To the Graduate Council:

I am submitting herewith a dissertation written by Brian J. Hoffman entitled "Disentangling the Meaning of Multisource Feedback: An Examination of the Nomological Network Surrounding Source and Dimension Factors." 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 Woehr
Major Professor

We have read this dissertation and
recommend its acceptance

Michael Rush

Lowell Gaertner

Lane Morris

Accepted for Council:

Anne Mayhew
Vice Chancellor and Dean of
Graduate Studies

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

DISENTANGELING THE MEANING OF MULTISOURCE FEEDBACK: AN EXAMINATION OF THE NOMOLOGICAL NETWORK SURROUNDING SOURCE AND DIMENSION FACTORS

A Dissertation Presented for the Doctor of Philosophy Degree

The University of Tennessee, Knoxville

Brian J. Hoffman

August 2006

DEDICATION

This culmination of my education is dedicated to my parents, Jim and Sharon Hoffman. Thank you for instilling in me the importance of hard work and for always standing behind me, whether I thought I needed it or not. Your constant encouragement has meant more than you know. This doctoral degree is as much your accolade as it is mine. Thank you.

ACKNOWLEDGMENTS

This research was funded by a Dissertation Grant Award provided jointly by the Society for Human Resource Management and the Academy of Management- Human Resources Division.

Many individuals have contributed to the completion of this dissertation and my academic career. First and foremost, I thank my fiancée, Myndi, who somehow managed to smilingly endure my coffee induced insomnia and matching mood as I finished this dissertation. Without her support, love, and occasional reality-checks, I am quite certain that I would have never made it this far.

I must also acknowledge my major advisor, Dave Woehr, whose professional competence is matched only by his infectious enthusiasm for our field. It is my sincerest hope that I can offer the same level of dedication to my students that Dave has shown me. I would also like to express my gratitude to Mike Rush, Lowell Gaertner, and Lane Morris for their thoughtful comments on the design, analyses, and implications of my dissertation. I would like to particularly thank Mike for working to beat "theory" into my consciousness; I am certainly a better scientist for it.

I would be remiss if I did not acknowledge Jeff Kudisch for introducing me to I/O, sharing with me his commitment to the field, and finding a place for me at UT. Without his guidance, I would not be here today. To this end, I must also thank Tom Ladd for taking a chance on an unknown prospect from Southern Mississippi. Finally, thanks to Kate Atchley for her kind words of encouragement and for taking the time to listen.

If my graduate career has been successful, I am firmly convinced that it is in large part due to the ability to work closely with people that I admire. I am grateful to Joy Oliver, Carrie Blair, and John Meriac for the hard work they laid on the Altar of the Meta-Analysis Gods, their eternal optimism, and most importantly, their friendship. I would also like to thank the senior students who, in various forms, acted as mentors to me. Specifically, thanks to Katie Helland, Cheryl Barksdale, Bill Walton, Jillian Peat, and Betsy Smith for their advice, support, and perhaps most importantly, taking me down a notch as needed. Thanks also to the Southern Miss Crew, Andy Goldblatt, Sam Hunter, Allen Watson, Vince Fortunato, and Sean Robson, for their patience and guidance during my formative years. Finally, I must give special thanks to Brian Frost, my closest confidante and a loyal friend. Among his many admirable attributes, he regularly demonstrated persistence by enduring beating upon beating at my hands while playing an assortment of PS2 titles.

ABSTRACT

This study combines internal and external approaches to construct validation in examining the construct validity of multisource feedback (MSF). First, consistent with prior MSF research, within source agreement was greater than across source agreement, the MSF instrument was equivalent across sources, and source and dimension latent factors characterized the MSF data. Next, existing MSF construct validity research was extended by examining the pattern of relationships between factor analytically derived source and dimension factors and externally measured constructs (e.g., assessment center dimensions, personality constructs, and intelligence). The pattern of relationships between MSF dimensions and conceptually similar and dissimilar external constructs suggested somewhat weak construct validity evidence for the MSF dimensions. In contrast, based on correlations with externally measured constructs, relatively strong evidence was provided for the construct validity of AC dimensions. Finally, the MSF source factors were differentially correlated with externally measured constructs, suggesting that MSF source effects represent substantively meaningful source specific variance, as opposed to error. These findings are discussed in the context of managerial skill diagnosis and the efficacy of collecting performance data from multiple sources.

TABLE OF CONTENTS

CHAPTER PAGE

I. INTRODUCTION……..1

II. REVIEW OF THE LITERATURE……7
Defining Managerial Performance.....……...7
Internal Approaches to Construct Validation....……..11
Interrater Agreement.....…….....11
Measurement Equivalence.....…….....…14
Latent Structure of MSF.....…….....…17
Potential Models Characterizing MSF.....20
A Comment on Internal Approaches to Construct Validation….....29
External Approaches to Construct Validation.....32
Dimension Effects.....….....41
Source Effects.....44
Summary….....….....…….....…....……..49

III. METHODOLOGY……..….....……..……52
Participants…….…….…….....52
Procedure.....…….…….....…..53
Measures.....…….…….…….....…54
Assessment Center Method……....…….…….....54
Personality.....…….....…….…….54
Cognitive Ability.....…….…….....55
Multisource Feedback....……..……..……..…....55
Preliminary Data Analyses and Procedures.....56
Preparation for Data Analysis.....…...…...…...……...…...56
Item Composites.....……..…….....…56
Structure of Performance Ratings.....58
Within Source Agreement.....…..……...…….…….....62
Evaluation of Model Fit...……...……..……..……....64

IV. RESULTS...…….…….....66
Research Questions 2-4: Agreement in MSF……...…..66
Research Questions 1, 5-7: Modeling MSF..……..68
The Nomological Network of MSF….....……79
Structure of AC Dimensions.....80
Research Question 8: The Construct Validity of MSF Dimensions…..…82
Research Question 9: The Meaning of MSF Source Factors…….....89
Construct Validity of AC Dimensions.....92

V. DISCUSSION…….....98

Summary of Results……..…….....98
Internal Approaches..…….....99
Rater Agreement.....99
Factor Structure of MSF.....101
External Approaches.....…….....106
Construct Validity Evidence for MSF Dimensions.....106
Construct Validity Evidence for AC Dimensions.....115
The Meaning of MSF Source Factors.....118
Limitations.....127
Implications.....130
Summary and Conclusions.....133

LIST OF REFERENCES……..……....135

APPENDICES…….....152
VITA……..…...163

LIST OF TABLES

TABLES		PAGE
Table 1:	Classification of Managerial Skills and External Constructs	59
Table 2:	Model Fit Statistics for Structural Models	70
Table 3:	Standardized Parameter Estimates for the Six Factor Model	73
Table 4:	Proportion of Variance Attributable to Dimension, Source, and Uniqueness Components from the Six Factor Model	75
Table 5:	Correlations Among Latent Factors	83
Table 6:	Correlations Among MSF Dimensions and External Constructs	86
Table 7:	Correlations Among AC Dimensions and External Constructs	95
Table 8:	Summary of Construct Validity Evidence for MSF and AC Dimensions	97

LIST OF FIGURES

FIGURES PAGE

Figure 1: One General Performance Factor…….....…….…..…...22

Figure 2: Three Dimension Latent Factors.....…….…..24

Figure 3: Three Source Latent Factors.....…….....…..…...25

Figure 4: One General Performance and Three Source Latent Factors….....…27

Figure 5: Three Source and Three Dimension Latent Factors…….….....28

Chapter 1

Introduction

Traditionally, organizations have primarily relied on employees' immediate supervisors to provide ratings of performance. In recent years however, organizations have started to evaluate work performance from multiple perspectives. To this end, multisource feedback (MSF) instruments have enjoyed increasing popularity as tools of performance evaluation (Carruthers, 2003; Church & Allen, 1997). Briefly, MSF entails obtaining ratings of job-related competencies from raters from multiple organizational levels. Although MSF is occasionally used for administrative purposes (e.g., organizational decision making), the preponderance of MSF tools are used for employee development (Church & Bracken, 1997; Timmereck & Bracken, 1997). Typically, organizational constituents' ratings are presented to the target, separated by skill dimension and the organizational level of the rater.

For MSF to be of use to managers as a developmental tool, the feedback derived from these tools must accurately reflect dimension-level performance (Borman, 1997; London & Beatty, 1993; London & Smither, 1995). For example, imagine a situation in which a MSF instrument indicates that a manager's problem-solving skills are deficient, when in reality the manager's problem solving skills are acceptable. In this situation, the manager may waste considerable time and resources working to develop this skill set.

The converse is also problematic. To the extent that inaccuracies in MSF ratings result in undiagnosed performance deficiencies, important skill deficiencies may go unnoticed. Recent reviews of the MSF literature (Seifert, Yukl, & McDonald, 2003; Smither, London, & Reilly, 2005) have indicated that the use of MSF tools has only a minimal impact on subsequent performance improvement. The authors of these reviews suggested that one particular factor that may inhibit MSF's capacity to facilitate performance improvement is questionable construct validity evidence associated with skill dimensions measured with MSF tools.

In accordance with the increasing popularity of MSF instruments, organizational scientists have sought to evaluate the psychometric properties of ratings taken from multiple sources. The vast majority of this research has focused on internal approaches to construct validity. For the purpose of this study, an internal approach to construct validity refers to an examination of the pattern of relationships between constructs assessed by the same measurement instrument. This research has relied on three primary approaches including: comparing agreement within to agreement across sources, the examination of the equivalence of MSF instruments across sources, and structural modeling to evaluate the latent structure of performance. This research has generally indicated that: 1) raters from different organizational levels agree less on the performance of a given target than do raters from the same organizational level (Conway & Huffcutt, 1997; Harris & Shaouebrock, 1988; Viswesvaran, Ones, & Schmidt, 1996), 2) MSF instruments are equivalent across sources, and 3) both dimension and source effects account for meaningful variance in MSF ratings (e.g., Lance, Teachout, & Donnelly, 1992; Lawler, 1967; Woehr, Sheehan, & Bennett, 2005).

