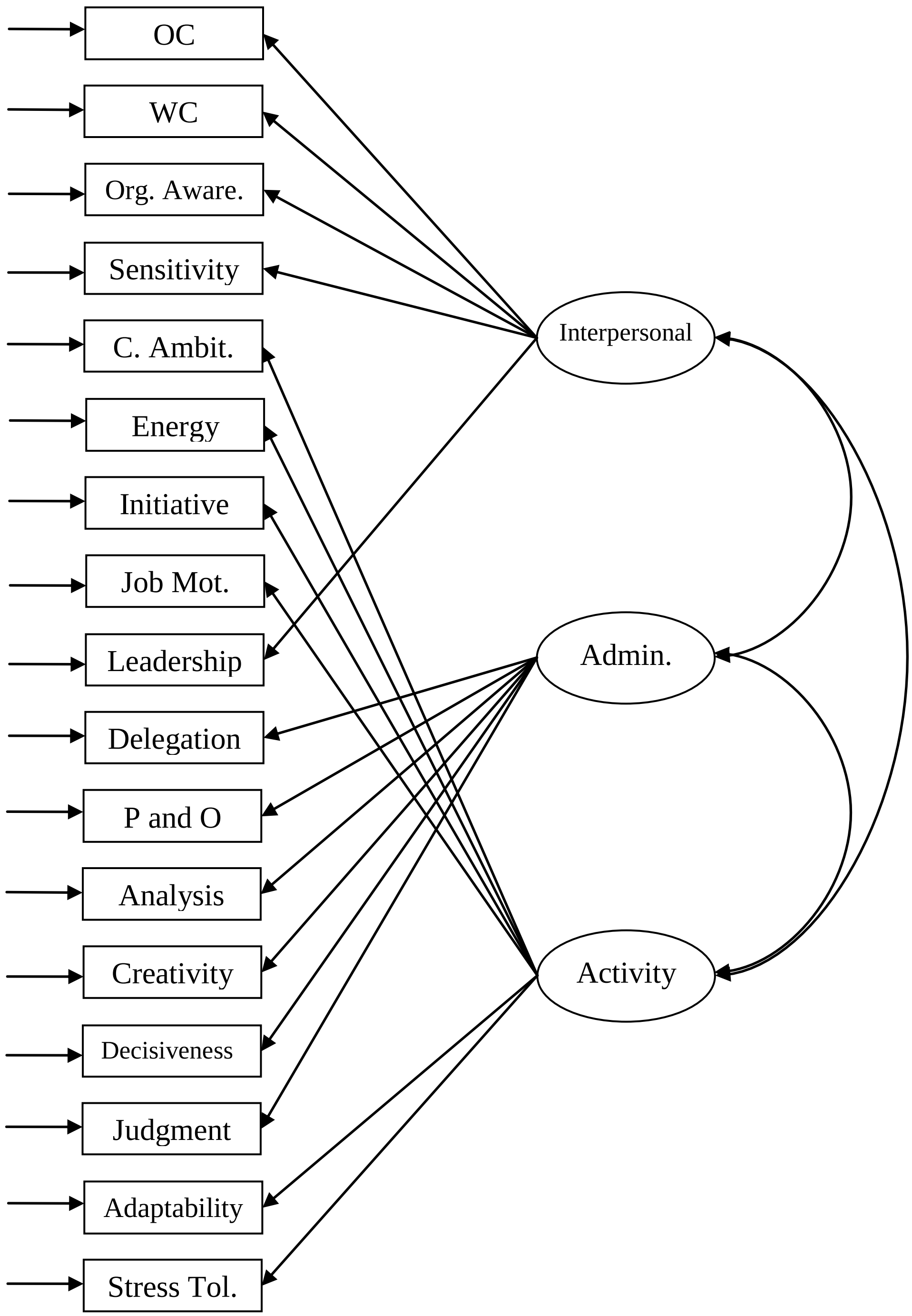


Figure 4. *CFA Model for Schmitt's (1977) Three-Factor Framework*

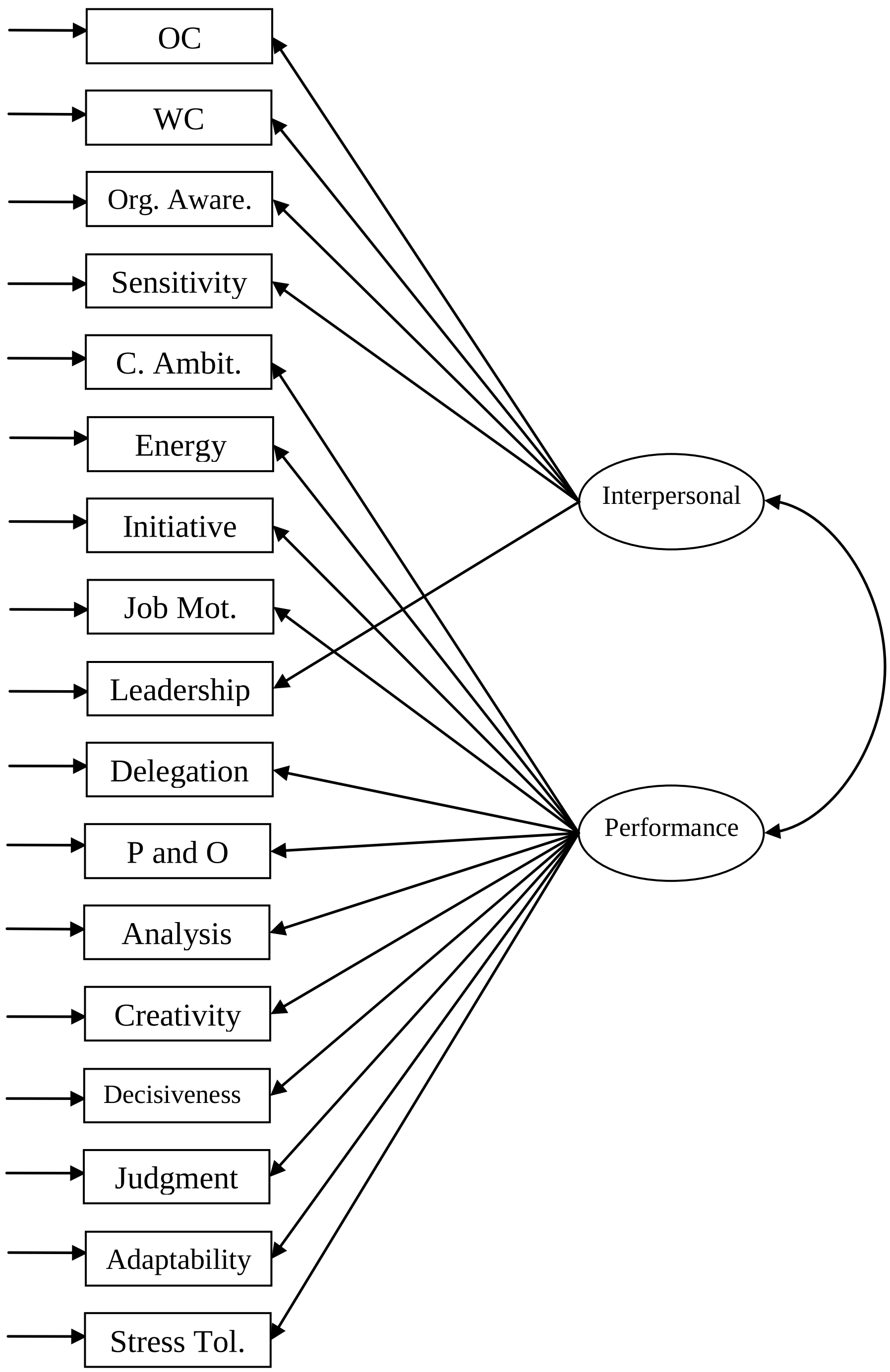


Figure 5. *CFA Model for Shore et al.'s (1992) Two-Factor Framework*

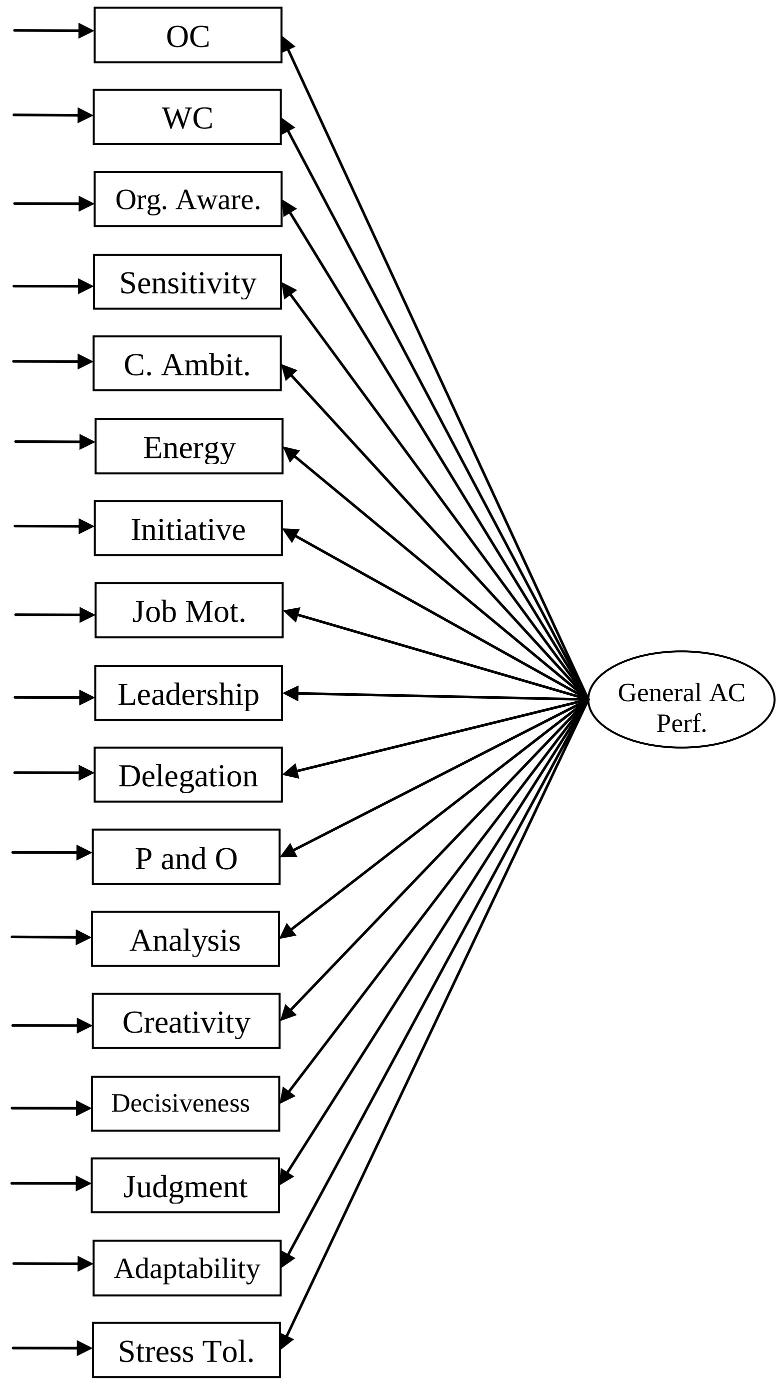


Figure 6. *CFA Model for Viswesvaran et al.'s (2005) Framework*

Appendix G

Instructions and Forms for Classifying Common AC Dimensions into A Priori Models

Assessment Center Dimension Rating Task

Subject Matter Expert: Thank you for taking the time to help with this AC dimension coding task. The purpose of this task is to decide how you think a list of commonly used AC dimensions fits into broader categories. Please read the names, definitions, and examples of each of the dimensions carefully. Once you understand the names of these dimensions, please take the list of [33] dimensions and sort them into the appropriate categories. To do this, simply write the numbers of the labels of the dimensions in the box corresponding to that particular category.

Please consider your responses clearly, and pay careful attention to the conceptual similarity of these dimensions. There does not need to be a perfect correspondence between dimension labels, the important consideration is where you think the dimensions best match. If you think the level one dimension is not related to any of the level two dimensions, place it in the ‘Unclassifiable’ box.

Please classify the following dimensions into the categories listed on the previous page into the categories below. The purpose of this exercise is to determine how these commonly used assessment center dimension labels group together.

Category	Primary Dimension Numbers
1. Communication	
2. Consideration / Awareness of Others	
3. Drive	
4. Influencing Others	
5. Organizing and Planning	
6. Problem Solving	
7. Stress Tolerance	
Unclassifiable	

Please classify the following dimensions into the categories listed on the previous page into the categories below. The purpose of this exercise is to determine how these commonly used assessment center dimension labels group together.

Category	Primary Dimension Numbers
1. Interpersonal Dealings and Communication	
2. Leadership and Supervision	
3. Technical Activities and the "Mechanics of Management"	
4. Useful Personal Behavior and Skills	
Unclassifiable	

Please classify the following dimensions into the categories listed on the previous page into the categories below. The purpose of this exercise is to determine how these commonly used assessment center dimension labels group together.