Despite general agreement that both source and dimension effects account for meaningful variance in MSF ratings, the meaning of these respective effects is open to interpretation (Borman, 1974; Borman, 1997; Woehr et al., 2005). That is, internal approaches to construct validation do not provide conclusive evidence as to the construct validity of dimension factors or the meaning of source factors (Arthur & Villado, in preparation; Farr, 2006; Murphy & DeShon, 2000). With respect to dimension effects, it is generally assumed that they represent information regarding the skill dimension being rated. For example, if problem solving is the dimension being rated, organizational scientists generally assume that ratings reflect problem-solving skills, and as a result, present targets with feedback on their problem-solving skills. However, sparse empirical evidence exists in support of this assumption (Borman, 1997).

Similarly, very little is understood as to the nature of source effects associated with MSF instruments. A variety of speculations have been forwarded as to the meaning of MSF source effects. For example, it has been proposed that differences in ratings across sources can be attributed to different conceptualizations of effective performance by different sources (Campbell & Lee, 1998), different opportunities to observe target behavior (Borman, 1974; Murphy & Cleveland, 1995), differential dimension weighting by different sources based on the benefit of the target's behavior to a given rater (Beauovis & Dubois, 2003), and the display of different behaviors in the presence of different groups of raters (Lance & Woehr, 1989). Finally, it is possible that source effects simply represent variance that cannot be attributed to a target's performance (e.g., error).

In the past, multitrait-multimethod (MTMM) research has generally assumed measurement method effects (e.g., source effects) represent "contamination introduced by the method used to measure a trait" (Lance, et al., 1992; pp. 448) and should be reduced as much as possible (Conway, 1996). However, researchers have recently suggested that method effects associated with performance ratings provided by raters from different organizational levels represents meaningful variance (Lance et al., 1992; Lance, Woehr, & Fisicaro, 1990). Existing research has not yet demonstrated the extent to which MSF source effects represent source specific performance relevant variance (SSPRV) or simply measurement error.

Although previous research incorporating internal approaches to construct validation provides useful information as to the quality of ratings made by multiple organizational constituents, these approaches leave important issues unresolved. Some suggest that this stream of research does not allow for a determination of the meaning of MSF source effects (Farr, 2006). For example, Viswesvaran, Schmidt, and Ones (2002) relied solely on the internal approaches to construct validation to provide evidence that source factors are meaningless. In contrast, some have argued that the results of studies incorporating these internal approaches should not be relied on when evaluating the validity of MSF instruments (Arthur & Villado, in preparation; Borman, 1997; Farr, 2006; Murphy & DeShon, 2000). In other words, although the internal approaches regularly demonstrate the presence of source and dimension effects, they are ill equipped to provide evidence regarding what this variance represents (e.g., the validity of these effects). Consequently, these authors argue that research relying solely on the internal

approaches results in erroneous conclusions with respect to the reliability and validity of MSF.

One potential method of determining the meaning of MSF source and dimension effects is through an external construct validation approach (Arthur & Villado, in preparation; Borman, 1997; Farr, 2006). An external construct validity approach refers to the examination of the pattern of relationships between a construct measured using one methodology and conceptually similar and dissimilar constructs assessed using different measurement methodologies. In the context of MSF, this approach would provide an indication of the degree to which MSF dimensions accurately assess the underlying performance construct of interest and MSF source effects represent SSPRV or measurement error. In fact, MSF researchers have called for an examination of the nomological network surrounding the source and dimension factors characteristic of multisource performance ratings (Borman, 1997; Conway, 2000; Farr, 2006; Woehr et al., 2005). To exemplify, Woehr and his colleagues (2005) noted that, "To the extent that these effects reflect systematic, performance-related variance, they should correlate with other criterion measures and to the extent that they represent random effects, they should not" (Woehr et al., 2005; pp. 598).

Although some research has investigated the relationship between ratings made from different sources and external constructs (e.g., Conway, 2000), this research has not explicitly modeled MSF source and dimension effects. Instead, this stream of research has typically compared the relationship between overall ratings made from one source and external constructs to the relationship between overall ratings made from a different source and external constructs (e.g., Conway, Lombardo, & Sanders, 2001). Thus,

research investigating the nomological network of MSF has confounded dimension and source effects. Hence, firm conclusions cannot be drawn regarding the construct validity of MSF dimension factors and the meaning of source factors.

Given the increasing prominence of MSF in organizations as a developmental tool, a clear understanding of *what* is actually being measured by MSF tools is crucial to progress in this stream of research. Although previous research has investigated the construct validity of MSF ratings using internal and external construct validity approaches, these approaches have not yet been examined together in a single study. Because of these limitations, existing empirical research does not adequately address the validity of MSF instruments. By first modeling both source and dimension factors and then by examining the relationship between these factors and the nomological network of performance-relevant individual differences (e.g., intelligence, personality constructs, and assessment center dimensions), this study represents the first attempt to disentangle the meaning of MSF source and dimension effects. The results of this study will give an indication of 1) the construct validity of MSF dimension factors, and 2) the extent to which MSF source factors represent SSPRV or error.

Chapter 2

Review of the Literature

Defining Managerial Performance

Before undertaking a discussion of the impact that different rating sources have on performance measurement, a discussion of relevant job performance models is necessary. That is, following Murphy's (1989) recommendations, I will begin by explicating the performance constructs of interest in the present study. The present approach may be contrasted with inductive approaches to performance measurement, wherein the nature of a given set of performance ratings is ascertained on the basis of statistical analyses such as factor analysis. According to Murphy, the a priori specification of a conceptual performance model is superior to inductive approaches of performance modeling in that the deductive approach does not place arbitrary statistical restrictions on the support of a performance model and consequently, avoids supporting a performance model on the basis of chance. Finally, this approach is more conducive to a priori hypothesis specification with respect to the relationship between specific performance dimensions and external constructs.

Clearly, before discussing the dimensionality of work performance, it is first necessary to delineate the relevant domain of performance. The present study is concerned with managerial performance. Specifically, the sample for this study focuses on managers enrolled in a leadership development program. These managers represent a wide variety of professions and are employed by a wide variety of organizations. In this

context, narrow technical skills specific to a given profession are less relevant than are more broad managerial competencies. For example, a leadership development program would be ill suited to train an accountant who lacks the necessary technical skills to adequately perform the tasks associated with his/her job. Hence, the performance model in this context must focus on a broad range of managerial competencies that generalize across organizations and job types. The discussion of managerial performance models will focus on sets of skills relevant to managers across organizational and job settings.

A host of taxonomies of managerial skills have proliferated the organizational literature over the years (cf., Bales, 1970; Borman & Brush, 1993; Fleishman, 1953; Fleishman, Mumford, Zaccaro, Levin, Korotkin, & Hein, 1991; Katz, 1974; Mann, 1965; Mintzberg, 1975; Tonrow & Pinto, 1976; Yukl, 1989). All of these frameworks have two dimensions in common. In particular, each framework specifies some form of interpersonal skills and some form of conceptual skills as essential components of managerial performance. Briefly, conceptual skills are broadly defined as skills that facilitate the accomplishment of job-relevant tasks. In the context of managerial jobs, conceptual skills involve administrative functions such as planning and decision-making. In contrast, interpersonal skills include behaviors directed at building and maintaining effective interpersonal relationships. The distinction between conceptual and interpersonal skills as overarching dimensions of performance is not restricted to managerial jobs. In fact, many performance taxonomies include some form of conceptual and interpersonal skills as essential components of job performance. For example, Borman and Motowidlo's (1993) widely researched model of task and contextual

performance may be mapped quite easily onto a performance model specifying conceptual and interpersonal skills (Conway & Huffcutt, 1997).

Although the majority of performance models specify conceptual and interpersonal skills as overarching factors of performance, another factor of performance is particularly relevant to effectiveness in managerial jobs. Specifically, leadership is a key component of managerial effectiveness. To this end, many taxonomies specific to managerial jobs specify a performance model composed of conceptual/administrative, interpersonal, and leadership skills. In contrast with nonmanagerial jobs, managers must accomplish organizational goals through the work of others. Essentially, neither conceptual/administrative nor interpersonal dimensions account for the essential component of managerial jobs of ensuring that "subordinates know what to do and are motivated to do it" (Conway, 1999, pp. 335). Thus, conceptual/administrative and interpersonal skills must be supplemented with leadership skills to adequately capture the construct domain of managerial performance (Borman & Brush, 1993; Conway, 1999).

The Michigan Leadership Studies were a series of studies designed to determine the content of managerial work that categorized managerial behavior based on a large sample of responses to interviews and questionnaires (Yukl, 2005). The results of this set of studies identified three primary managerial functions including: task-oriented behavior, relations-oriented behavior, and participative leadership (Katz, Maccoby, & Morse, 1950; Katz & Khan, 1952). Similarly, Bales (1970) proposed that managerial performance could be classified into three categories including: task, interpersonal, and leadership performance. More recently, Borman and Brush (1993) proposed a model of managerial performance consisting of four "mega-dimensions" including: interpersonal

relations/communication skills, leadership/supervision, administrative skills/mechanics of management, and a useful behaviors classification that essentially represents an "other" category. Similarly, Conway (1999) provided meta-analytic support for a model of managerial performance consisting of task performance, interpersonal facilitation, and leadership.