Category	Primary Dimension Numbers
1. Administrative Skills	
2. Interpersonal Skills	
3. Activity / Forcefulness	
Unclassifiable	

Please classify the following dimensions into the categories listed on the previous page into the categories below. The purpose of this exercise is to determine how these commonly used assessment center dimension labels group together.

Category	Primary Dimension Numbers
1. Interpersonal-Style	
2. Performance-Style	
Unclassifiable	

Please read the following dimension names and refer to this sheet as necessary when making your ratings.

Dimension Name	Definition
1. Oral communication	Effective expression in individual or group situations (includes gestures and nonverbal communications)
2. Oral presentation	Effective expression when presenting ideas or tasks to an individual or to a group when given time for preparation (includes gestures and nonverbal communication)
3. Written communication	Clear expression of ideas in writing and use of good grammatical form
4. Planning and Organizing	Establishing a course of action for self and/or others to accomplish a specific goal; planning proper assignments of personnel and appropriate allocation of resources
5. Delegation	Utilizing subordinates effectively; allocating decision making and other responsibilities to the appropriate subordinates
6. Control	Establishing procedures to monitor and/or regulate processes, tasks, or activities of subordinates and job activities and responsibilities; taking action to monitor the results of delegated assignments or projects
7. Development of Subordinates	Developing the skills and competencies of subordinates through training and development activities related to current and future jobs
8. Organizational Sensitivity	Action that indicates an awareness of the impact and the implications of decisions on other components of the organization
9. Extra-Organizational Sensitivity	Action that indicates an awareness of the impact and implications of decisions relevant to societal and governmental factors
10. Extra-Organizational Awareness	Use of knowledge of changing societal and governmental pressures outside the organization to identify potential problems and opportunities
11. Organizational Awareness	Use of knowledge of changing situations and pressures inside the organization to identify potential organizational problems and opportunities
12. Sensitivity	Actions that indicate a consideration for the feelings and needs of others
13. Leadership	Utilization of appropriate interpersonal styles and methods in guiding individuals (subordinates, peers, superiors) or groups toward task accomplishment
14. Recognition of Employee Safety Needs	Awareness of conditions that affect employees’ safety needs and taking action to resolve inadequacies and discrepancies

15. Analysis	Identifying problems, securing relevant information, relating data from different sources, and identifying possible causes of problems
16. Judgment	Developing alternative courses of action and making decisions based on logical assumptions that reflect factual information
17. Creativity	Generating and/or recognizing imaginative solutions and innovations in work-related situations
18. Risk-Taking	Taking or initiating action that involves a deliberate gamble in order to achieve a recognized benefit or advantage
19. Decisiveness	Readiness to make decisions, render judgments, take action or commit oneself
20. Technical and Professional Knowledge	Level of understanding of relevant technical and professional information
21. Energy	Maintaining a high activity level
22. Range of Interests	Breadth and diversity of general business related knowledge – well informed
23. Initiative	Active attempts to influence events to achieve goals; self-starting rather than passive acceptance. Taking action to achieve goals beyond those called for; originating action.
24. Tolerance for Stress	Stability of performance under pressure and/or opposition
25. Adaptability	Maintaining effectiveness in varying environments, with various tasks, responsibilities or people
26. Independence	Taking action in which the dominant influence is one's own convictions rather than the influence of others' opinions
27. Tenacity	Staying with a position or plan of action until the desired objective is achieved or is no longer reasonably attainable
28. Job Motivation	The extent to which activities and responsibilities available in the job overlap with activities and responsibilities that result in personal satisfaction
29. Career Ambition	The expressed desire to advance to higher job levels with active efforts toward self-development and advancement
30. Integrity	Maintaining social, ethical, and organizational norms in job-related activities
31. Work Standards	Setting high goals or standards of performance for self, subordinates, others and organization. Dissatisfied with average performance
32. Resilience	Handling disappointment and/or rejection while maintaining effectiveness
33. Practical Learning	Assimilating and applying new, job-related information, taking into consideration rate and complexity

Appendix H

Rater Agreement for Classifying Common AC Dimensions into A Priori Models

Table 13. Arthur et al. (2003) Rating Agreement

	CM	CAO	DR	IO	OP	PS	ST	X
1. Oral communication	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2. Oral presentation	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
3. Written communication	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
4. Planning and Organizing	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00
5. Delegation	0.00	0.00	0.00	0.63	0.25	0.13	0.00	0.00
6. Control	0.00	0.13	0.00	0.38	0.38	0.13	0.00	0.00
7. Development of Subordinates	0.00	0.13	0.13	0.63	0.13	0.00	0.00	0.00
8. Organizational Sensitivity	0.00	0.75	0.00	0.00	0.13	0.13	0.00	0.00
9. Extra-Organizational Sensitivity	0.00	0.75	0.00	0.00	0.13	0.13	0.00	0.00
10. Extra-Organizational Awareness	0.00	0.25	0.00	0.00	0.00	0.75	0.00	0.00
11. Organizational Awareness	0.00	0.25	0.00	0.00	0.00	0.75	0.00	0.00
12. Sensitivity	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00
13. Leadership	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00
14. Rec. of Employee Safety Needs	0.00	0.88	0.00	0.00	0.00	0.13	0.00	0.00
15. Analysis	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00
16. Judgment	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00
17. Creativity	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00
18. Risk-Taking	0.00	0.00	0.50	0.00	0.00	0.13	0.25	0.13
19. Decisiveness	0.00	0.00	0.38	0.13	0.13	0.38	0.00	0.00
20. Tech. and Prof. Knowledge	0.00	0.00	0.00	0.00	0.00	0.38	0.00	0.63
21. Energy	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00
22. Range of Interests	0.00	0.00	0.13	0.00	0.00	0.25	0.00	0.63
23. Initiative	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00
24. Tolerance for Stress	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00
25. Adaptability	0.00	0.00	0.00	0.00	0.13	0.25	0.50	0.13
26. Independence	0.00	0.00	0.75	0.13	0.00	0.13	0.00	0.00
27. Tenacity	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00
28. Job Motivation	0.00	0.00	0.63	0.00	0.00	0.00	0.00	0.38
29. Career Ambition	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00
30. Integrity	0.00	0.63	0.00	0.13	0.00	0.00	0.00	0.25
31. Work Standards	0.00	0.00	0.75	0.13	0.00	0.00	0.00	0.13
32. Resilience	0.00	0.00	0.13	0.00	0.00	0.00	0.88	0.00
33. Practical Learning	0.00	0.00	0.00	0.00	0.00	0.88	0.00	0.13

Note. Underlined values represent primary dimension labels where the proportion of raters was .75 or greater. CM = Communication, CAO = Consideration and Awareness of Others, DR = Drive, IO = Influencing Others, OP = Organizing and Planning, PS = Problem Solving, ST = Stress Tolerance, X = Unclassifiable.

Table 14. Borman and Brush (1993) Rating Agreement

	IntpDeal	LeadSup	TechAct	PersBeh	X
1. Oral communication	1.00	0.00	0.00	0.00	0.00
2. Oral presentation	1.00	0.00	0.00	0.00	0.00
3. Written communication	0.88	0.00	0.13	0.00	0.00
4. Planning and Organizing	0.00	0.13	0.75	0.13	0.00
5. Delegation	0.00	0.88	0.13	0.00	0.00
6. Control	0.00	0.63	0.38	0.00	0.00
7. Development of Subordinates	0.00	1.00	0.00	0.00	0.00
8. Organizational Sensitivity	0.25	0.13	0.50	0.00	0.13
9. Extra-Organizational Sensitivity	0.13	0.13	0.63	0.00	0.13
10. Extra-Organizational Awareness	0.00	0.00	0.75	0.00	0.25
11. Organizational Awareness	0.00	0.13	0.63	0.13	0.13
12. Sensitivity	0.63	0.00	0.00	0.38	0.00
13. Leadership	0.13	0.88	0.00	0.00	0.00
14. Rec. of Employee Safety Needs	0.25	0.38	0.38	0.00	0.00
15. Analysis	0.00	0.00	0.75	0.25	0.00
16. Judgment	0.00	0.00	0.75	0.25	0.00
17. Creativity	0.00	0.13	0.13	0.75	0.00
18. Risk-Taking	0.00	0.00	0.13	0.63	0.25
19. Decisiveness	0.00	0.25	0.38	0.38	0.00
20. Tech. and Prof. Knowledge	0.00	0.00	0.75	0.25	0.00
21. Energy	0.00	0.00	0.00	1.00	0.00
22. Range of Interests	0.00	0.00	0.38	0.63	0.00
23. Initiative	0.00	0.00	0.00	1.00	0.00
24. Tolerance for Stress	0.00	0.00	0.00	1.00	0.00
25. Adaptability	0.13	0.00	0.13	0.75	0.00
26. Independence	0.00	0.13	0.00	0.88	0.00
27. Tenacity	0.00	0.00	0.13	0.88	0.00
28. Job Motivation	0.00	0.00	0.00	0.75	0.25
29. Career Ambition	0.00	0.00	0.00	0.88	0.13
30. Integrity	0.50	0.13	0.00	0.38	0.00
31. Work Standards	0.00	0.38	0.13	0.38	0.13
32. Resilience	0.00	0.00	0.00	1.00	0.00
33. Practical Learning	0.00	0.00	0.50	0.50	0.00

Note. Underlined values represent primary dimension labels where the proportion of raters was .75 or greater. IntpDeal = Interpersonal Dealings and Communication, LeadSup = Leadership and Supervision, TechAct = Technical Activities and the Mechanics of Management, PersBeh = Useful Personal Behavior, X = Unclassifiable.

Table 15. Schmitt (1977) Rating Agreement

	Admin.	Interpersonal	Activity	X
1. Oral communication	0.00	1.00	0.00	0.00
2. Oral presentation	0.00	1.00	0.00	0.00
3. Written communication	0.25	0.63	0.00	0.13
4. Planning and Organizing	1.00	0.00	0.00	0.00
5. Delegation	0.75	0.13	0.13	0.00
6. Control	0.88	0.00	0.13	0.00
7. Development of Subordinates	0.38	0.50	0.00	0.13
8. Organizational Sensitivity	0.50	0.25	0.13	0.13
9. Extra-Organizational Sensitivity	0.50	0.13	0.13	0.25
10. Extra-Organizational Awareness	0.63	0.00	0.13	0.25
11. Organizational Awareness	0.63	0.00	0.13	0.25
12. Sensitivity	0.00	1.00	0.00	0.00
13. Leadership	0.13	0.75	0.00	0.13
14. Rec. of Employee Safety Needs	0.63	0.25	0.13	0.00
15. Analysis	0.63	0.00	0.25	0.13
16. Judgment	0.63	0.00	0.25	0.13
17. Creativity	0.25	0.13	0.63	0.00
18. Risk-Taking	0.00	0.13	0.75	0.13
19. Decisiveness	0.25	0.13	0.50	0.13
20. Tech. and Prof. Knowledge	0.75	0.00	0.13	0.13
21. Energy	0.00	0.13	0.75	0.13
22. Range of Interests	0.38	0.00	0.13	0.50
23. Initiative	0.00	0.00	1.00	0.00
24. Tolerance for Stress	0.00	0.25	0.38	0.38
25. Adaptability	0.13	0.13	0.63	0.13
26. Independence	0.13	0.00	0.50	0.38
27. Tenacity	0.00	0.00	1.00	0.00
28. Job Motivation	0.00	0.00	0.38	0.63
29. Career Ambition	0.00	0.00	0.88	0.13
30. Integrity	0.00	0.50	0.00	0.50
31. Work Standards	0.63	0.13	0.25	0.00
32. Resilience	0.13	0.13	0.50	0.25
33. Practical Learning	0.75	0.00	0.13	0.13

Note. Underlined values represent primary dimension labels where the proportion of raters was .75 or greater. Admin. = Administrative Skills, Interpersonal = Interpersonal Skills, Activity = Activity / Forcefulness, X = Unclassifiable.

Table 16. Shore et al. (1992) Rating Agreement

	Interpersonal	Performance	X
1. Oral communication	1.00	0.00	0.00
2. Oral presentation	0.88	0.13	0.00
3. Written communication	0.63	0.25	0.13
4. Planning and Organizing	0.00	1.00	0.00
5. Delegation	0.38	0.63	0.00
6. Control	0.00	1.00	0.00
7. Development of Subordinates	0.50	0.50	0.00
8. Organizational Sensitivity	0.25	0.63	0.13
9. Extra-Organizational Sensitivity	0.13	0.75	0.13
10. Extra-Organizational Awareness	0.00	0.88	0.13
11. Organizational Awareness	0.00	0.88	0.13
12. Sensitivity	1.00	0.00	0.00
13. Leadership	0.88	0.13	0.00
14. Rec. of Employee Safety Needs	0.13	0.63	0.25
15. Analysis	0.00	0.88	0.13
16. Judgment	0.00	0.88	0.13
17. Creativity	0.00	0.88	0.13
18. Risk-Taking	0.13	0.88	0.00
19. Decisiveness	0.13	0.88	0.00
20. Tech. and Prof. Knowledge	0.00	0.63	0.38
21. Energy	0.13	0.50	0.38
22. Range of Interests	0.00	0.63	0.38
23. Initiative	0.00	1.00	0.00
24. Tolerance for Stress	0.00	0.88	0.13
25. Adaptability	0.00	0.88	0.13
26. Independence	0.13	0.50	0.38
27. Tenacity	0.00	0.88	0.13
28. Job Motivation	0.00	0.38	0.63
29. Career Ambition	0.00	0.38	0.63
30. Integrity	0.63	0.38	0.00
31. Work Standards	0.00	1.00	0.00
32. Resilience	0.00	0.88	0.13
33. Practical Learning	0.00	0.88	0.13

Note. Underlined values represent primary dimension labels where the proportion of raters was .75 or greater. Interpersonal = Interpersonal-Style, Performance = Performance-Style, X = Unclassifiable.

Appendix I

Hierarchical CFA Models

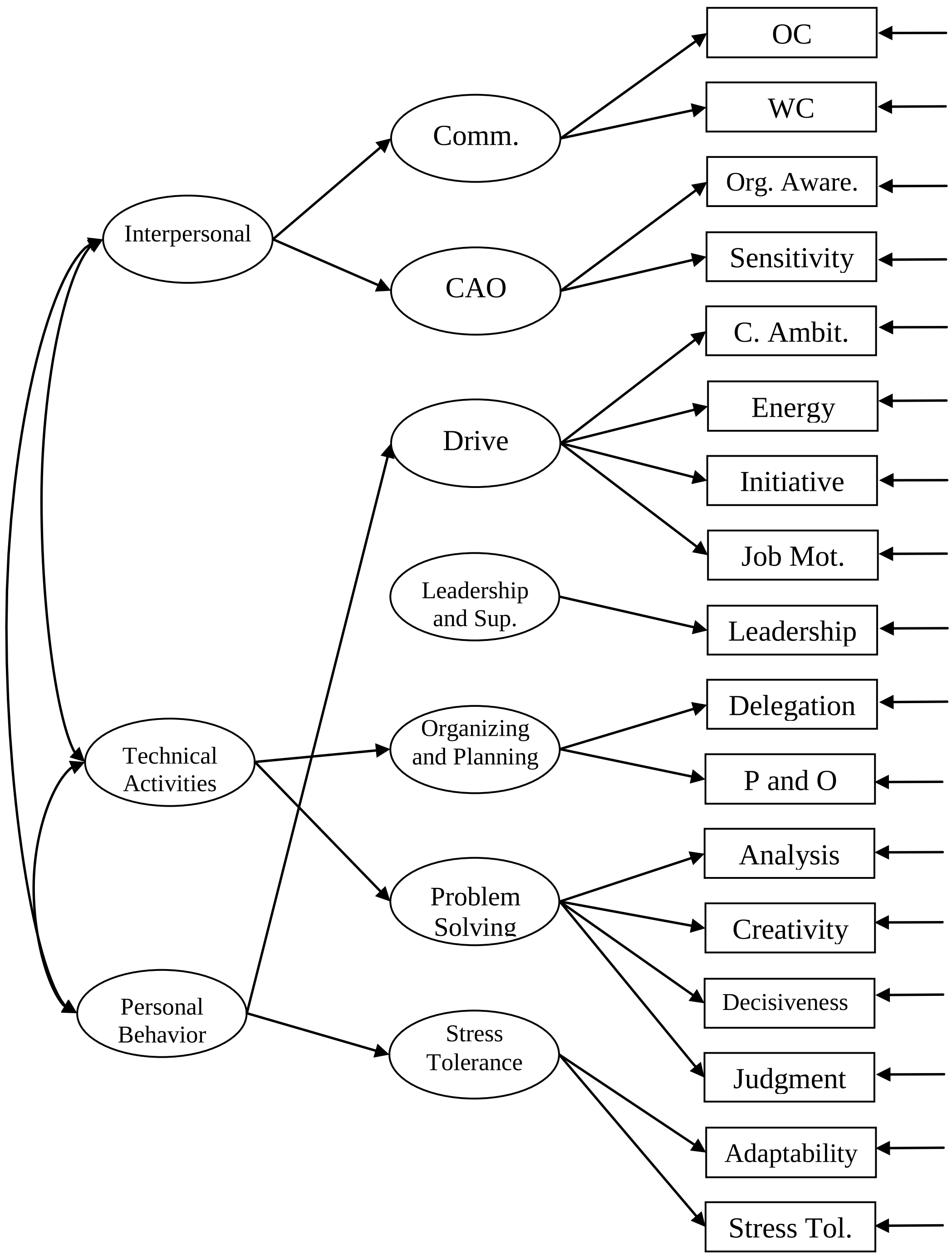


Figure 7. *Hierarchical CFA Model for Borman and Brush's (1993) Four-Factor Framework*

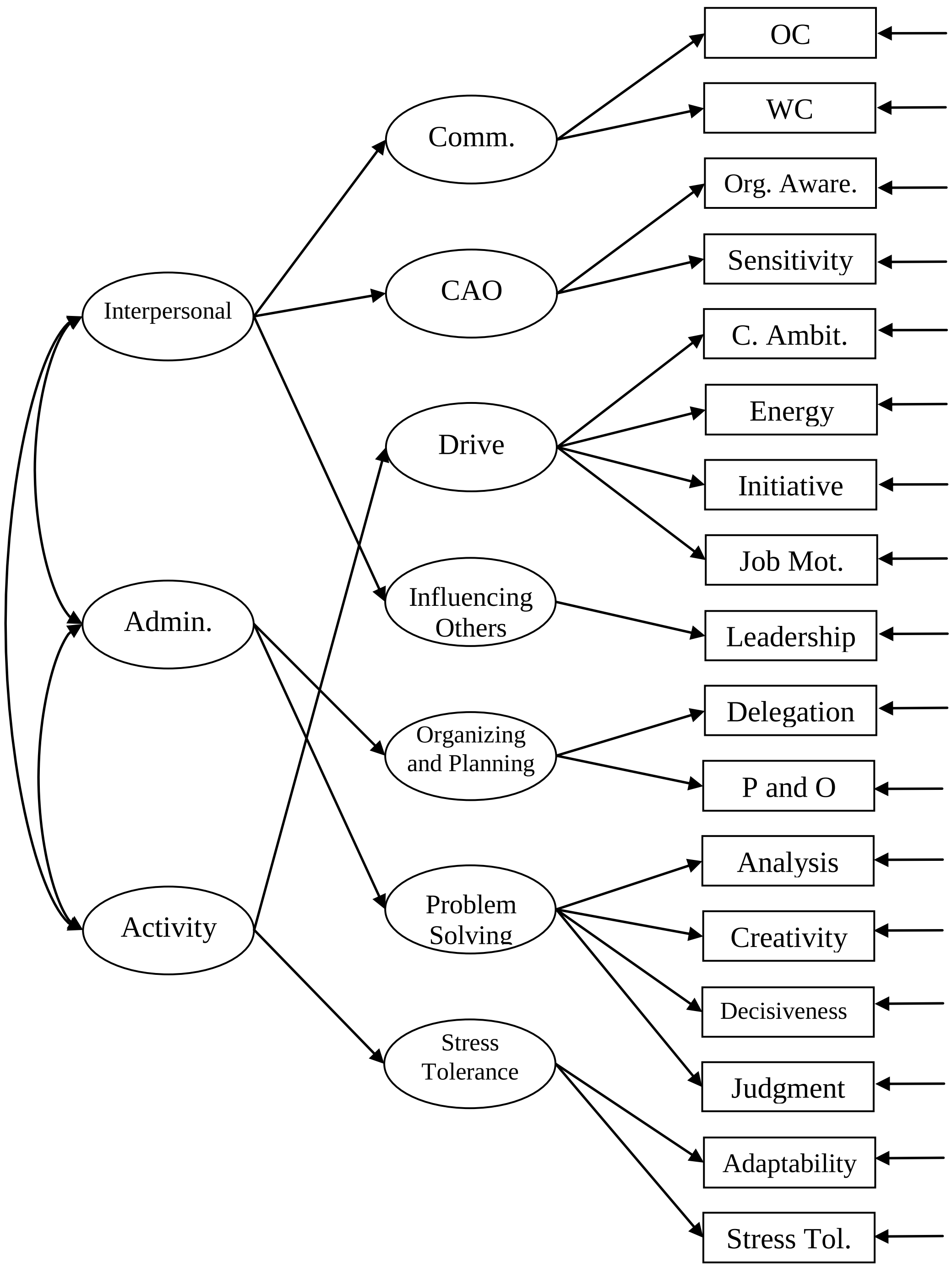


Figure 8. *Hierarchical CFA Model for Schmitt's (1977) Three-Factor Framework*

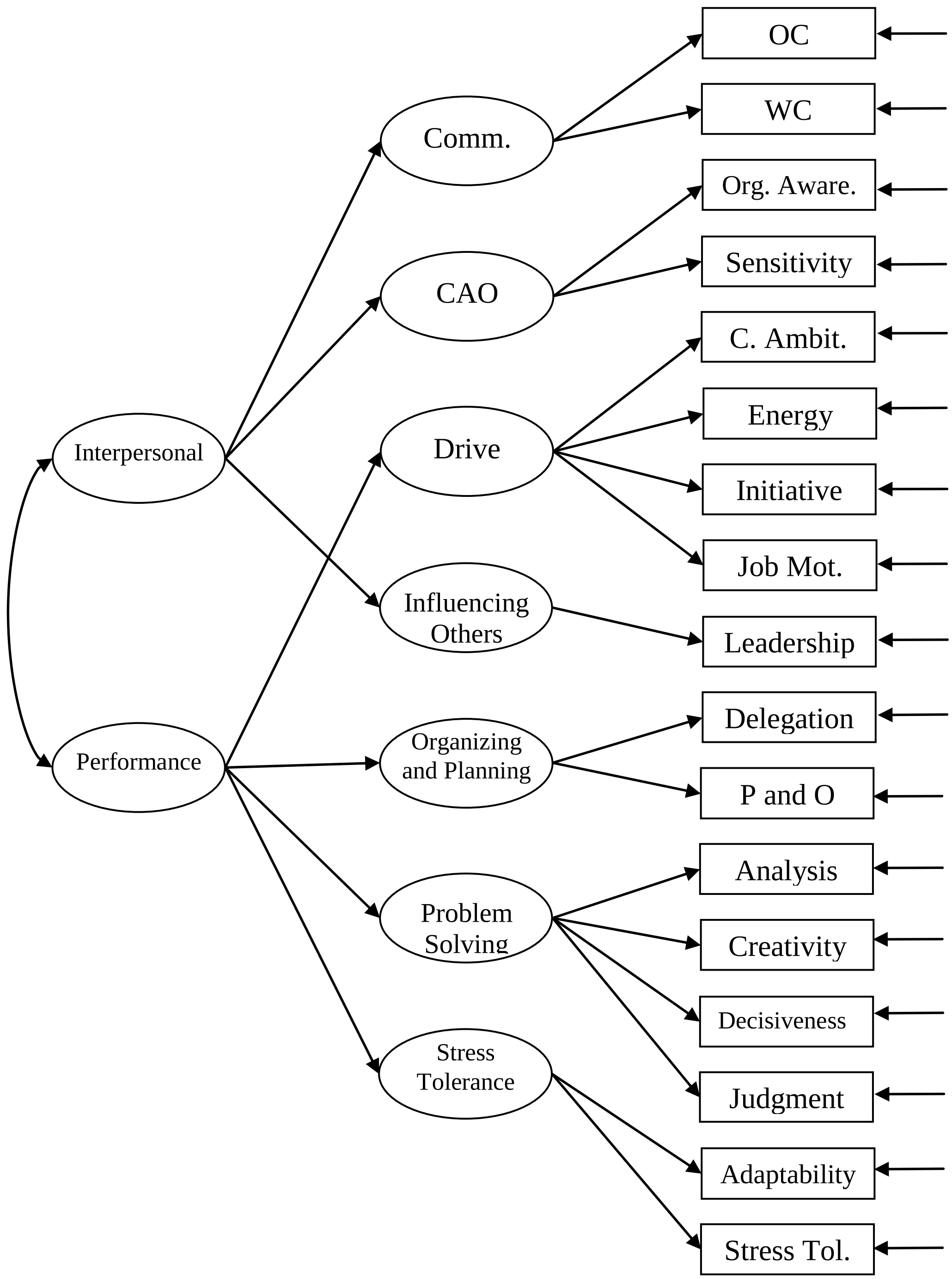


Figure 9. *Hierarchical CFA Model for Shore et al.'s (1992) Two-Factor Framework*