Consistent with previous research on the nature and structure of managerial performance (Bales, 1958; Borman & Brush, 1993; Conway, 1999; Katz, Maccoby, & Morse, 1950; Katz & Khan, 1952), the present study will conceptualize managerial performance as composed of three dimensions including: conceptual/administrative, interpersonal, and leadership skills. In this context, conceptual/administrative skills are relevant to the performance of the specific tasks of a manger's job such as planning, problem solving, and decision-making. Interpersonal skills are defined as skills that are relevant to the building and maintenance of interpersonal relationships, showing personal concern for others, and effectively communicating with others. Finally, leadership skills include behaviors directed toward influencing the work performance of others by providing guidance and direction, motivating by providing them with recognition, encouragement, and facilitating the work of others to help meet organizational goals. Importantly, although both leadership and interpersonal skills are primarily interpersonal in nature, they differ in that leadership skills are more involved with the achievement of goals than are interpersonal skills, which are somewhat less goal-oriented (Borman & Brush, 1993).

Research Question1: Will managerial performance be best represented by a three-factor model including conceptual/administrative, interpersonal, and leadership skills?

Internal Approaches to Construct Validation

The preponderance of MSF construct validity research has focused on internal approaches to construct validity. In essence, this approach entails examining the relationship between ratings provided by raters from different organizational levels on the same rating instrument. Internal construct validation approaches have persisted on two fronts in the performance literature. First, this research has examined the relative degree of correspondence in ratings completed by raters at the same organizational level and raters at different organizational levels (e.g., Conway & Huffcutt, 1997). Second, research has incorporated structural modeling approaches to assess the equivalence of ratings taken from different sources (e.g., Woehr et al., 2005) and the latent factor structure of performance ratings (e.g., Lance et al., 1992). Research relevant to each of these approaches will be discussed in turn.

Interrater Agreement

A wealth of MSF research has focused on the level of agreement within and across rating sources to ascertain the psychometric properties of ratings taken from multiple sources (Borman, 1997; Murphy, Cleveland, & Mohler, 2001). This stream of research consistently indicates that raters from the same organizational level display greater levels of agreement than raters from different organizational levels (Conway & Huffcutt, 1997; Harris & Schaubroeck, 1988; Viswesvaran, Ones, & Schmidt, 1996). Research indicating that individuals at the same organizational level agree more strongly

than raters from different organizational levels suggests systematic differences in ratings depending on the organizational level of the rater. In essence, this research provides initial evidence for the importance of considering rater organizational level when interpreting performance rating data.

A meta-analysis by Conway and Huffcutt (1997) indicated that the average relationship between ratings provided by raters from different organizational levels is typically somewhat weak (average r = .22). Other meta-analyses investigating the relationship between ratings provided by multiple sources have evidenced relationships commensurate with those reported by Conway and Huffcutt (Harris & Shaubroeck, 1988; Viswesvaran, Ones, & Schmidt, 1996). Although at first glance, the disagreement across sources is somewhat troubling, MSF researchers have interpreted poor agreement across sources as evidence for the usefulness of MSF (Borman, 1974; Lance, Woehr, & Fisicaro, 1990). For example, if different sources completely agree as to the nature of a target's performance, collecting performance ratings from different sources would be unnecessary. In fact, the utility of multisource ratings is predicated on different sources providing somewhat different perspectives on a target's performance (Borman, 1997).

Research Question 2: What will be the level of correspondence between ratings provided by raters from different organizational levels?

In contrast to the weak level of agreement between raters from different organizational levels, research has demonstrated moderate levels of agreement between raters from the same organizational level (Conway & Huffcutt, 1997; Harris & Schaubroeck, 1988; Viswesvaran, Ones, & Schmidt, 1996). Together, research consistently indicates that the agreement of ratings within a single source (e.g., peers) is

greater than the agreement of ratings provided by distinct sources (e.g., supervisors and peers). For example, the average within source correlation reported by Conway and Huffcutt (.40) was greater than the average relationship of ratings across sources (.22). To the extent that within source agreement indices are greater than across source agreement indices, preliminary support is provided for the importance of considering the rater's organizational perspective when evaluating MSF ratings. Specifically, research indicating that individuals at the same organizational level agree more strongly than raters from different organizational levels suggests systematic differences in ratings depending on the organizational level of the rater. The findings of previous interrater agreement research lead to the following research question:

Research Question 3: Will ratings made by the same source be more strongly related than ratings made by different sources?

Moreover, the level of agreement across sources may differ depending on the performance dimension being rated. To illustrate, it is possible that supervisor and peer raters will display the lowest levels of agreement on a target's leadership and interpersonal skills, because supervisors may be in less of a position to observe these behaviors. For example, a supervisor rarely has the opportunity to see a manager work with his/her direct reports. Or, a person with poor interpersonal skills may not display these skills when his/her supervisor is present (Lance & Woehr, 1989). Similarly, subordinate ratings of conceptual skills may be different from peer or supervisor ratings because direct reports are often in a position that they need to defer to supervisors on technical issues or important decisions. Thus, a manager's subordinates may be in less of a position to evaluate the quality of their managers' conceptual skills. In contrast, peers

and supervisors may be a more accurate barometer of a manager's conceptual skills because to be at a similar organizational level or higher, the same set of skills is often necessary. Taken together, there is reason to believe that the level of agreement across sources will differ depending on the underlying dimension of performance being rated.

Despite the intuitive appeal of this proposition, a recent meta-analysis by Viswesvaran, Schmidt, and Ones (2002) suggested that existing research does not support the proposition that the relationship between supervisor and peer ratings is moderated by the content of the dimension being rated. In contrast, the meta-analysis by Conway and Huffcutt (1997) indicated that across source agreement was higher for interpersonal skills than cognitive skills. Given the conflicting evidence provided by previous research, the extent to which performance dimension being rated moderates across source agreement is unresolved.

Research Question 4: Will the relationship between different sources' ratings differ depending on the performance dimension being rated?

Measurement Equivalence

Other researchers have incorporated a structural modeling approach to evaluating the psychometric properties of MSF scales (e.g., Cheung, 1999; Conway, 1996; Coovert, Craiger, & Teachout, 1997; Diefendorff, Silverman, & Greguras, 2005; Facteau & Craig, 2001; Lance et al., 1992; Lance & Bennett, 1997; Lawler, 1967; Maurer, Raju, & Collins, 1998; Woehr, Sheehan, & Bennett, 2005; Zedeck & Baker, 1972). Structural modeling has been used in MSF investigations in two ways including: 1) the assessment of measurement equivalence, and 2) the examination of the latent

structure of performance ratings using multitrait-multimethod matrices (MTMM; Campbell & Fiske, 1959). Each of these approaches will be discussed in turn.

Measurement equivalence researchers have typically specified separate models of ratings from different sources and examined whether rating source moderates the pattern or magnitude of loadings on latent performance factors (Cheung, 1999). In essence, this approach seeks to determine the degree to which the dimensions of performance load on the same underlying performance factor, and the dimensions rated across sources have equivalent loadings on latent factors. Researchers use this type of study to determine the extent to which ratings on a given MSF instrument are comparable across sources. Stated differently, "...if the underlying characteristics being measured in these rating systems are not on the same psychological measurement scale, then observed differences across [rating sources] are possibly artifactual, contaminated, or misleading" (Maurer, Raju, & Collins, 1998; pp. 700). Together, measurement equivalence studies attempt to indicate the degree to which different rater groups conceptualize the dimensions of job performance similarly.

With one notable exception (Lance & Bennett, 1997), this stream of research has consistently indicated that performance ratings are equivalent across sources (Diefendorff, Silverman, & Greguras, 2005; Facteau & Craig, 2001; Maurer et al., 1998; Scullen, Mount, & Judge, 2003). Despite the contributions of traditional MSF measurement equivalence research, the typical methodologies used to assess equivalence leave important questions as to the construct validity of multisource ratings unresolved. That is, by specifying separate performance models for each source and examining

source as a moderator of factor structure and factor loadings, this research confounds the impact of source and dimension effects.

A more appropriate method of modeling MSF ratings is to specify models including both dimension and source effects. To this end, Woehr and his colleagues (2005) incorporated just such an approach. Specifically, these authors examined the equivalence of ratings across sources by modeling both dimension and source factors derived from a MTMM matrix. Consistent with prior equivalence research, the results of their study indicated that the underlying performance dimension being rated was equivalent across rating sources, whereas the impact of rating source differed across sources. In sum, the state of the art of the MSF measurement equivalence literature consists of modeling MSF using both source and dimension factors. Although the study by Woehr and his colleagues (2005) represents the only investigation of the measurement equivalence of MSF which has specified source and dimension factors in a single model, this approach is the superior method of examining the measurement equivalence of MSF because it specifies a theoretically meaningful model. However, Woehr and his colleagues drew their sample from military mechanics and engineers, as opposed to managers. Importantly, MSF is typically used to evaluate managers, and the study by Woehr et al. represents the only attempt to examine equivalence using this approach. Consequently, the present study will investigate the measurement equivalence of a MSF instrument in a managerial sample by specifying a model consisting of both source and dimension factors in an effort to replicate the findings of Woehr and his colleagues (2005).

Research Question 5: Will ratings from different sources be equivalent?

Latent Structure of MSF

Although the study by Woehr et al. (2005) represents the only attempt to explore the equivalence of MSF ratings using a model including both source and dimension latent factors, a wealth of research has explored the construct validity of performance ratings provided by multiple organizational sources from a MTMM perspective. The original MTMM framework was designed to examine the construct validity of measures by examining the degree to which conceptually similar traits measured by different methods were related (convergent validity) and conceptually dissimilar traits were unrelated (discriminant validity; Campbell & Fiske, 1959). More specifically, the MTMM methodology proposes that the relationship between the same trait measured using different methods (monotrait-heteromethod relationship; MTHM) should be greater than the relationship between different traits assessed with the same method (heterotrait-monomethod relationship; HTMM) and the relationship between different traits using different methods (heterotrait-heteromethod; HTHM). Additionally, the method effect (HTMM) has previously been conceptualized as representing a major source of systematic error (Cote & Buckley, 1987; Fiske, 1982). Previous MTMM research investigating a variety of different methods and traits has frequently indicated that the relationship between the same trait measured by different methods is weaker than the relationship between different traits measured by the same method (Cote & Buckley, 1987).

Advances in statistical techniques over they years has resulted in alternate methods of analyzing MTMM data. Of the modern approaches to examining MTMM data, confirmatory factor analysis has received the most widespread use (Goffin &

Jackson, 1992). Confirmatory factor analytic methodologies typically entail specifying both trait and method latent factors and examining the relative factor loadings of dimension and method effects (Conway, 1996). This approach allows for the estimation of the relative proportion of variance that traits and methods explain in performance ratings.

In the context of MSF ratings, the dimension being rated serves as the trait and the source providing the rating serves as the method. In essence, MSF ratings represent a special case of the MTMM methodology that can be referred to as a multitrait-multisource matrix (MTMS). These approaches begin by generating a matrix containing the correlation (or covariance) among each dimension measured by each rating source. So, for example, if the three dimensions evaluated in the present study (conceptual/administrative, interpersonal, and leadership skills) were rated by three sources (peers, supervisors, and subordinates), the resulting matrix would contain nine lower diagonal correlations, one for each rating source's rating on each dimension.

Lawler's (1967) investigation of the relative impact of performance dimension and rating source on performance ratings represented one of the first attempts to examine multisource performance ratings using a MTMS methodology. Since Lawler's initial evaluation of multisource ratings using a MTMS methodology, substantial research has investigated the relative impact of source and dimension factors on ratings provided by raters from different organizational levels. Although this research has incorporated a variety of methodologies (e.g., average HTMM and MTHM correlations and CFA), the results consistently indicate that performance ratings made by raters from different organizational levels are characterized by both source and dimension effects (Campbell,

McHenry, & Wise, 1990; Coovert, Craiger, & Teachout, 1997; Holzbach, 1978; King, Hunter, & Schmidt, 1980; Klimoski & London, 1974; Lance, Teachout, & Donnelly, 1992; Mount, Judge, Scullen, Sytsma, and Hezlett, 1998; Scullen, Mount, & Goff, 2000; Vance, MacCallum, Coovert, and Hedge, 1988; Woehr et al., 2005; Zedeck & Baker, 1972). Moreover, previous MTMS research has been conducted in a variety of samples, with a variety of rating sources, and multiple different performance dimensions.

Conway (1996) provided a review of twenty studies that have used a MTMS methodology to investigate performance ratings. The results of this study suggested that across studies, source effects accounted for an average of 25% of the variance in performance ratings, while dimension effects accounted for an average of 16% of the variance in performance ratings. Finally, uniqueness (idiosyncratic error) accounted for the majority of variance in performance ratings (29%). The findings of Conway's review are consistent with other research investigating the relative impact of dimensions, sources, and uniqueness on multisource performance ratings (e.g., Woehr et al., 2005).

Together, existing research exploring the structure of MSF ratings indicates that both source and dimension effects account for meaningful portions of variance in the performance ratings provided by raters from different organizational levels. This research also supports the practice of examining the factor structure of performance ratings by subjecting MTMS matrices to confirmatory factor analysis (Schmitt & Stults, 1986; Lance, Noble, & Scullen, 2002). Thus, an initial purpose of this study is to examine the structure of multisource ratings using a CFA of MTMS data. The results of these analyses will also be used to estimate the relative proportion of variance performance dimension, rating source, and uniqueness explain in MSF ratings.

Potential Models Characterizing MSF

Although previous research typically indicates that source and dimension factors characterize multisource ratings, researchers have argued that when evaluating structural models, multiple competing theoretical models should be specified and tested. This approach allows for an indication of the most appropriate model among multiple theoretically plausible models and avoids the acceptance of a model when more appropriate models characterize the data (Lance et al., 1992; Mulaik, James, VanAlstine, Bennett, Lind, & Stilwell, 1989; Widman, 1985). Accordingly, I will outline five latent factor models that potentially characterize performance ratings taken from multiple organizational sources.

First, a model where all dimension ratings from all sources load on a single latent performance factor could characterize MSF data (Model 1). Support for this model would indicate that raters from different organizational levels view performance in similar terms (e.g., with a general performance evaluation), irrespective of the dimension being rated. In other words, neither source not dimension effects are specified in this model. Multiple studies have indicated that job performance can be characterized with a general factor (Cooper, 1981; Hulin, 1982; Viswesvaran, 1993). However, this research also indicates that although the domains of job performance are interrelated, they share an average of approximately 50% of the variance (Viswesvaran, 1993). In other words, the dimensions of performance are strongly related, yet distinct constructs. To this end, a meta-analysis by Hoffman, Blair, Meriac, and Woehr (in press) demonstrated that task performance and organizational citizenship behaviors are empirically distinct, albeit strongly intercorrelated performance domains. Together, the results of previous research suggest

that work performance is best represented as a multidimensional construct. Given that previous research consistently indicates differences in ratings across sources and provides support for the multi-dimensionality of job performance, a model consisting of a single performance factor is unlikely to provide the best representation of job performance data. A graphical representation of this model is presented in Figure 1.

A model composed of underlying performance dimensions with different sources' ratings serving as indicators of conceptually similar performance dimensions is also a possibility (Model 2). In the context of the present study, this model specifies that each of the three sources' ratings of the three performance dimensions load on the same underlying factors representing conceptual/administrative, interpersonal, and leadership performance. That is, all three sources' leadership performance ratings would load on a single latent factor, all three sources' conceptual/administrative performance ratings would load on a single latent factor, and all three sources' interpersonal performance ratings would load on a single latent factor. Importantly, this model proposes a similar result as do traditional external construct validity approaches. Specifically, independent ratings of the same construct (conceptual/administrative, interpersonal, or leadership performance) should load on the same latent factor (convergent validity) and ratings of different constructs should not load on the same factor (discriminant validity).

By not modeling a latent source factor, this type of model suggests that measurement source is irrelevant and that ratings made by each source are indicators of the same underlying performance dimensions. However, given previous research

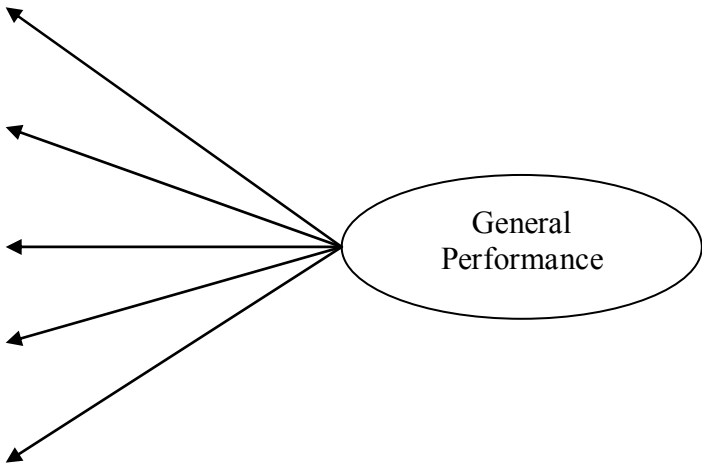

Figure 1. One General Performance Factor

demonstrating the presence of source effects associated with performance ratings, the impact of measurement source is an important omission. Still, a multi-dimensional model will probably be better supported by MSF data than will a general performance model (Model 1) because of the explication of theoretically derived performance dimensions. However, the consistent support for source effects in previous research suggests that a model failing to specify latent source factors will not provide the best representation of MSF ratings. This model is presented graphically in Figure 2.

The third model suggests that the structure of ratings from multiple sources is represented by a factor corresponding to each source's ratings (Model 3). For example, this model specifies peer ratings on each of the three dimensions of interest combine to form a single latent factor, as do ratings provided by supervisor and subordinate ratings. Essentially, this model suggests that ratings of the same dimension across sources will not share meaningful variance and that all dimensions rated by a single source will form a single latent factor. This model is consistent with perspective that each rating source has a general impression of a target's performance. Despite consistent findings demonstrating that source effects explain variance in performance ratings, research also indicates that dimension effects explain meaningful variance in performance ratings. Thus, it is unlikely that a structure consisting solely of source effects will provide the best fit to MSF data. Figure 3 provides an illustration of this model.

Model 4 combines Model 1 and Model 3 to posit a MSF model consisting of three source factors and one overall performance factor. In essence, this model posits a general performance factor that impacts performance ratings across sources and three factors that