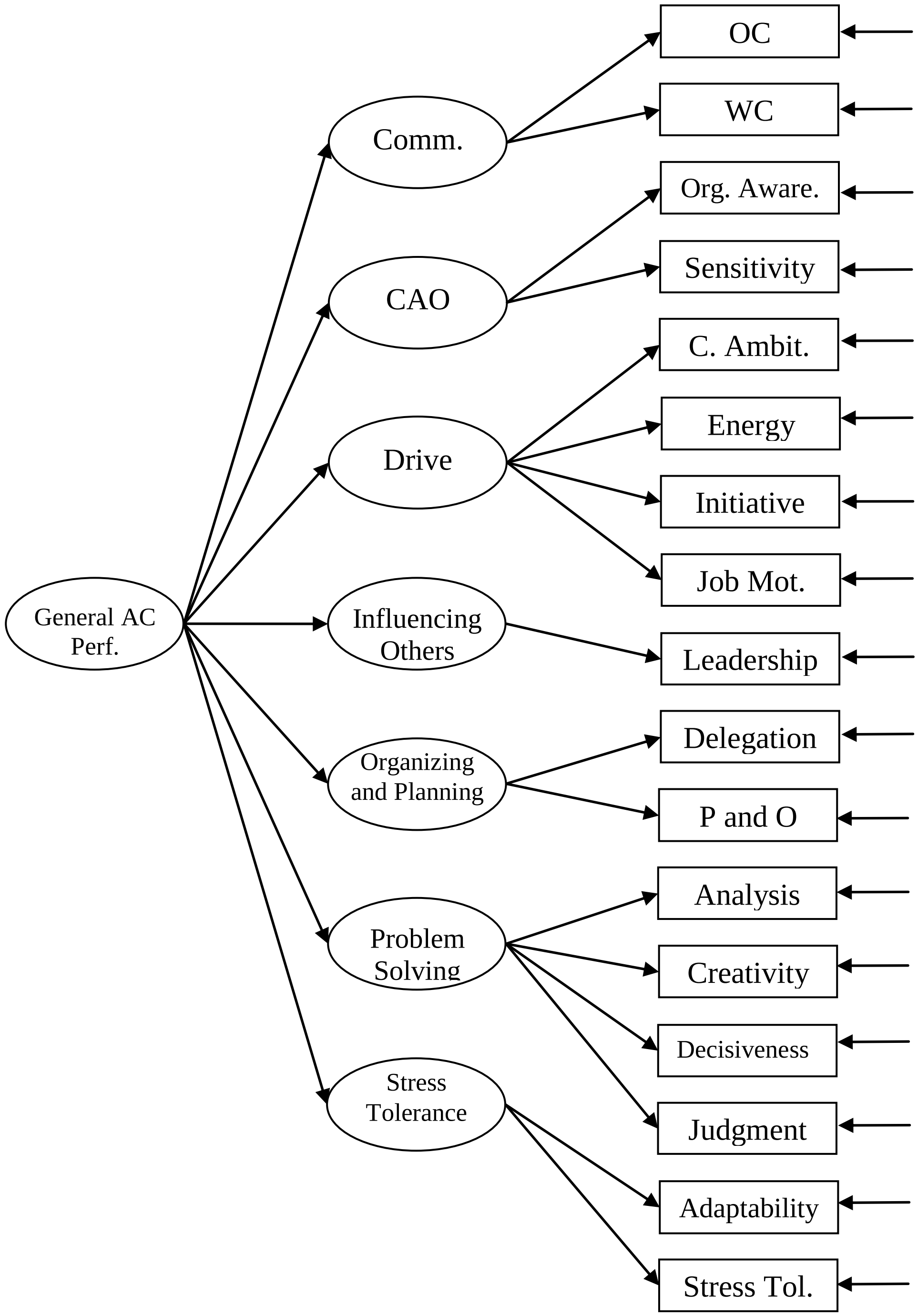


Figure 10. *Hierarchical CFA Model for Viswesvaran et al.'s (2005) One-Factor Framework*

Appendix J

LISREL Syntax and Output for the CFA Models

Model FO1 - Arthur et al. (2003)
DA NI = 17 NO = 918
km sy

1.00

0.41 1.00

0.43 0.25 1.00

0.45 0.37 0.46 1.00

0.45 0.30 0.31 0.26 1.00

0.56 0.28 0.43 0.43 0.48 1.00

0.45 0.18 0.61 0.33 0.64 0.52 1.00

0.49 0.37 0.68 0.44 0.64 0.53 0.64 1.00

0.54 0.32 0.50 0.46 0.33 0.64 0.57 0.41 1.00

0.31 0.43 0.61 0.21 0.62 0.56 0.59 0.61 0.46 1.00

0.46 0.42 0.40 0.38 0.35 0.50 0.44 0.43 0.51 0.45 1.00

0.41 0.36 0.40 0.41 0.34 0.43 0.34 0.47 0.47 0.60 0.59 1.00

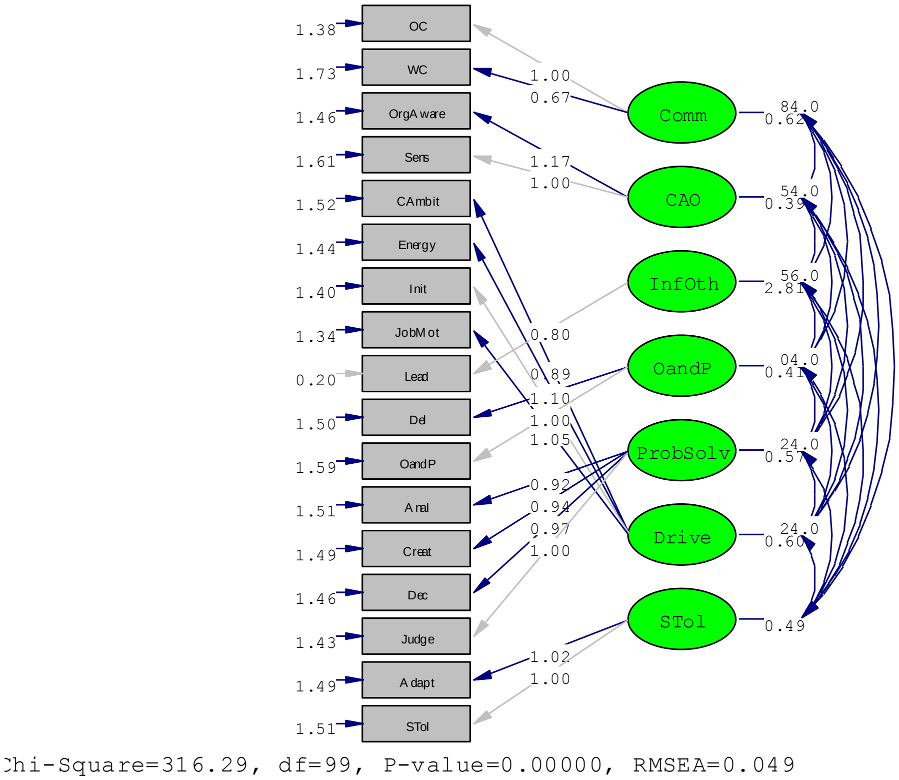
0.46 0.28 0.64 0.41 0.50 0.50 0.38 0.57 0.49 0.57 0.45 0.28 1.00

0.49 0.32 0.21 0.36 0.33 0.34 0.51 0.44 0.40 0.63 0.56 0.48 0.67
1.00
0.45 0.38 0.44 0.43 0.36 0.44 0.44 0.44 0.52 0.48 0.65 0.61 0.53
0.57 1.00
0.45 0.23 0.63 0.52 0.34 0.51 0.40 0.34 0.56 0.45 0.36 0.38 0.39
0.28 0.41 1.00
0.55 0.36 0.39 0.44 0.32 0.50 0.38 0.54 0.49 0.42 0.46 0.47 0.32
0.50 0.49 0.50 1.00

LA
OC WC OrgAware Sens CAmbit Energy Init JobMot Lead Del OandP Anal Creat Dec Judge Adapt STol

MO NX = 17 NK = 7 TD = FR PH = SY, FR

```
LK
Comm CAO InfOth OandP ProbSolv Drive STol

PA LX
1 0 0 0 0 0 0
1 0 0 0 0 0 0
0 1 0 0 0 0 0
0 1 0 0 0 0 0
0 0 0 0 0 1 0
0 0 0 0 0 1 0
0 0 0 0 0 1 0
0 0 0 0 0 1 0
0 0 1 0 0 0 0
0 0 0 1 0 0 0
0 0 0 1 0 0 0
0 0 0 0 1 0 0
0 0 0 0 1 0 0
0 0 0 0 1 0 0
0 0 0 0 1 0 0
0 0 0 0 0 0 1
0 0 0 0 0 0 1

st 1.0 lx 1 1 lx 4 2 lx 7 6 lx 11 4 lx 15 5 lx 17 7
fi lx 1 1 lx 4 2 lx 7 6 lx 11 4 lx 15 5 lx 17 7
st .80 lx 9 3
fi lx 9 3
st .20 td 9 9
fi td 9 9

PD

OU MI EF MR SS SC AD=OFF  IT= 5000
```


OC
WC
OrgAware
Sens
CAmbit
Energy
Init
JobMot
Lead
Del
OandP
Anal
Creat
Dec
Judge
Adapt
STol
Comm
CAO
InfOth
OandP
ProbSolv
Drive
STol
1.38
1.73
1.46
1.61
1.52
1.44
1.40
1.34
0.20
1.50
1.59
1.51
1.49
1.46
1.43
1.49
1.51
1.00
0.67
1.17
1.00
0.80
0.89
1.10
1.00
1.05
0.92
0.94
0.97
1.00
1.02
1.00
0.62
0.39
2.81
0.41
0.57
0.60
0.49
Chi-Square=316.29, df=99, P-value=0.00000, RMSEA=0.049

Goodness of Fit Statistics

Degrees of Freedom = 99
Minimum Fit Function Chi-Square = 329.96 (P = 0.0)
Normal Theory Weighted Least Squares Chi-Square = 316.29 (P = 0.0)
Estimated Non-centrality Parameter (NCP) = 217.29
90 Percent Confidence Interval for NCP = (167.32 ; 274.88)

Minimum Fit Function Value = 0.36
Population Discrepancy Function Value (F0) = 0.24
90 Percent Confidence Interval for F0 = (0.18 ; 0.30)
Root Mean Square Error of Approximation (RMSEA) = 0.049
90 Percent Confidence Interval for RMSEA = (0.043 ; 0.055)
P-Value for Test of Close Fit (RMSEA < 0.05) = 0.60

Expected Cross-Validation Index (ECVI) = 0.46
90 Percent Confidence Interval for ECVI = (0.41 ; 0.53)
ECVI for Saturated Model = 0.33
ECVI for Independence Model = 7.29

Chi-Square for Independence Model with 136 Degrees of Freedom = 6652.90
Independence AIC = 6686.90
Model AIC = 424.29
Saturated AIC = 306.00
Independence CAIC = 6785.88
Model CAIC = 738.69
Saturated CAIC = 1196.80

Normed Fit Index (NFI) = 0.95
Non-Normed Fit Index (NNFI) = 0.95
Parsimony Normed Fit Index (PNFI) = 0.69
Comparative Fit Index (CFI) = 0.96
Incremental Fit Index (IFI) = 0.96
Relative Fit Index (RFI) = 0.93

Critical N (CN) = 375.20

Root Mean Square Residual (RMR) = 0.074
Standardized RMR = 0.037
Goodness of Fit Index (GFI) = 0.96
Adjusted Goodness of Fit Index (AGFI) = 0.94
Parsimony Goodness of Fit Index (PGFI) = 0.62

Model FO2 - Direct Test of Borman and Brush (1993) From Figure 1
DA NI = 17 NO = 918
km sy