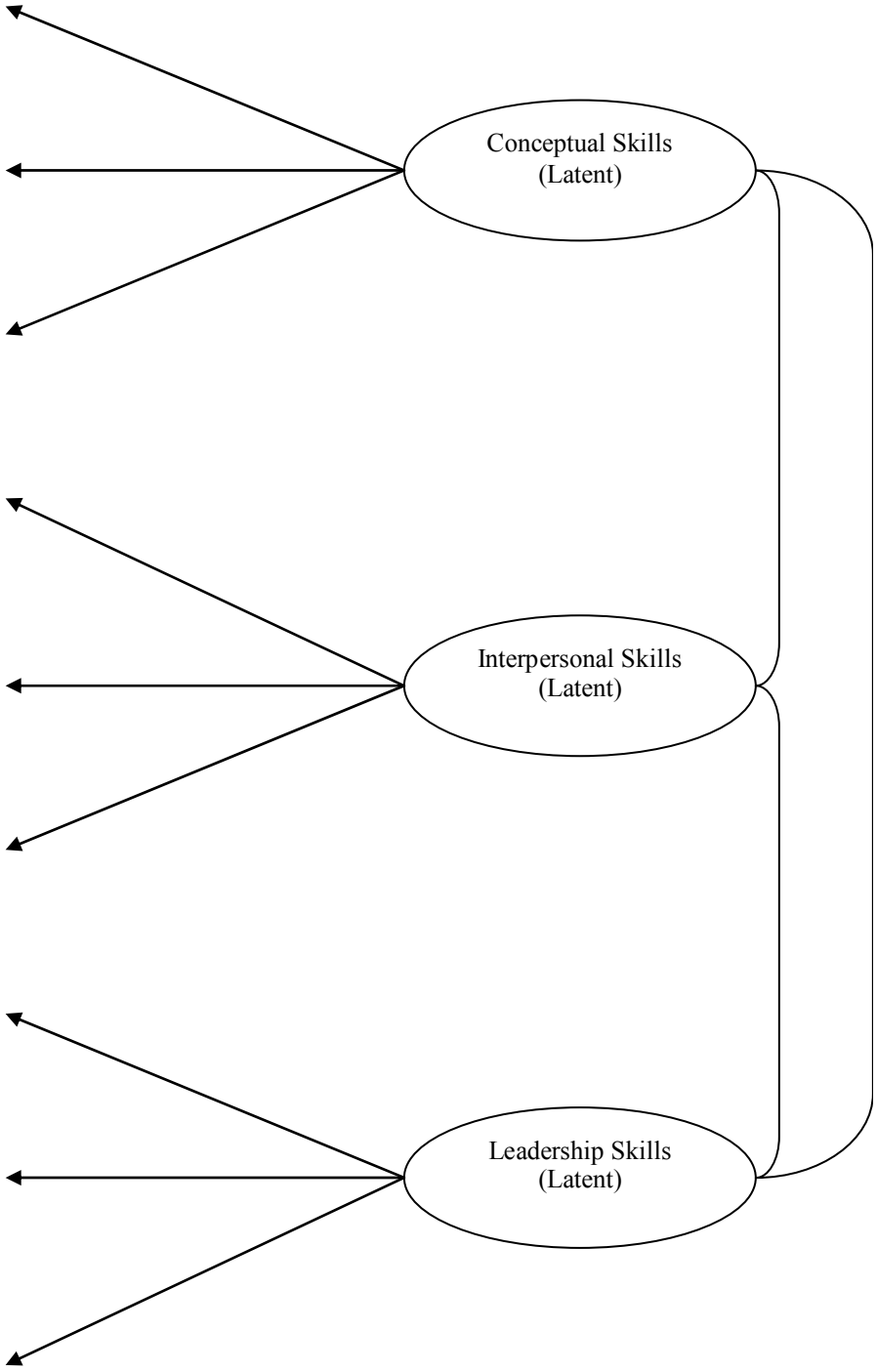

Figure 2. Three Dimension Latent Factors

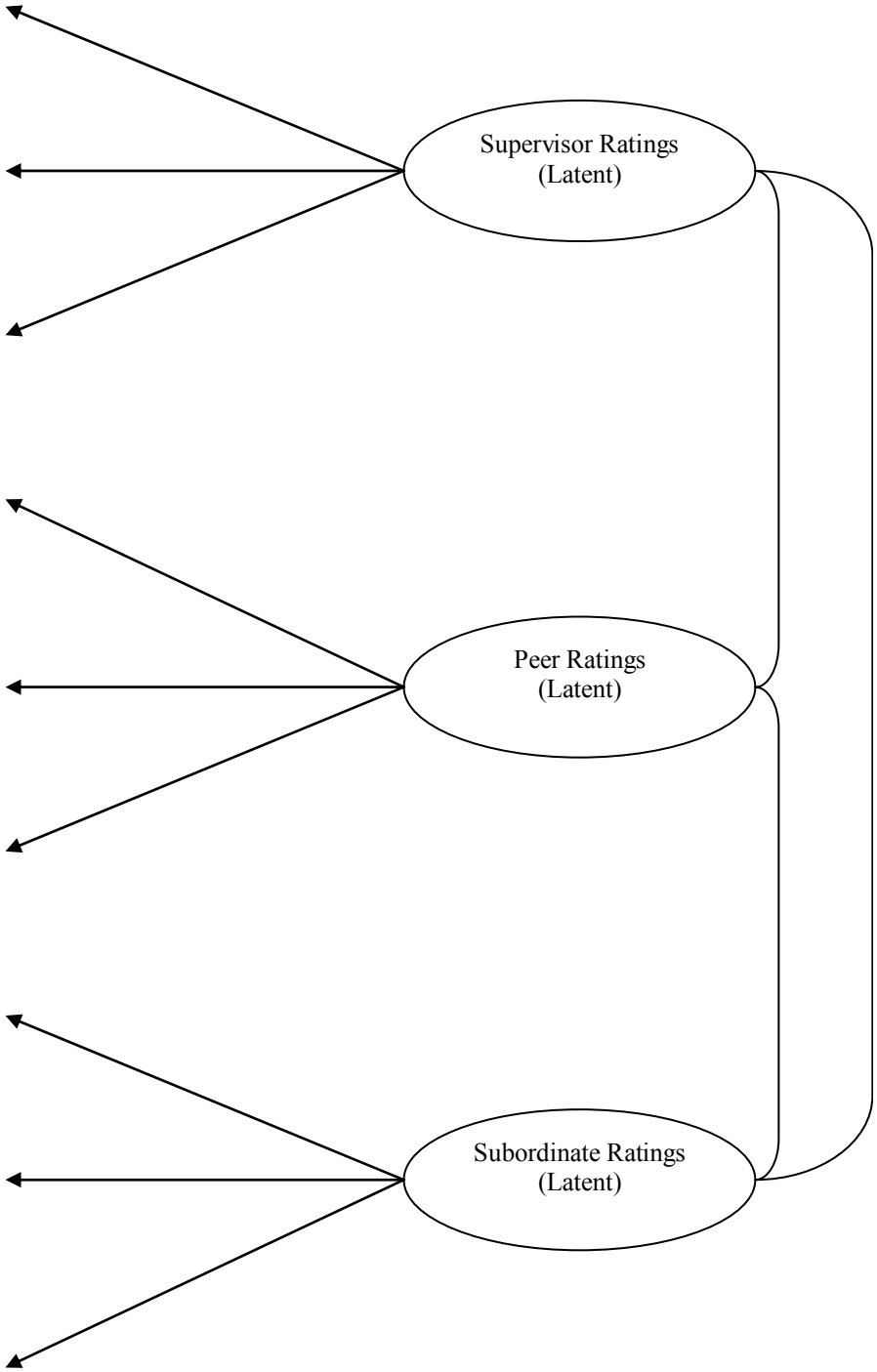

Figure 3. Three Source Latent Factors

represent performance specific to each source (Lance et al., 2000).The three source factors take on the same meaning as in Model 3. The general performance factor is akin to general mental ability in that it is an empirically derived general factor than influences performance across many different dimensions (Lance et al., 2000). Although Lance and his colleagues have supported this model when examining the factor structure of assessment center (AC) ratings, research has not yet applied this model to MSF. This model is graphically depicted in Figure 4.

Model 5 specifies six factors in the present context, where ratings from different source of the same dimension load on a latent factor and ratings on all dimensions for each of the three sources load on three separate latent factors. In other words, latent factors representing conceptual/administrative, interpersonal, and leadership skills and latent factors representing each of the three sources are modeled to form this structure. As previously discussed, substantial MSF research has been conducted supporting a model consisting of source and dimension effects. Thus, I expect this model to provide the best representation of the data in the present study. This model is graphically depicted in Figure 5.

Research Question 6: Does a model consisting of both source and dimension factors best characterize MSF ratings?

Research Question 7: What will be the relative proportion of variance explained in performance ratings by dimension factors, source factors, and uniqueness?

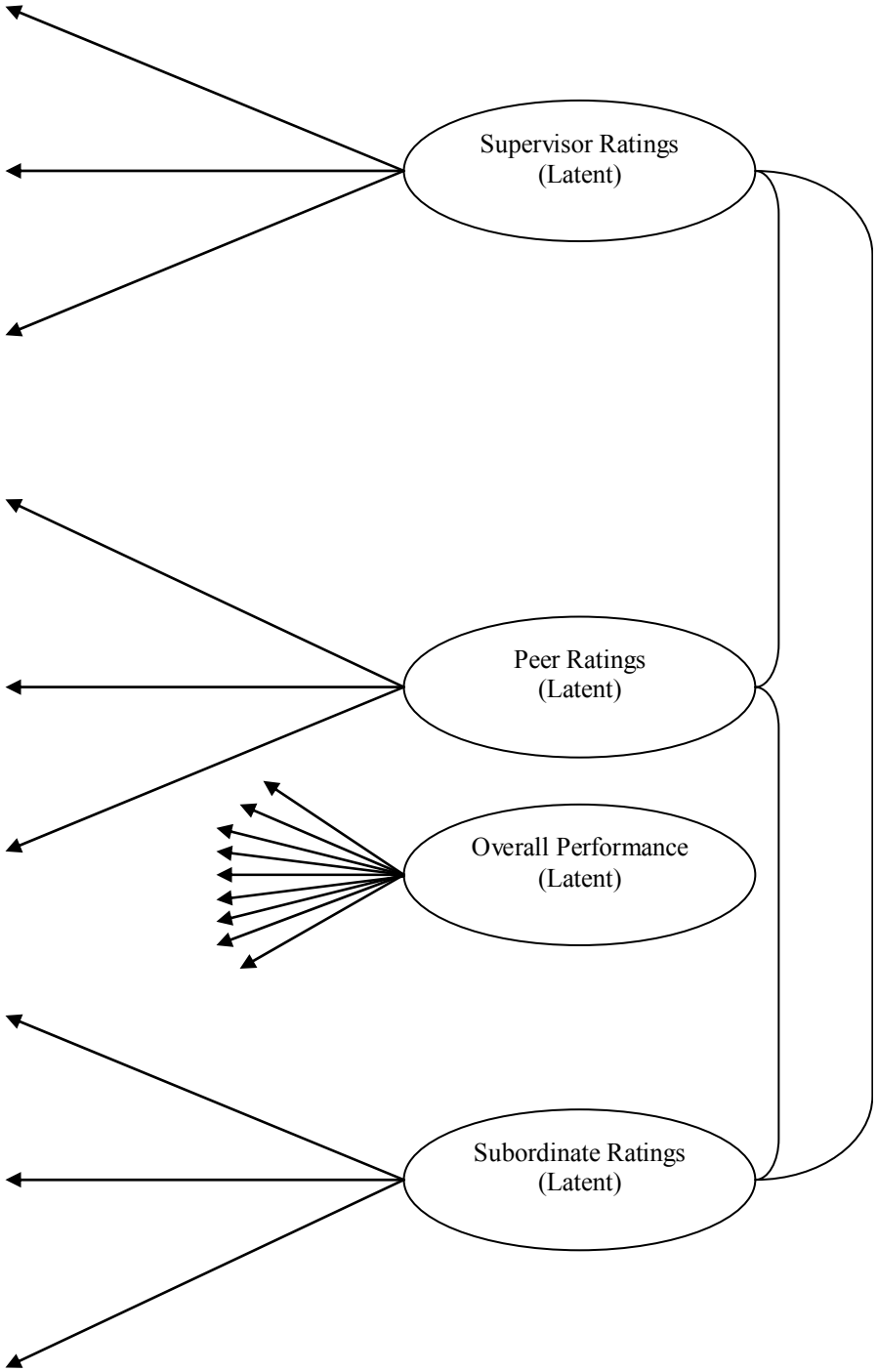

Figure 4. One General Performance and Three Source Latent Factors

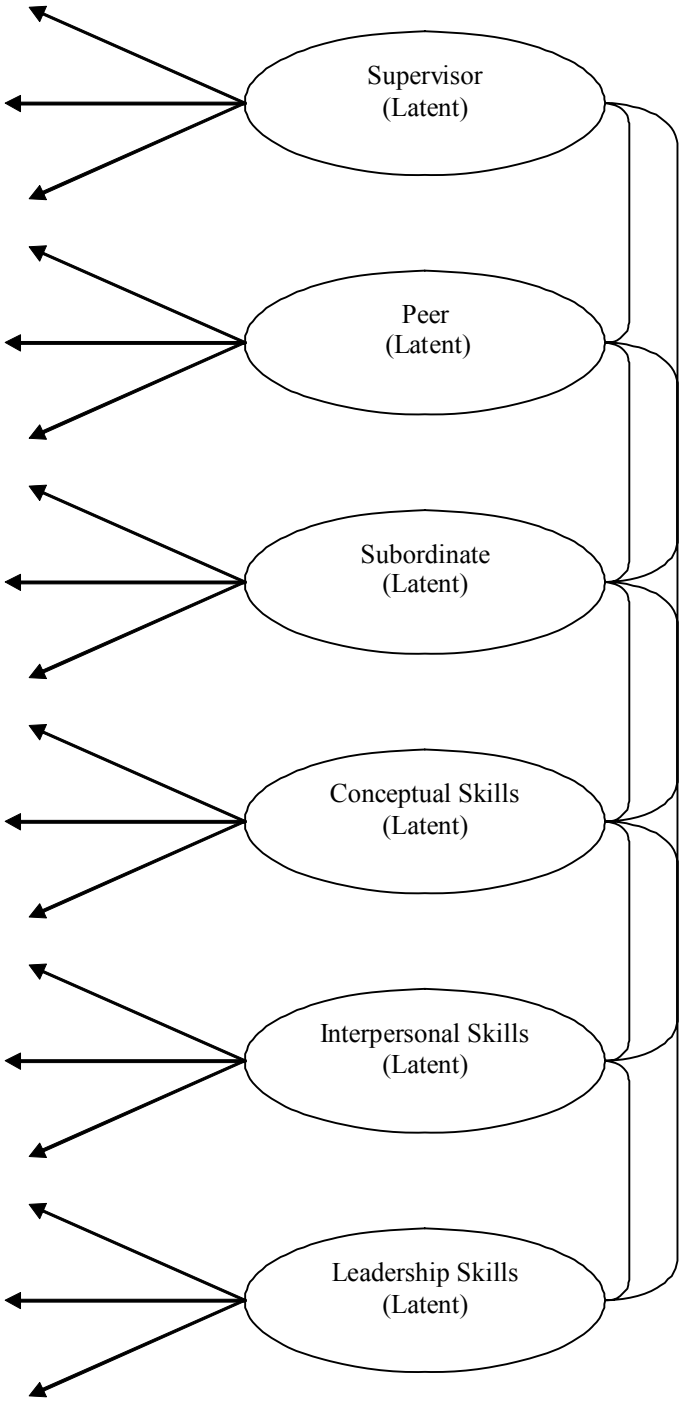

Figure 5. Three Source and Three Dimension Latent Factors

A Comment on Internal Approaches to Construct Validation

One final point is in order regarding the internal approaches to construct validation. In the preceding section, research incorporating the internal approaches was presented in the framework of the construct validity evidence of MSF. However, this may not be the case. Although methodological approaches such as interrater agreement, measurement equivalence, and MTMS have often been purported to yield construct validity evidence of MSF, such internal approaches may be more appropriately viewed as yielding evidence pertaining to the reliability of MSF (Arthur & Villado, in preparation). To illustrate, although agreement studies have been used to demonstrate that MSF operates in accordance with its philosophical purpose of providing unique perspectives on a target's performance, this type of methodology does not provide direct evidence for what is actually being measured by MSF instruments (Borman, 1997). Additionally, MTMS studies may be viewed as a special case of interrater reliability. More concretely, typical MTMS analyses examine the correspondence among dimensions rated by raters from different organizational levels. This process is the very essence of assessing interrater reliability. In fact, the resulting relationship is viewed by some as a case of parallel forms reliability, where each rater serves as a parallel form (Schmidt, et al., 2000). Thus, although MTMS and agreement studies provide evidence regarding the consistency of MSF instruments, this type of research design offers little concrete evidence with respect to *what* is actually being measured (Arthur & Villado, in preparation; Farr, 2006; Woehr et al., 2005).

Similarly, measurement equivalence research seeks to examine the consistency (e.g., reliability) of a given scale across rating sources. In other words, although

equivalence research examines whether a scale operates similarly across populations, it gives little indication as to what the scale actually measures. Together, although internal approaches provide important information with respect to the consistency of MSF instruments, this stream of research gives little indication of what MSF instruments actually measure. Stated differently, research incorporating internal approaches does not adequately gauge the validity of MSF instruments.

For example, imagine a situation where a MSF dimension correlated perfectly within and across sources and was equivalent across sources. Further, imagine that this MSF dimension was also uncorrelated with other conceptually similar constructs measured using different methodologies. In this example, the hypothetical MSF dimension is perfectly reliable according to the results of the internal approaches but is not a valid indicator of the construct of interest. This extreme example was given to demonstrate that a strict reliance on internal approaches does not shed light on the validity of a given MSF dimension.

Importantly, others have argued that interpreting the results of internal approaches as reliability evidence of MSF results in an erroneous conclusion that MSF instruments are not very reliable (Murphy & DeShon, 2000). These authors suggest that performance relevant variance may be collapsed into error when using the internal approaches. For example, disagreement across sources would be considered error with agreement studies typical of the internal approaches. However, there is reason to believe that this variance is substantively meaningful (e.g., not error) and in part, represents unique information on a target's performance (Borman, 1974; Borman, 1997). Consequently, collapsing all of the variance not shared across two sources' ratings into error may result in erroneous

estimates of MSF's true and error variance and subsequently underestimate reliability (Murphy & DeShon, 2000).

A primary tenet of psychometrics is that a test's reliability sets the upper limit on that test's validity, such that a test can be no more valid than it is reliable. If the results from the internal approaches were taken as an accurate reflection of MSF's reliability, the cap for observed validity would be quite low indeed. But, if agreement and MTMS studies are underestimating the reliability of MSF due to the specification of performance relevant variance into the error term, the validity of MSF may actually exceed that of observed reliabilities (Murphy & DeShon, 2000). In sum, some argue that studies using internal approaches are best viewed as efforts to gather evidence for the reliability of performance ratings (Schmidt et al., 2000), while others argue that internal approaches result in erroneous conclusions with respect to MSF's reliability due to variance misspecification (Murphy & DeShon, 2000). Existing empirical research does not speak to the extent to which variance unique to different sources represents error versus true score variance. Thus, previous research stands mute with respect to the debate over whether across source agreement indices erroneously partition performance relevant variance into the error term.

A more appropriate method of investigating the construct validity of MSF is to explore the relationship between source and dimension effects and constructs assessed using external measures (Arthur & Villado, in preparation; Borman, 1997; Conway, 2000; Farr, 2006; Woehr et al., 2005). In fact, Campbell and Fiske's (1959) original conceptualization of the MTMM methodology specified that to interpret the results of MTMM analyses in the context of construct validity, the constructs of interest must be

measured using maximally dissimilar methods. None of the aforementioned research incorporating the internal approach meets this requirement. Consequently, previous research sheds little light on the construct-related validity of MSF dimensions and the meaning of MSF source effects.

A method where the relationship between MSF and constructs measured using disparate measurement methodologies is ideally suited to facilitate inferences as to whether source specific variance reflects performance relevant variance or error and whether dimension specific variance accurately captures the proposed performance dimension. So, the question remains: what do dimension and source communalities represent? In order to answer this question, the covariance between dimension and source factors and constructs assessed using measurement systems external from the MSF instrument (e.g., paper and pencil measures and assessment centers) will be examined in this study (Campbell & Fiske, 1959). This approach may be referred to as an external approach to construct validation.