1.00

0.41 1.00

0.43 0.25 1.00

0.45 0.37 0.46 1.00

0.45 0.30 0.31 0.26 1.00

0.56 0.28 0.43 0.43 0.48 1.00

0.45 0.18 0.61 0.33 0.64 0.52 1.00

0.49 0.37 0.68 0.44 0.64 0.53 0.64 1.00

0.54 0.32 0.50 0.46 0.33 0.64 0.57 0.41 1.00

0.31 0.43 0.61 0.21 0.62 0.56 0.59 0.61 0.46 1.00

0.46 0.42 0.40 0.38 0.35 0.50 0.44 0.43 0.51 0.45 1.00

0.41 0.36 0.40 0.41 0.34 0.43 0.34 0.47 0.47 0.60 0.59 1.00

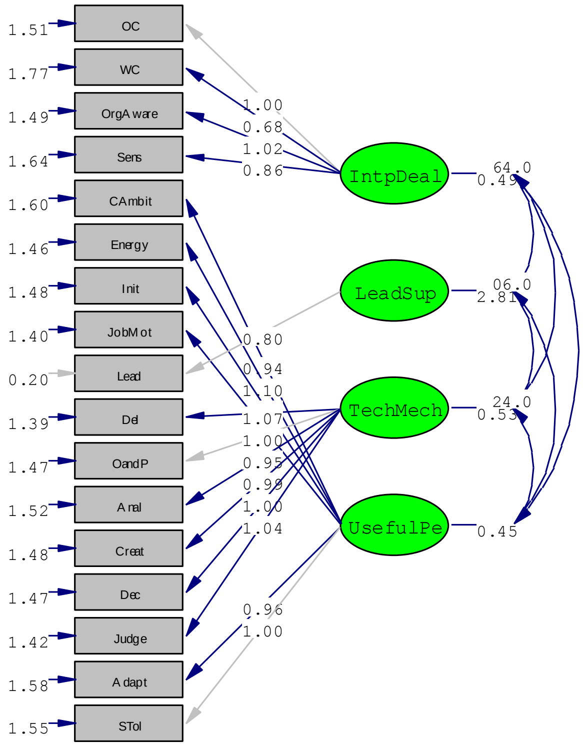
0.46 0.28 0.64 0.41 0.50 0.50 0.38 0.57 0.49 0.57 0.45 0.28 1.00

0.49 0.32 0.21 0.36 0.33 0.34 0.51 0.44 0.40 0.63 0.56 0.48 0.67
1.00
0.45 0.38 0.44 0.43 0.36 0.44 0.44 0.44 0.52 0.48 0.65 0.61 0.53
0.57 1.00
0.45 0.23 0.63 0.52 0.34 0.51 0.40 0.34 0.56 0.45 0.36 0.38 0.39
0.28 0.41 1.00
0.55 0.36 0.39 0.44 0.32 0.50 0.38 0.54 0.49 0.42 0.46 0.47 0.32
0.50 0.49 0.50 1.00

LA
OC WC OrgAware Sens CAmbit Energy Init JobMot Lead Del OandP Anal Creat Dec Judge
Adapt STol

MO NX = 17 NK = 4 TD = FR PH = SY, FR

```
LK
IntpDealComm LeadSup TechMechMgt UsefulPersB

PA LX
1 0 0 0
1 0 0 0
1 0 0 0
1 0 0 0
0 0 0 1
0 0 0 1
0 0 0 1
0 0 0 1
0 1 0 0
0 0 1 0
0 0 1 0
0 0 1 0
0 0 1 0
0 0 1 0
0 0 1 0
0 0 0 1
0 0 0 1

st 1.0 lx 1 1 lx 11 3 lx 17 4
fi lx 1 1 lx 11 3 lx 17 4
st .80 lx 9 2
fi lx 9 2
st .20 td 9 9
fi td 9 9

PD

OU MI EF MR SS SC AD=OFF  IT= 5000
```



Chi-Square=374.60, df=114, P-value=0.00000, RMSEA=0.050

Goodness of Fit Statistics

Degrees of Freedom = 114
Minimum Fit Function Chi-Square = 380.62 (P = 0.0)
Normal Theory Weighted Least Squares Chi-Square = 374.60 (P = 0.0)
Estimated Non-centrality Parameter (NCP) = 260.60
90 Percent Confidence Interval for NCP = (205.73 ; 323.07)

Minimum Fit Function Value = 0.42
Population Discrepancy Function Value (F0) = 0.28
90 Percent Confidence Interval for F0 = (0.22 ; 0.35)
Root Mean Square Error of Approximation (RMSEA) = 0.050
90 Percent Confidence Interval for RMSEA = (0.044 ; 0.056)
P-Value for Test of Close Fit (RMSEA < 0.05) = 0.50

Expected Cross-Validation Index (ECVI) = 0.49
90 Percent Confidence Interval for ECVI = (0.43 ; 0.56)
ECVI for Saturated Model = 0.33
ECVI for Independence Model = 7.29

Chi-Square for Independence Model with 136 Degrees of Freedom = 6652.90
Independence AIC = 6686.90
Model AIC = 452.60
Saturated AIC = 306.00
Independence CAIC = 6785.88
Model CAIC = 679.66
Saturated CAIC = 1196.80

Normed Fit Index (NFI) = 0.94
Non-Normed Fit Index (NNFI) = 0.95
Parsimony Normed Fit Index (PNFI) = 0.79
Comparative Fit Index (CFI) = 0.96
Incremental Fit Index (IFI) = 0.96
Relative Fit Index (RFI) = 0.93

Critical N (CN) = 367.29

Root Mean Square Residual (RMR) = 0.079
Standardized RMR = 0.040
Goodness of Fit Index (GFI) = 0.95
Adjusted Goodness of Fit Index (AGFI) = 0.94
Parsimony Goodness of Fit Index (PGFI) = 0.71

Model FO3 - Direct Test of Schmitt (1977) From Figure 1
DA NI = 17 NO = 918
km sy

1.00

0.41 1.00

0.43 0.25 1.00

0.45 0.37 0.46 1.00

0.45 0.30 0.31 0.26 1.00

0.56 0.28 0.43 0.43 0.48 1.00

0.45 0.18 0.61 0.33 0.64 0.52 1.00

0.49 0.37 0.68 0.44 0.64 0.53 0.64 1.00

0.54 0.32 0.50 0.46 0.33 0.64 0.57 0.41 1.00

0.31 0.43 0.61 0.21 0.62 0.56 0.59 0.61 0.46 1.00

0.46 0.42 0.40 0.38 0.35 0.50 0.44 0.43 0.51 0.45 1.00

0.41 0.36 0.40 0.41 0.34 0.43 0.34 0.47 0.47 0.60 0.59 1.00

0.46 0.28 0.64 0.41 0.50 0.50 0.38 0.57 0.49 0.57 0.45 0.28 1.00

0.49 0.32 0.21 0.36 0.33 0.34 0.51 0.44 0.40 0.63 0.56 0.48 0.67
1.00
0.45 0.38 0.44 0.43 0.36 0.44 0.44 0.44 0.52 0.48 0.65 0.61 0.53
0.57 1.00
0.45 0.23 0.63 0.52 0.34 0.51 0.40 0.34 0.56 0.45 0.36 0.38 0.39
0.28 0.41 1.00
0.55 0.36 0.39 0.44 0.32 0.50 0.38 0.54 0.49 0.42 0.46 0.47 0.32
0.50 0.49 0.50 1.00

LA
OC WC OrgAware Sens CAmbit Energy Init JobMot Lead Del OandP Anal Creat Dec Judge
Adapt STol

MO NX = 17 NK = 3 TD = FR PH = SY, FR

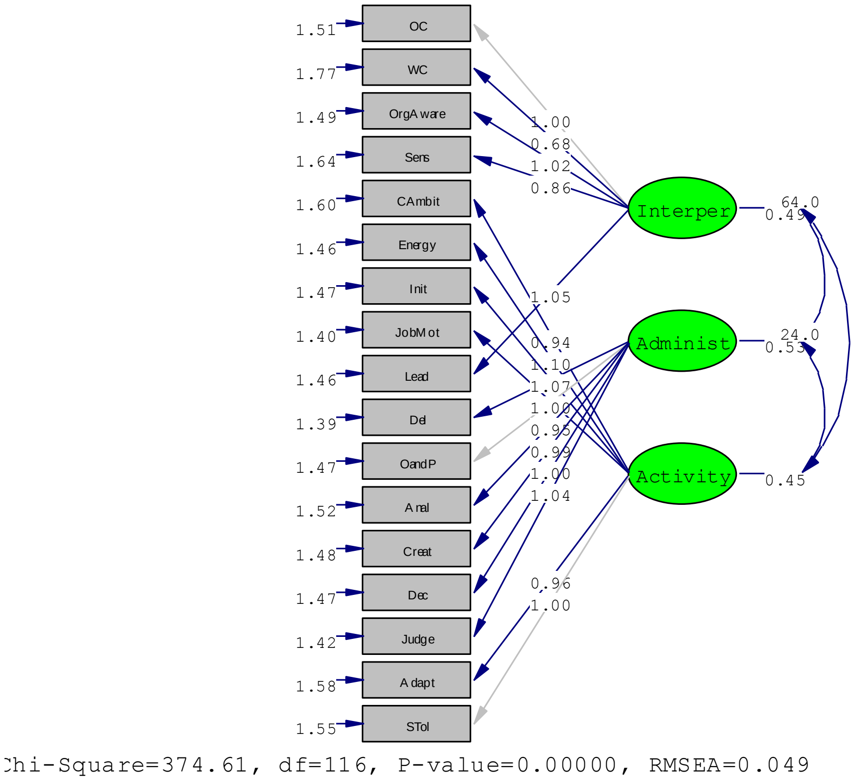
LK
Interpersonal Administrative Activity

PA LX
1 0 0
1 0 0
1 0 0
1 0 0
0 0 1
0 0 1
0 0 1
0 0 1
1 0 0
0 1 0
0 1 0
0 1 0
0 1 0
0 1 0
0 1 0
0 0 1
0 0 1

st 1.0 lx 1 1 lx 11 2 lx 17 3
fi lx 1 1 lx 11 2 lx 17 3

PD

OU MI EF MR SS SC AD=OFF IT= 5000


OC
WC
OrgAware
Sens
CAmbit
Energy
Init
JobMot
Lead
Del
OandP
Anal
Creat
Dec
Judge
Adapt
STol
Interper
Administ
Activity
1.51
1.77
1.49
1.64
1.60
1.46
1.47
1.40
1.46
1.39
1.47
1.52
1.48
1.47
1.42
1.58
1.55
1.00
0.68
1.02
0.86
1.05
0.94
1.10
1.07
1.00
0.95
0.99
1.00
1.04
0.96
1.00
64.0
0.49
24.0
0.53
0.45
Chi-Square=374.61, df=116, P-value=0.00000, RMSEA=0.049

Goodness of Fit Statistics

Degrees of Freedom = 116
Minimum Fit Function Chi-Square = 380.72 (P = 0.0)
Normal Theory Weighted Least Squares Chi-Square = 374.61 (P = 0.0)
Estimated Non-centrality Parameter (NCP) = 258.61
90 Percent Confidence Interval for NCP = (203.84 ; 321.00)

Minimum Fit Function Value = 0.42
Population Discrepancy Function Value (F0) = 0.28
90 Percent Confidence Interval for F0 = (0.22 ; 0.35)
Root Mean Square Error of Approximation (RMSEA) = 0.049
90 Percent Confidence Interval for RMSEA = (0.044 ; 0.055)
P-Value for Test of Close Fit (RMSEA < 0.05) = 0.57

Expected Cross-Validation Index (ECVI) = 0.49
90 Percent Confidence Interval for ECVI = (0.43 ; 0.56)
ECVI for Saturated Model = 0.33
ECVI for Independence Model = 7.29

Chi-Square for Independence Model with 136 Degrees of Freedom = 6652.90
Independence AIC = 6686.90
Model AIC = 448.61
Saturated AIC = 306.00
Independence CAIC = 6785.88
Model CAIC = 664.03
Saturated CAIC = 1196.80

Normed Fit Index (NFI) = 0.94
Non-Normed Fit Index (NNFI) = 0.95
Parsimony Normed Fit Index (PNFI) = 0.80
Comparative Fit Index (CFI) = 0.96
Incremental Fit Index (IFI) = 0.96
Relative Fit Index (RFI) = 0.93

Critical N (CN) = 372.75

Root Mean Square Residual (RMR) = 0.079
Standardized RMR = 0.040
Goodness of Fit Index (GFI) = 0.95
Adjusted Goodness of Fit Index (AGFI) = 0.94
Parsimony Goodness of Fit Index (PGFI) = 0.72

Model FO4 - Direct Test of Shore et al. (1992) From Figure 1
DA NI = 17 NO = 918
km sy

1.00

0.41 1.00

0.43 0.25 1.00

0.45 0.37 0.46 1.00

0.45 0.30 0.31 0.26 1.00

0.56 0.28 0.43 0.43 0.48 1.00

0.45 0.18 0.61 0.33 0.64 0.52 1.00

0.49 0.37 0.68 0.44 0.64 0.53 0.64 1.00

0.54 0.32 0.50 0.46 0.33 0.64 0.57 0.41 1.00

0.31 0.43 0.61 0.21 0.62 0.56 0.59 0.61 0.46 1.00

0.46 0.42 0.40 0.38 0.35 0.50 0.44 0.43 0.51 0.45 1.00

0.41 0.36 0.40 0.41 0.34 0.43 0.34 0.47 0.47 0.60 0.59 1.00

0.46 0.28 0.64 0.41 0.50 0.50 0.38 0.57 0.49 0.57 0.45 0.28 1.00

0.49 0.32 0.21 0.36 0.33 0.34 0.51 0.44 0.40 0.63 0.56 0.48 0.67
1.00
0.45 0.38 0.44 0.43 0.36 0.44 0.44 0.44 0.52 0.48 0.65 0.61 0.53
0.57 1.00
0.45 0.23 0.63 0.52 0.34 0.51 0.40 0.34 0.56 0.45 0.36 0.38 0.39
0.28 0.41 1.00
0.55 0.36 0.39 0.44 0.32 0.50 0.38 0.54 0.49 0.42 0.46 0.47 0.32
0.50 0.49 0.50 1.00

LA
OC WC OrgAware Sens CAmbit Energy Init JobMot Lead Del OandP Anal Creat Dec Judge
Adapt STol

MO NX = 17 NK = 2 TD = FR PH = SY, FR

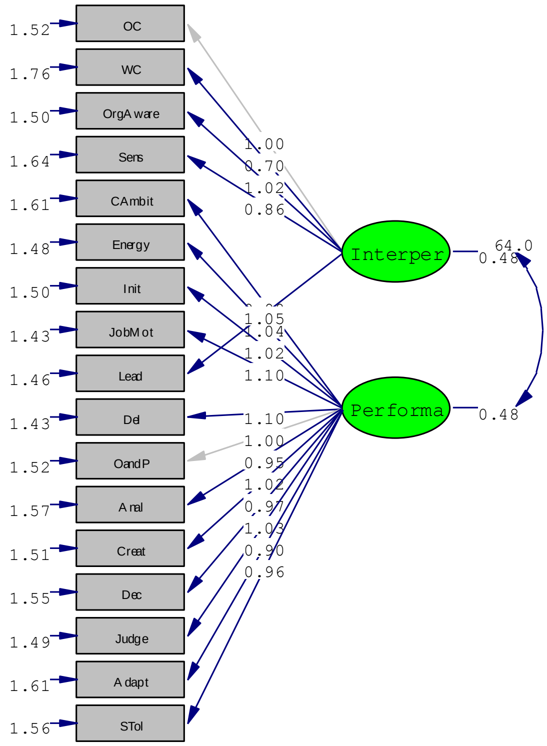
LK
Interpersonal Performance

PA LX
1 0
1 0
1 0
1 0
0 1
0 1
0 1
0 1
1 0
0 1
0 1
0 1
0 1
0 1
0 1
0 1
0 1

st 1.0 lx 1 1 lx 11 2
fi lx 1 1 lx 11 2

PD

OU MI EF MR SS SC AD=OFF IT= 5000



Chi-Square=411.98, df=118, P-value=0.00000, RMSEA=0.052

Goodness of Fit Statistics

Degrees of Freedom = 118
Minimum Fit Function Chi-Square = 418.67 (P = 0.0)
Normal Theory Weighted Least Squares Chi-Square = 411.98 (P = 0.0)
Estimated Non-centrality Parameter (NCP) = 293.98
90 Percent Confidence Interval for NCP = (235.93 ; 359.63)

Minimum Fit Function Value = 0.46
Population Discrepancy Function Value (F0) = 0.32
90 Percent Confidence Interval for F0 = (0.26 ; 0.39)
Root Mean Square Error of Approximation (RMSEA) = 0.052
90 Percent Confidence Interval for RMSEA = (0.047 ; 0.058)
P-Value for Test of Close Fit (RMSEA < 0.05) = 0.25

Expected Cross-Validation Index (ECVI) = 0.53
90 Percent Confidence Interval for ECVI = (0.46 ; 0.60)
ECVI for Saturated Model = 0.33
ECVI for Independence Model = 7.29

Chi-Square for Independence Model with 136 Degrees of Freedom = 6652.90
Independence AIC = 6686.90
Model AIC = 481.98
Saturated AIC = 306.00
Independence CAIC = 6785.88
Model CAIC = 685.76
Saturated CAIC = 1196.80

Normed Fit Index (NFI) = 0.94
Non-Normed Fit Index (NNFI) = 0.95
Parsimony Normed Fit Index (PNFI) = 0.81
Comparative Fit Index (CFI) = 0.95
Incremental Fit Index (IFI) = 0.95
Relative Fit Index (RFI) = 0.93

Critical N (CN) = 344.10

Root Mean Square Residual (RMR) = 0.083
Standardized RMR = 0.042
Goodness of Fit Index (GFI) = 0.95
Adjusted Goodness of Fit Index (AGFI) = 0.93
Parsimony Goodness of Fit Index (PGFI) = 0.73

```
Model FO5 - Direct Test of Viswesvaran et al. (2005) From Figure 1
DA NI = 17 NO = 918
km sy

1.00

0.41 1.00

0.43 0.25 1.00

0.45 0.37 0.46 1.00

0.45 0.30 0.31 0.26 1.00

0.56 0.28 0.43 0.43 0.48 1.00

0.45 0.18 0.61 0.33 0.64 0.52 1.00

0.49 0.37 0.68 0.44 0.64 0.53 0.64 1.00

0.54 0.32 0.50 0.46 0.33 0.64 0.57 0.41 1.00

0.31 0.43 0.61 0.21 0.62 0.56 0.59 0.61 0.46 1.00

0.46 0.42 0.40 0.38 0.35 0.50 0.44 0.43 0.51 0.45 1.00

0.41 0.36 0.40 0.41 0.34 0.43 0.34 0.47 0.47 0.60 0.59 1.00

0.46 0.28 0.64 0.41 0.50 0.50 0.38 0.57 0.49 0.57 0.45 0.28 1.00

0.49 0.32 0.21 0.36 0.33 0.34 0.51 0.44 0.40 0.63 0.56 0.48 0.67
     1.00
0.45 0.38 0.44 0.43 0.36 0.44 0.44 0.44 0.52 0.48 0.65 0.61 0.53
     0.57 1.00
0.45 0.23 0.63 0.52 0.34 0.51 0.40 0.34 0.56 0.45 0.36 0.38 0.39
     0.28 0.41 1.00
0.55 0.36 0.39 0.44 0.32 0.50 0.38 0.54 0.49 0.42 0.46 0.47 0.32
     0.50 0.49 0.50 1.00

LA
OC WC OrgAware Sens CAmbit Energy Init JobMot Lead Del OandP Anal Creat Dec Judge
Adapt STol

MO NX = 17 NK = 1 TD = FR PH = SY, FR
```

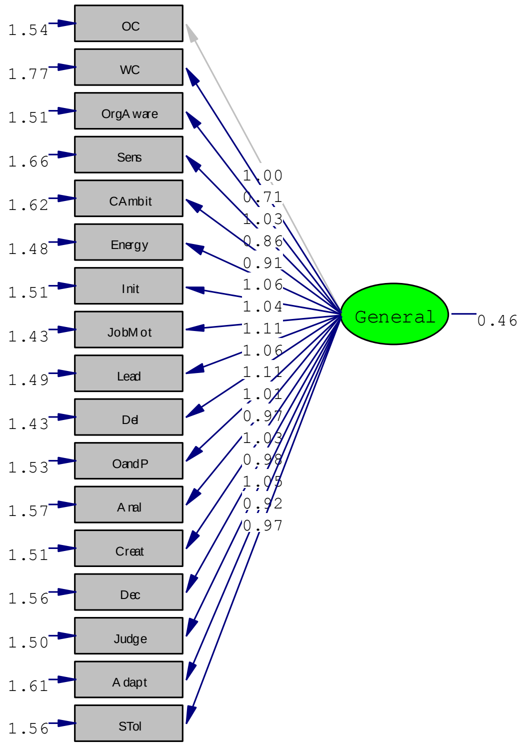
LK
General

PA LX
1
1
1
1
1
1
1
1
1
1
1
1
1
1
1
1
1

st 1.0 lx 1 1
fi lx 1 1

PD

OU MI EF MR SS SC AD=OFF IT= 5000



Chi-Square=415.59, df=119, P-value=0.00000, RMSEA=0.052

Goodness of Fit Statistics

Degrees of Freedom = 119
Minimum Fit Function Chi-Square = 421.11 (P = 0.0)
Normal Theory Weighted Least Squares Chi-Square = 415.59 (P = 0.0)
Estimated Non-centrality Parameter (NCP) = 296.59
90 Percent Confidence Interval for NCP = (238.26 ; 362.50)

Minimum Fit Function Value = 0.46
Population Discrepancy Function Value (F0) = 0.32
90 Percent Confidence Interval for F0 = (0.26 ; 0.40)
Root Mean Square Error of Approximation (RMSEA) = 0.052
90 Percent Confidence Interval for RMSEA = (0.047 ; 0.058)
P-Value for Test of Close Fit (RMSEA < 0.05) = 0.25

Expected Cross-Validation Index (ECVI) = 0.53
90 Percent Confidence Interval for ECVI = (0.46 ; 0.60)
ECVI for Saturated Model = 0.33
ECVI for Independence Model = 7.29

Chi-Square for Independence Model with 136 Degrees of Freedom = 6652.90
Independence AIC = 6686.90
Model AIC = 483.59
Saturated AIC = 306.00
Independence CAIC = 6785.88
Model CAIC = 681.54
Saturated CAIC = 1196.80

Normed Fit Index (NFI) = 0.94
Non-Normed Fit Index (NNFI) = 0.95
Parsimony Normed Fit Index (PNFI) = 0.82
Comparative Fit Index (CFI) = 0.95
Incremental Fit Index (IFI) = 0.95
Relative Fit Index (RFI) = 0.93

Critical N (CN) = 344.62

Root Mean Square Residual (RMR) = 0.084
Standardized RMR = 0.042
Goodness of Fit Index (GFI) = 0.95
Adjusted Goodness of Fit Index (AGFI) = 0.93
Parsimony Goodness of Fit Index (PGFI) = 0.74

```
Model H1: Hierarchical 4-Factor
DA NI = 17 NO = 918
km sy

1.00

0.41 1.00

0.43 0.25 1.00

0.45 0.37 0.46 1.00

0.45 0.30 0.31 0.26 1.00

0.56 0.28 0.43 0.43 0.48 1.00

0.45 0.18 0.61 0.33 0.64 0.52 1.00

0.49 0.37 0.68 0.44 0.64 0.53 0.64 1.00

0.54 0.32 0.50 0.46 0.33 0.64 0.57 0.41 1.00

0.31 0.43 0.61 0.21 0.62 0.56 0.59 0.61 0.46 1.00

0.46 0.42 0.40 0.38 0.35 0.50 0.44 0.43 0.51 0.45 1.00

0.41 0.36 0.40 0.41 0.34 0.43 0.34 0.47 0.47 0.60 0.59 1.00

0.46 0.28 0.64 0.41 0.50 0.50 0.38 0.57 0.49 0.57 0.45 0.28 1.00

0.49 0.32 0.21 0.36 0.33 0.34 0.51 0.44 0.40 0.63 0.56 0.48 0.67
     1.00
0.45 0.38 0.44 0.43 0.36 0.44 0.44 0.44 0.52 0.48 0.65 0.61 0.53
     0.57 1.00
0.45 0.23 0.63 0.52 0.34 0.51 0.40 0.34 0.56 0.45 0.36 0.38 0.39
     0.28 0.41 1.00
0.55 0.36 0.39 0.44 0.32 0.50 0.38 0.54 0.49 0.42 0.46 0.47 0.32
     0.50 0.49 0.50 1.00

LA
OC WC OrgAware Sens CAmbit Energy Init JobMot Lead Del OandP Anal Creat Dec Judge
Adapt STol

MO Ny = 17 Nk = 3 Ne = 7  PH = St be=fu,fi ga=fu,fi ps=di
```

Le
Comm CAO InfOth OandP ProbSolv Drive STol

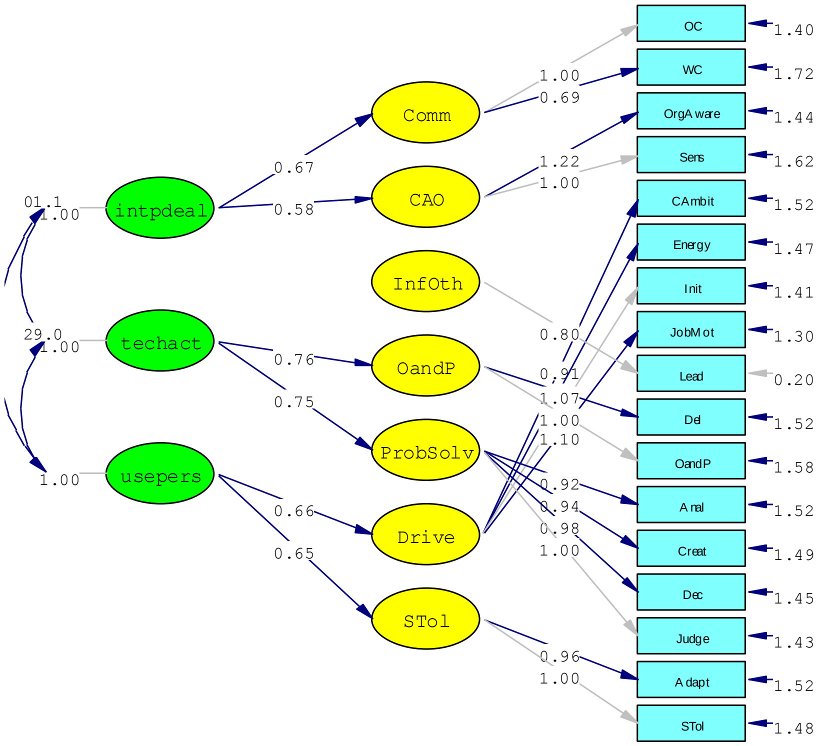
lk
intpdeal techact usepers
pa ga
1 0 0
1 0 0
0 0 0
0 1 0
0 1 0
0 0 1
0 0 1