External Approaches to Construct Validation

Certainly, the usefulness of MSF tools is predicated on different sources offering different perspectives of a target's performance (Borman, 1974; Borman, 1997; Lance & Woehr, 1989; Lance, Woehr, & Fisicaro, 1990). Again, if each source provided identical ratings (agreed perfectly), collecting performance ratings from multiple sources would provide redundant information. However, as previously discussed, empirical research on MSF generally indicates that ratings provided from different sources are less strongly related than are rating provided by raters from the same organizational level (e.g., Conway & Huffcutt, 1997) and that the factor structure of MSF is best characterized by dimension and source latent factors (e.g., Lance et al., 1992; Woehr et al., 2005).

Therefore, these findings provide initial evidence that MSF instruments perform in accordance with their philosophical purpose. That is, this research provides evidence that raters from different organizational levels provide different perspectives on ratees' performance. However, empirical research incorporating the internal approaches is less clear as to the meaning of MSF source effects or dimension effects for that matter. Started differently, the wealth of prior research incorporating internal approaches does not speak to the validity of MSF (Borman, 1997; Farr, 2006).

As previously mentioned, one method of examining the meaning of MSF source and dimension effects is to examine the nomological network surrounding these effects. This external approach to construct validity consists of examining the pattern of relationships between a construct and conceptually similar and dissimilar constructs measured using distinct methodologies (Cronbach & Mehl, 1955). This type of research method is used to accumulate evidence focusing on *what* a construct measures (e.g., validity). If one hopes to understand what MSF instruments are actually measuring, this type of evidence is a crucial adjunct to the evidence provided by the internal approaches. In addition, this type of research is crucial in determining whether source specific communality is more appropriately specified as true score or error in MSF studies.

A wealth of research has examined the relationship between performance ratings and a variety of external constructs. This research, largely stemming from the personnel selection literature, has primarily examined the relationship between supervisor ratings and external constructs such as intelligence and personality. To a lesser extent, existing research has also examined the relationship between ratings provided by other sources (e.g., peers and subordinates) and external constructs.

A brief discussion of general trends with respect to the relationship between individual differences and peer, supervisor, and subordinate performance ratings is warranted. In general, research indicates that performance ratings are related to intelligence, conscientiousness, and management skills (Arthur, Day, McNelly, and Edens, 2003; Barrick & Mount, 1991; Schmidt and Hunter, 1998; Tett, Jackson, & Rothstein, 1991). Because different studies use different measures of performance and individual differences, it is very difficult to compare the relationship between individual differences and performance for different rating sources. For example, some research has indicated that intelligence is related to subordinate ratings of performance (e.g., Hoffman & Frost, 2006), while other research has not (e.g., Atwater & Yammarino, 1993; Yammarino & Bass, 1990). With some exceptions, very little research has compared the relationship between individual differences and ratings provided by different sources. Thus, existing research has not demonstrated the degree to which the relationship between individual differences and performance is moderated by the source providing the performance ratings.

To date, Conway, Lombardo, and Sanders (2001) conducted the only quantitative review examining the nomological network of ratings taken from multiple sources. In particular, these authors meta-analytically summarized the relationship between peer and subordinate ratings and external constructs (intelligence, need for affiliation, potency, achievement, dependability, adjustment, agreeableness, intellectance, and rugged individualism). The results of their study indicated that of the individual difference constructs they examined, subordinate ratings weakly related to agreeableness. However, the results of their study also indicated substantial variability in previously reported

relationships between subordinate ratings and individual differences. Finally, peer ratings were weakly related to intelligence, agreeableness, dependability, and affiliation. These authors also compared these results to research investigating the relationship between individual differences and supervisor ratings of performance. Conway and his colleagues (2001) suggested that agreeableness was more strongly related to subordinate and peer ratings than to supervisor ratings, but supervisor ratings were more strongly related to intelligence than were peer or subordinate ratings. Finally, dependability was more strongly related to peer and supervisor ratings than to subordinate ratings in this meta-analysis. Together, the results of this meta-analysis suggest that the relationship between individual differences and performance is moderated by the source providing the ratings.

Although the meta-analysis by Conway and his colleagues provides a useful first step in consolidating the literature of the relationship between peer and subordinate ratings and ratee traits, some important limitations should be noted. Importantly, these authors aggregated all performance measures into an overall performance category. This practice may attenuate relationships between external constructs and performance ratings (Hogan & Holland, 2003). For example, although ability may be unrelated to subordinate ratings of overall performance, it may relate to conceptually similar performance dimensions, such as conceptual/administrative skills. In fact, a meta-analysis by Hogan and Holland (2003) found a similar effect such that the relationship between personality and performance was much stronger when personality scales and performance constructs were matched on the basis of conceptual similarity. Moreover, most of the studies included in the meta-analysis by Conway and his colleagues were based on studies that examined the relationship between traits and performance ratings provided by a single

source (e.g., peer ratings only). In contrast to traditional performance rating instruments, MSF instruments typically contain more items, more dimensions, more ambiguous dimensions, are not based on a job analysis, are used for developmental purposes only, and are not completed by trained raters (Rogelberg & Waclawski, 2001). Consequently, the results of the meta-analysis by Conway and his colleagues may not generalize to settings in which MSF is used as the job performance criteria.

Atkins and Wood (2002) examined the relationship between assessment center (AC) exercise ratings and ratings taken from multiple sources (self, supervisor, subordinate, and peer ratings) in a MSF system. These authors examined the relationship between the overall rating for each rating source (all items aggregated within each rating source), AC exercise ratings, and overall AC performance. Their results indicated that supervisor ratings were the most strongly related to the overall AC rating ($r = .29$), followed by peers ($r = .20$), subordinate ratings ($r = .15$), and finally self-ratings ($r = -.24$). Moreover, when all three sources were combined in analyses, peers and subordinate ratings explained significant variance in AC performance beyond supervisor ratings, lending some evidence for the importance of examining work performance from multiple perspectives.

Despite the contributions of the study by Atkins and Wood, it has some important limitations. Specifically, these authors examined the relationship between AC exercise scores and MSF. By failing to examine the relationship between AC dimension ratings and MSF, this study suffers from a similar limitation as studies examining the relationship between MSF and objective performance (Hazucha, Hezlett, and Schneider, 1993). That is, AC exercise ratings shed little light as to the strengths and weaknesses of

the assesse. Although they maybe a useful index to gauge one's overall performance, exercise ratings do not give an indication of specific skills (Arthur & Villado, in preparation). Hence, the study by Atkins and Wood does not illuminate the relationship between substantively meaningful individual differences and MSF ratings.

Conway (2000) examined the relationship between managerial developmental dimensions rated by multiple sources and conceptually similar personality constructs. Specifically, Conway examined the relationship between three sources' (peers, supervisor, and subordinates) ratings of three dimensions of managerial performance (technical skills, interpersonal effectiveness, and leadership) and twenty different personality constructs. For all three sources, empathy displayed equitable relationships with interpersonal effectiveness. However, socialization was related to supervisor ratings of interpersonal effectiveness but not peer or subordinate ratings. Self-control was significantly related to supervisor and subordinate ratings of interpersonal effectiveness but not peer ratings. Interestingly, none of the personality constructs were related to subordinate ratings of leadership. However, social presence, sociability, and self-acceptance were each similarly related to supervisor and peer ratings of leadership. In contrast, capacity for status, independence, well-being, and achievement via independence were related to supervisor ratings of leadership but not subordinate or peer ratings of leadership. In sum, the study by Conway (2000) demonstrated that personality constructs were differentially related to conceptually similar and dissimilar performance dimensions and that differences in rating source occasionally moderated these relationships.

Despite the contribution of previous research examining the nomological network of ratings taken from different rating sources, each of the previously discussed studies has an important limitation. Specifically, this research has typically aggregated ratings from different sources and examined the relationship between these aggregated scales and external constructs. By examining the relationship between overall ratings and external constructs, the relative impact of individual differences on source and dimension factors is confounded (Arthur & Villado, in preparation; Farr, 2006). That is, the analytic technique that has been relied on by previous research does not allow for the estimation of the relationship among individual differences and the source and dimension factors characteristic of MSF.

For example, the correlation of .15 between ability and peer ratings reported in the meta-analysis by Conway and his colleagues (2001) does not elucidate the degree to which the covariance between peer ratings and ability is attributable to source effects versus dimension effects. Ability could be related to performance dimensions designed to assess conceptual/administrative skills, providing convergent validity evidence for peer ratings of conceptual skills. In contrast, if ability were related to dimensions designed to assess interpersonal skills, this would indicate poor discriminant validity for interpersonal skills dimensions. Similarly, if ability were related to a peer source factor, this would indicate that peers form an overall impression of performance based on coworkers' ability, irrespective of the performance dimension being assessed. However, by confounding the effect of sources and dimensions, previous research examining the nomological network of MSF does not allow for such a determination.

Given the limitations of previous research examining the relationship between performance ratings taken from multiple sources and external constructs, MSF researchers have called for an examination of the relationship between MSF source and dimension effects and externally measured constructs (Arthur & Villado, in preparation; Farr, 2006). It is the purpose of the present study to answer this call by examining the relationship between MSF source and dimension latent factors and external constructs including: AC dimensions, personality constructs, and intelligence. In other words, this study combines external and internal construct validation methodologies to examine the construct validity of MSF.

Importantly, this study will include two methods of measuring of external constructs: AC dimensions and paper and pencil measures of personality and cognitive ability. The incorporation of two types of measures is essential to the present study's methodology. Specifically, research incorporating only one external measure of relevant constructs would be unable to answer the relevant research question if the measure of the construct included was unrelated to MSF ratings. For example, if conceptually similar AC dimensions and MSF dimension effects were unrelated, a researcher would be unable to ascertain the extent to which this finding is caused by poor AC construct validity or poor MSF construct validity. However, given that paper and pencil measures of personality and intelligence have accumulated substantial evidence supporting their construct validity (Arthur & Villado, in preparation), these measures can serve as an additional reference point with which to gauge the construct validity of both MSF ratings and AC dimensions.