PA Ly
1 0 0 0 0 0 0
1 0 0 0 0 0 0
0 1 0 0 0 0 0
0 1 0 0 0 0 0
0 0 0 0 0 1 0
0 0 0 0 0 1 0
0 0 0 0 0 1 0
0 0 0 0 0 1 0
0 0 1 0 0 0 0
0 0 0 1 0 0 0
0 0 0 1 0 0 0
0 0 0 0 1 0 0
0 0 0 0 1 0 0
0 0 0 0 1 0 0
0 0 0 0 1 0 0
0 0 0 0 0 0 1
0 0 0 0 0 0 1

st 1.0 ly 1 1 ly 4 2 ly 7 6 ly 11 4 ly 15 5 ly 17 7
fi ly 1 1 ly 4 2 ly 7 6 ly 11 4 ly 15 5 ly 17 7
st .80 ly 9 3
fi ly 9 3
st .20 te 9 9
fi te 9 9

PD

OU MI EF MR SS SC AD=OFF IT= 5000



Chi-Square=529.13, df=111, P-value=0.00000, RMSEA=0.064

Goodness of Fit Statistics

Degrees of Freedom = 111
Minimum Fit Function Chi-Square = 567.36 (P = 0.0)
Normal Theory Weighted Least Squares Chi-Square = 529.13 (P = 0.0)
Estimated Non-centrality Parameter (NCP) = 418.13
90 Percent Confidence Interval for NCP = (350.26 ; 493.53)

Minimum Fit Function Value = 0.62
Population Discrepancy Function Value (F0) = 0.46
90 Percent Confidence Interval for F0 = (0.38 ; 0.54)
Root Mean Square Error of Approximation (RMSEA) = 0.064
90 Percent Confidence Interval for RMSEA = (0.059 ; 0.070)
P-Value for Test of Close Fit (RMSEA < 0.05) = 0.00

Expected Cross-Validation Index (ECVI) = 0.67
90 Percent Confidence Interval for ECVI = (0.59 ; 0.75)
ECVI for Saturated Model = 0.33
ECVI for Independence Model = 7.29

Chi-Square for Independence Model with 136 Degrees of Freedom = 6652.90
Independence AIC = 6686.90
Model AIC = 613.13
Saturated AIC = 306.00
Independence CAIC = 6785.88
Model CAIC = 857.66
Saturated CAIC = 1196.80

Normed Fit Index (NFI) = 0.91
Non-Normed Fit Index (NNFI) = 0.91
Parsimony Normed Fit Index (PNFI) = 0.75
Comparative Fit Index (CFI) = 0.93
Incremental Fit Index (IFI) = 0.93
Relative Fit Index (RFI) = 0.90

Critical N (CN) = 241.13

Root Mean Square Residual (RMR) = 0.17
Standardized RMR = 0.087
Goodness of Fit Index (GFI) = 0.94
Adjusted Goodness of Fit Index (AGFI) = 0.91
Parsimony Goodness of Fit Index (PGFI) = 0.68

Model H2: Hierarchical 3-Factor
DA NI = 17 NO = 918
km sy

1.00

0.41 1.00

0.43 0.25 1.00

0.45 0.37 0.46 1.00

0.45 0.30 0.31 0.26 1.00

0.56 0.28 0.43 0.43 0.48 1.00

0.45 0.18 0.61 0.33 0.64 0.52 1.00

0.49 0.37 0.68 0.44 0.64 0.53 0.64 1.00

0.54 0.32 0.50 0.46 0.33 0.64 0.57 0.41 1.00

0.31 0.43 0.61 0.21 0.62 0.56 0.59 0.61 0.46 1.00

0.46 0.42 0.40 0.38 0.35 0.50 0.44 0.43 0.51 0.45 1.00

0.41 0.36 0.40 0.41 0.34 0.43 0.34 0.47 0.47 0.60 0.59 1.00

0.46 0.28 0.64 0.41 0.50 0.50 0.38 0.57 0.49 0.57 0.45 0.28 1.00

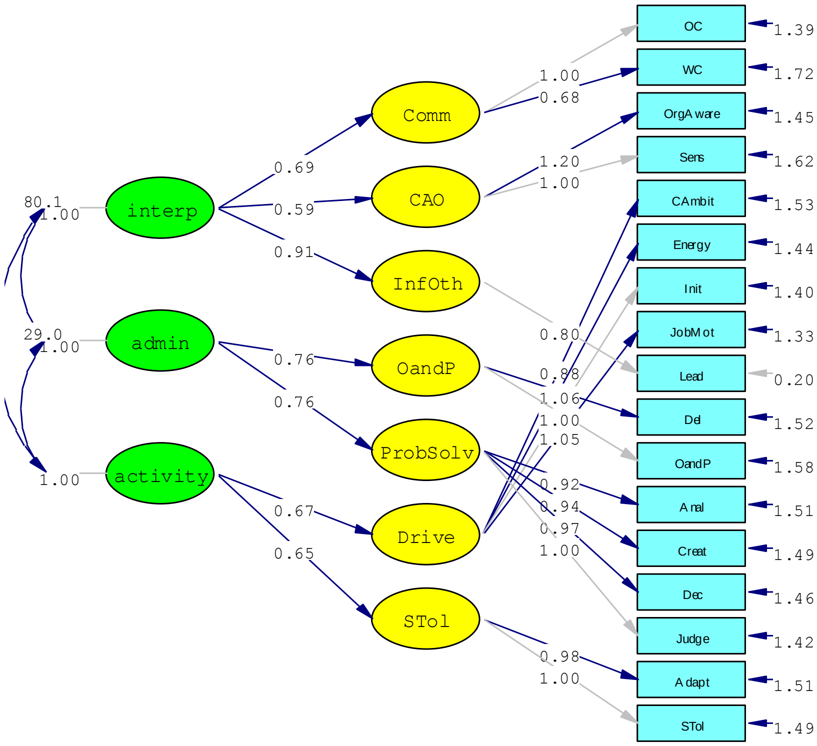
0.49 0.32 0.21 0.36 0.33 0.34 0.51 0.44 0.40 0.63 0.56 0.48 0.67
1.00
0.45 0.38 0.44 0.43 0.36 0.44 0.44 0.44 0.52 0.48 0.65 0.61 0.53
0.57 1.00
0.45 0.23 0.63 0.52 0.34 0.51 0.40 0.34 0.56 0.45 0.36 0.38 0.39
0.28 0.41 1.00
0.55 0.36 0.39 0.44 0.32 0.50 0.38 0.54 0.49 0.42 0.46 0.47 0.32
0.50 0.49 0.50 1.00

LA
OC WC OrgAware Sens CAmbit Energy Init JobMot Lead Del OandP Anal Creat Dec Judge
Adapt STol

MO Ny = 17 Nk = 3 Ne = 7 PH = St be=fu,fi ga=fu,fi ps=di

```
Le
Comm CAO InfOth OandP ProbSolv Drive STol

lk
interp admin activity
pa ga
1 0 0
1 0 0
1 0 0
0 1 0
0 1 0
0 0 1
0 0 1

PA Ly
1 0 0 0 0 0 0
1 0 0 0 0 0 0
0 1 0 0 0 0 0
0 1 0 0 0 0 0
0 0 0 0 0 1 0
0 0 0 0 0 1 0
0 0 0 0 0 1 0
0 0 0 0 0 1 0
0 0 1 0 0 0 0
0 0 0 1 0 0 0
0 0 0 1 0 0 0
0 0 0 0 1 0 0
0 0 0 0 1 0 0
0 0 0 0 1 0 0
0 0 0 0 1 0 0
0 0 0 0 0 0 1
0 0 0 0 0 0 1

st 1.0 ly 1 1 ly 4 2 ly 7 6 ly 11 4 ly 15 5 ly 17 7
fi  ly 1 1 ly 4 2 ly 7 6 ly 11 4 ly 15 5 ly 17 7
st .80 ly 9 3
fi ly 9 3
st .20 te 9 9
fi te 9 9

PD

OU MI EF MR SS SC AD=OFF  IT= 5000
```

AC Dimension Structure 140



Chi-Square=337.34, df=110, P-value=0.00000, RMSEA=0.047

Goodness of Fit Statistics

Degrees of Freedom = 110
Minimum Fit Function Chi-Square = 347.87 (P = 0.0)
Normal Theory Weighted Least Squares Chi-Square = 337.34 (P = 0.0)
Estimated Non-centrality Parameter (NCP) = 227.34
90 Percent Confidence Interval for NCP = (175.83 ; 286.48)

Minimum Fit Function Value = 0.38
Population Discrepancy Function Value (F0) = 0.25
90 Percent Confidence Interval for F0 = (0.19 ; 0.31)
Root Mean Square Error of Approximation (RMSEA) = 0.047
90 Percent Confidence Interval for RMSEA = (0.042 ; 0.053)
P-Value for Test of Close Fit (RMSEA < 0.05) = 0.76

Expected Cross-Validation Index (ECVI) = 0.46
90 Percent Confidence Interval for ECVI = (0.41 ; 0.53)
ECVI for Saturated Model = 0.33
ECVI for Independence Model = 7.29

Chi-Square for Independence Model with 136 Degrees of Freedom = 6652.90
Independence AIC = 6686.90
Model AIC = 423.34
Saturated AIC = 306.00
Independence CAIC = 6785.88
Model CAIC = 673.69
Saturated CAIC = 1196.80

Normed Fit Index (NFI) = 0.95
Non-Normed Fit Index (NNFI) = 0.95
Parsimony Normed Fit Index (PNFI) = 0.77
Comparative Fit Index (CFI) = 0.96
Incremental Fit Index (IFI) = 0.96
Relative Fit Index (RFI) = 0.94

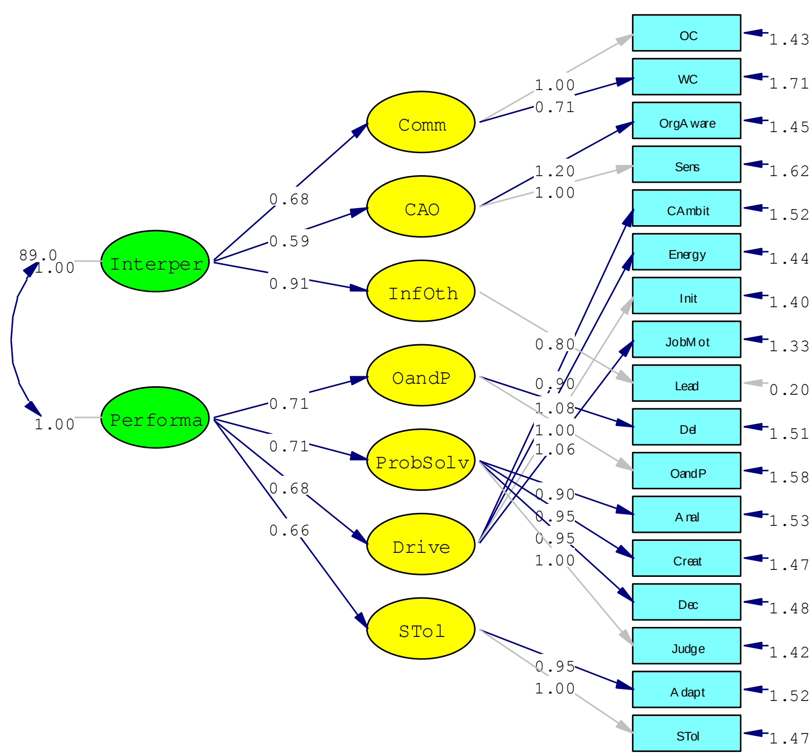
Critical N (CN) = 389.59

Root Mean Square Residual (RMR) = 0.076
Standardized RMR = 0.038
Goodness of Fit Index (GFI) = 0.96
Adjusted Goodness of Fit Index (AGFI) = 0.94
Parsimony Goodness of Fit Index (PGFI) = 0.69

```
Model H3: Hierarchical 2-Factor
DA NI = 17 NO = 918
km sy

1.00

0.41	1.00

0.43	0.25	1.00

0.45	0.37	0.46	1.00

0.45	0.30	0.31	0.26	1.00

0.56	0.28	0.43	0.43	0.48	1.00

0.45	0.18	0.61	0.33	0.64	0.52	1.00

0.49	0.37	0.68	0.44	0.64	0.53	0.64	1.00

0.54	0.32	0.50	0.46	0.33	0.64	0.57	0.41	1.00

0.31	0.43	0.61	0.21	0.62	0.56	0.59	0.61	0.46	1.00

0.46	0.42	0.40	0.38	0.35	0.50	0.44	0.43	0.51	0.45	1.00

0.41	0.36	0.40	0.41	0.34	0.43	0.34	0.47	0.47	0.60	0.59	1.00

0.46	0.28	0.64	0.41	0.50	0.50	0.38	0.57	0.49	0.57	0.45	0.28	1.00

0.49	0.32	0.21	0.36	0.33	0.34	0.51	0.44	0.40	0.63	0.56	0.48	0.67
	1.00
0.45	0.38	0.44	0.43	0.36	0.44	0.44	0.44	0.52	0.48	0.65	0.61	0.53
	0.57	1.00
0.45	0.23	0.63	0.52	0.34	0.51	0.40	0.34	0.56	0.45	0.36	0.38	0.39
	0.28	0.41	1.00
0.55	0.36	0.39	0.44	0.32	0.50	0.38	0.54	0.49	0.42	0.46	0.47	0.32
	0.50	0.49	0.50	1.00

LA
OC WC OrgAware Sens CAmbit Energy Init JobMot Lead Del OandP Anal Creat Dec Judge
Adapt STol

MO Ny = 17 Nk = 2 Ne = 7  PH = St be=fu,fi ga=fu,fi ps=di
```

```
Le
Comm CAO InfOth OandP ProbSolv Drive STol

lk
Interpersonal Performance
pa ga
1 0
1 0
1 0
0 1
0 1
0 1
0 1

PA Ly
1 0 0 0 0 0 0
1 0 0 0 0 0 0
0 1 0 0 0 0 0
0 1 0 0 0 0 0
0 0 0 0 0 1 0
0 0 0 0 0 1 0
0 0 0 0 0 1 0
0 0 0 0 0 1 0
0 0 1 0 0 0 0
0 0 0 1 0 0 0
0 0 0 1 0 0 0
0 0 0 0 1 0 0
0 0 0 0 1 0 0
0 0 0 0 1 0 0
0 0 0 0 1 0 0
0 0 0 0 0 0 1
0 0 0 0 0 0 1

st 1.0 ly 1 1 ly 4 2 ly 7 6 ly 11 4 ly 15 5 ly 17 7
fi  ly 1 1 ly 4 2 ly 7 6 ly 11 4 ly 15 5 ly 17 7
st .80 ly 9 3
fi ly 9 3
st .20 te 9 9
fi te 9 9

PD

OU MI EF MR SS SC AD=OFF  IT= 5000
```

AC Dimension Structure 144



Chi-Square=371.97, df=112, P-value=0.00000, RMSEA=0.050

Goodness of Fit Statistics

Degrees of Freedom = 112
Minimum Fit Function Chi-Square = 380.23 (P = 0.0)
Normal Theory Weighted Least Squares Chi-Square = 371.97 (P = 0.0)
Estimated Non-centrality Parameter (NCP) = 259.97
90 Percent Confidence Interval for NCP = (205.26 ; 322.30)

Minimum Fit Function Value = 0.41
Population Discrepancy Function Value (F0) = 0.28
90 Percent Confidence Interval for F0 = (0.22 ; 0.35)
Root Mean Square Error of Approximation (RMSEA) = 0.050
90 Percent Confidence Interval for RMSEA = (0.045 ; 0.056)
P-Value for Test of Close Fit (RMSEA < 0.05) = 0.45

Expected Cross-Validation Index (ECVI) = 0.50
90 Percent Confidence Interval for ECVI = (0.44 ; 0.56)
ECVI for Saturated Model = 0.33
ECVI for Independence Model = 7.29

Chi-Square for Independence Model with 136 Degrees of Freedom = 6652.90
Independence AIC = 6686.90
Model AIC = 453.97
Saturated AIC = 306.00
Independence CAIC = 6785.88
Model CAIC = 692.68
Saturated CAIC = 1196.80

Normed Fit Index (NFI) = 0.94
Non-Normed Fit Index (NNFI) = 0.95
Parsimony Normed Fit Index (PNFI) = 0.78
Comparative Fit Index (CFI) = 0.96
Incremental Fit Index (IFI) = 0.96
Relative Fit Index (RFI) = 0.93

Critical N (CN) = 362.09

Root Mean Square Residual (RMR) = 0.079
Standardized RMR = 0.040
Goodness of Fit Index (GFI) = 0.95
Adjusted Goodness of Fit Index (AGFI) = 0.94
Parsimony Goodness of Fit Index (PGFI) = 0.70

```
Model H4: Hierarchical 1-Factor
DA NI = 17 NO = 918
km sy

1.00

0.41   1.00

0.43   0.25   1.00

0.45   0.37   0.46   1.00

0.45   0.30   0.31   0.26   1.00

0.56   0.28   0.43   0.43   0.48   1.00

0.45   0.18   0.61   0.33   0.64   0.52   1.00

0.49   0.37   0.68   0.44   0.64   0.53   0.64   1.00

0.54   0.32   0.50   0.46   0.33   0.64   0.57   0.41   1.00

0.31   0.43   0.61   0.21   0.62   0.56   0.59   0.61   0.46   1.00

0.46   0.42   0.40   0.38   0.35   0.50   0.44   0.43   0.51   0.45   1.00

0.41   0.36   0.40   0.41   0.34   0.43   0.34   0.47   0.47   0.60   0.59   1.00

0.46   0.28   0.64   0.41   0.50   0.50   0.38   0.57   0.49   0.57   0.45   0.28   1.00

0.49   0.32   0.21   0.36   0.33   0.34   0.51   0.44   0.40   0.63   0.56   0.48   0.67
       1.00
0.45   0.38   0.44   0.43   0.36   0.44   0.44   0.44   0.52   0.48   0.65   0.61   0.53
       0.57   1.00
0.45   0.23   0.63   0.52   0.34   0.51   0.40   0.34   0.56   0.45   0.36   0.38   0.39
       0.28   0.41   1.00
0.55   0.36   0.39   0.44   0.32   0.50   0.38   0.54   0.49   0.42   0.46   0.47   0.32
       0.50   0.49   0.50   1.00

LA
OC WC OrgAware Sens CAmbit Energy Init JobMot Lead Del OandP Anal Creat Dec Judge
Adapt STol

MO Ny = 17 Nk = 1 Ne = 7  PH = St be=fu,fi ga=fu,fr ps=di
```

Le
Comm CAO InfOth OandP ProbSolv Drive STol

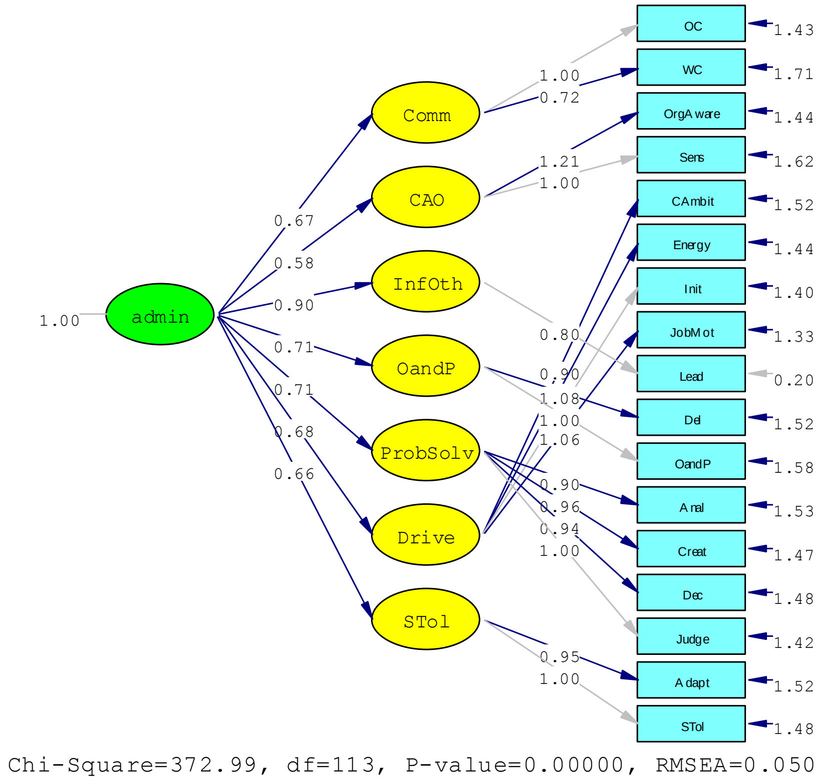
lk
admin interp activity

PA Ly
1 0 0 0 0 0 0
1 0 0 0 0 0 0
0 1 0 0 0 0 0
0 1 0 0 0 0 0
0 0 0 0 0 1 0
0 0 0 0 0 1 0
0 0 0 0 0 1 0
0 0 0 0 0 1 0
0 0 1 0 0 0 0
0 0 0 1 0 0 0
0 0 0 1 0 0 0
0 0 0 0 1 0 0
0 0 0 0 1 0 0
0 0 0 0 1 0 0
0 0 0 0 1 0 0
0 0 0 0 0 0 1
0 0 0 0 0 0 1

st 1.0 ly 1 1 ly 4 2 ly 7 6 ly 11 4 ly 15 5 ly 17 7
fi ly 1 1 ly 4 2 ly 7 6 ly 11 4 ly 15 5 ly 17 7
st .80 ly 9 3
fi ly 9 3
st .20 te 9 9
fi te 9 9

PD

OU MI EF MR SS SC AD=OFF IT= 5000


OC
WC
OrgAware
Sens
CAmbit
Energy
Init
JobMot
Lead
Del
OandP
Anal
Creat
Dec
Judge
Adapt
STol
1.43
1.71
1.44
1.62
1.52
1.44
1.40
1.33
0.20
1.52
1.58
1.53
1.47
1.48
1.42
1.52
1.48
Comm
CAO
InfOth
OandP
ProbSolv
Drive
STol
admin
1.00
0.67
0.58
0.90
0.71
0.71
0.68
0.66
1.00
0.72
1.21
1.00
0.80
0.90
1.08
1.00
1.06
0.90
0.96
0.94
1.00
0.95
1.00
Chi-Square=372.99, df=113, P-value=0.00000, RMSEA=0.050

Goodness of Fit Statistics

Degrees of Freedom = 113
Minimum Fit Function Chi-Square = 380.61 (P = 0.0)
Normal Theory Weighted Least Squares Chi-Square = 372.99 (P = 0.0)
Estimated Non-centrality Parameter (NCP) = 259.99
90 Percent Confidence Interval for NCP = (205.22 ; 322.36)

Minimum Fit Function Value = 0.42
Population Discrepancy Function Value (F0) = 0.28
90 Percent Confidence Interval for F0 = (0.22 ; 0.35)
Root Mean Square Error of Approximation (RMSEA) = 0.050
90 Percent Confidence Interval for RMSEA = (0.045 ; 0.056)
P-Value for Test of Close Fit (RMSEA < 0.05) = 0.48

Expected Cross-Validation Index (ECVI) = 0.49
90 Percent Confidence Interval for ECVI = (0.43 ; 0.56)
ECVI for Saturated Model = 0.33
ECVI for Independence Model = 7.29

Chi-Square for Independence Model with 136 Degrees of Freedom = 6652.90
Independence AIC = 6686.90
Model AIC = 452.99
Saturated AIC = 306.00
Independence CAIC = 6785.88
Model CAIC = 685.87
Saturated CAIC = 1196.80

Normed Fit Index (NFI) = 0.94
Non-Normed Fit Index (NNFI) = 0.95
Parsimony Normed Fit Index (PNFI) = 0.78
Comparative Fit Index (CFI) = 0.96
Incremental Fit Index (IFI) = 0.96
Relative Fit Index (RFI) = 0.93

Critical N (CN) = 364.52

Root Mean Square Residual (RMR) = 0.079
Standardized RMR = 0.040
Goodness of Fit Index (GFI) = 0.95
Adjusted Goodness of Fit Index (AGFI) = 0.94
Parsimony Goodness of Fit Index (PGFI) = 0.70

VITAE

John P. Meriac received his Bachelor of Arts degree in Psychology with a minor in Business Administration from East Carolina University in 2001. He then proceeded to obtain his Master of Arts degree in Industrial-Organizational Psychology and Human Resource Management from Appalachian State University in 2004. His Doctor of Philosophy degree in Industrial and Organizational Psychology from The University of Tennessee was conferred in 2008. John's research has been published in several peer-reviewed academic journals and edited books. In addition, he has presented the results of his research at several academic meetings. John will pursue a career as a professor, and is currently employed as an Assistant Professor in the Industrial-Organizational Psychology program at The University of Missouri – Saint Louis.