A brief discussion of external approaches to addressing the construct validity evidence associated with AC research is relevant here. AC research has typically demonstrated construct validity evidence for AC dimensions when using external construct validation approaches (e.g., Hoffman & Kudisch, 2002; Shore, Thornton, & Shore, 1990; Craik, Ware, Kamp, O'Reilly, Staw, & Zedeck, 2002). For example, Hoffman and Kudisch (2002) and Shore, Thornton, and Shore (1990) provided construct validity evidence for AC dimensions by examining their intercorrelations with conceptually similar and dissimilar measures of personality and cognitive ability. In particular, the study by Hoffman and Kudisch indicated that intelligence was more strongly related to the conceptual skills dimensions measured by the AC than to conceptually dissimilar AC dimensions (e.g., interpersonal and leadership skills) and that dominance was more strongly related to a leadership cluster of AC dimensions than to conceptually dissimilar dimensions (e.g., conceptual and interpersonal skills).

Similarly, Shore, Thornton, and Shore (1990) demonstrated construct validity evidence for performance-style and interpersonal-style AC dimensions. In their study, intelligence was more strongly related to performance-style AC dimensions than to the interpersonal-style dimensions. Similarly, AC interpersonal-style dimensions were more strongly related to conceptually similar personality constructs (e.g., agreeableness) than to conceptually dissimilar personality constructs (e.g., intellectance). Taken together, this stream of research provides construct validity evidence for AC dimensions when a nomological network approach to construct validation is taken. Previous research indicating support for the nomological network of ACs using paper and pencil measures as the external constructs suggests that including both ACs and paper and pencil

measures may be useful to the purpose of "triangulating" evidence for the construct validity of MSF dimension effects.

Each of the three measurement methodologies incorporated in the present study contain scales that can be classified as conceptually similar to the three broad dimensions of managerial skills defined earlier. As previously discussed, AC research has indicated convergence between AC dimensions and conceptually similar constructs assessed using paper and pencil measures of cognitive ability and personality (Kudisch & Hoffman, 2002; Shore et al., 1995). Similarly, MSF research has indicated some degree of convergence between MSF dimensions and conceptually similar personality constructs measured with paper and pencil instruments (Conway, 2000). Together, this research suggests that constructs assessed using performance ratings, paper and pencil measures, and AC dimensions can be categorized on the basis of conceptual similarity. And, construct validity evidence can be evaluated by comparing the relationship between conceptually similar and conceptually dissimilar constructs across the three measurement methodologies incorporated in the present study.

Dimension Effects

The accumulation of construct related validity evidence of constructs assessed using paper and pencil measures is based on rigorous testing to establishing construct validity by investigating the relationship with external measures of conceptually similar and dissimilar constructs (Arthur & Villado, in preparation). In contrast, researchers generally assume that MSF dimensions accurately reflect the underlying performance dimension ostensibly being assessed. That is, researchers rarely provide construct-related validity evidence to support the assumption that MSF dimensions accurately reflect the

dimensions they were designed to assess (Borman, 1997; Farr, 2006). When the construct-related validity evidence of MSF is examined, researchers often rely on the evidence provided by internal approaches as sufficient evidence to support the construct validity of MSF dimensions. As previously discussed, the evidence derived from the internal approaches may be more appropriately viewed as an examination of the reliability of MSF dimensions (Arthur & Villado, in preparation).

Importantly, presenting targets with accurate dimension-level feedback is essential for MSF to be of use in developmental contexts (London & Smither, 1995). Thus, an understanding of the validity of MSF is crucial to the utility of MSF (Borman, 1997). Again, the method taken in the present study of examining the covariance between the dimensions measured by MSF and conceptually similar and dissimilar constructs assessed using external measurement methodologies will be used to give an indication of the validity of MSF dimensions. Given this method, four broad outcomes with respect to the pattern of relationships between MSF dimension effects and external constructs are possible. Each of these possibilities will be discussed in turn.

First, conceptually similar constructs may converge across all three measurement strategies. In the context of the present study, support would be provided for this possibility if indices of interpersonal, leadership, and conceptual/administrative skills measured with different methodologies converged. For example, the interpersonal skills factor from the MSF would be positively related to AC dimensions that assess interpersonal skills and personality constructs indicative of interpersonal skills. The same finding would be expected with respect to alternate measures of conceptual/administrative and leadership dimensions. In fact, this pattern of results would

be both the optimal and the expected finding. That is, convergence across similar constructs assessed using different measurement methodologies is at the crux of evaluating construct validity evidence (Campbell & Fiske, 1959; Cronbach & Meehl, 1955). Practically, demonstrating construct validity evidence for both MSF and AC dimensions would reinforce the incorporation of these measurement strategies into developmental settings.

It is also possible that indicators of conceptual/administrative, interpersonal, and leadership skills will converge across two of the three measurement methods. For example, AC dimensions of conceptual/administrative skills and the paper and pencil measure of intelligence may converge with each other but not correspond to the MSF conceptual/administrative skills factor. This pattern of results would indicate support for the construct validity of the AC dimensions, but not support for the construct validity of the MSF dimensions. Conversely, conceptually similar constructs measured by the MSF instrument and the paper and pencil measures might converge, while the AC dimensions do not converge with conceptually similar constructs assessed with the paper and pencil instruments. This finding would provide construct validity evidence for the MSF dimensions but not for the AC dimensions. It is worthwhile to note that although previous AC research provides evidence for the correspondence between AC dimensions and conceptually similar constructs assessed using paper and pencil measures, research has not examined the correspondence between latent MSF dimension factors and external constructs (Craik et al., 2002; Hoffman & Kudisch, 2002; Shore et al., 1990). Based on the findings of previous research, if this pattern of results is supported, it is likely that the

convergence will be found with respect to AC dimensions and conceptually similar constructs assessed using paper and pencil measures.

An additional possibility is that conceptually similar constructs will not converge across any of the three approaches. For example, indicators of leadership may not be related across any of the three measurement methodologies. This pattern of results would indicate poor construct validity evidence for each of the three measurement strategies. Indeed, such a finding would be quite discouraging for organizational scientists.

For the sake of parsimony, the possibility of support for specific dimensions and not others was omitted from the preceding discussion. Clearly, it is quite possible that conceptually similar constructs will converge across different methods for one of the overarching performance dimensions but not for the other two performance dimensions. For example, conceptual/administrative skills may converge across all three measurement strategies whereas interpersonal and leadership skills do not. To this end, multiple combinations of any of the aforementioned possibilities could occur with respect to the level of convergent and discriminant validity evidence associated with each of the MSF dimension factors. I certainly acknowledge this possibility; however, the purpose of the preceding discussion was to elucidate general trends that could emerge in the data.

Research Question 8: What will the pattern of relationships be between MSF dimension effects and conceptually similar and dissimilar constructs measured with different methods?

Source Effects

The reader will notice that the previous discussion focused solely on the correspondence between MSF dimension effects and external constructs. However, the

covariance between source effects and external measures of individual differences also represents an important area for research. That raters from different organizational sources have different perspectives on performance is an implicit assumption of MSF (Borman, 1997; Woehr et al., 2005). Unfortunately, existing research does not elucidate the extent to which source effects represent SSPRV or error. That is, existing research has not critically evaluated the assumption that source effects actually reflect unique perspective of a target's performance. For the purpose of this study, SSPRV refers to variance relevant to a target's "true performance" that is attributable an interaction between the source providing the ratings and the target's actual performance. In this context, error refers to any variance that is not related to a target's "true performance." In other words, error can be a reflection of either random error or systematic bias that is unrelated to a target's true performance.

Although traditional MTMM research assumes that method effects represent measurement error, source effects in the context of MSF may actually represent SSPRV. In fact, multiple propositions have been forwarded to explain the reason for the presence of source effects and how they may represent performance-relevant information. Borman (1974) was the first to suggest that differences in performance ratings provided by raters from different organizational levels represent SSPRV. In particular, Borman (1974) argued that raters from different organizational levels have different opportunities to observe coworkers' performance and as a result, disagreements across sources provide meaningful, performance relevant information. Similarly, targets may intentionally display different types of behavior in the presence of different raters (Lance & Woehr, 1989). To exemplify, an individual may interact differently with his supervisors than his

subordinates or peers. If this were the case, disagreement across sources would also be considered important information when interpreting ratings of one's performance. Finally, it is also possible that raters occupying different levels in the organizational hierarchy attend to different ratee behaviors because of the value of those behaviors to raters from a given level (e.g., Beauovis & Dubois, 2003). Although all levels may actually see similar behaviors, a particular source may be more likely to attend to and subsequently recall behaviors that are most important to that particular source. For example, supervisors may more closely attend to a target's conceptual/administrative skills than do subordinates because supervisors may feel that the quality of the output of his/her work group (and hence his/her effectiveness) is a reflection of the quality of work conducted by his subordinates. Thus, when supervisors report a target's performance in a MSF setting, they will be more likely to recall performance related to conceptual/administrative skills. Here again, the disagreement across sources represents performance-relevant variance.

Despite multiple propositions that source effects reflect performance relevant variance, previous research has not empirically examined the meaning of source effects. In other words, existing research does not speak to the degree to which source effects represent SSPRV versus error variance. Understanding the meaning of MSF source effects would be of great value to practitioners. Specifically, to the extent that different sources disagree as to the nature of an employees' performance, feedback acceptance and subsequent effort at improving performance may be hindered (Conway, 2000; Smither et al., 2005). Moreover, disparate ratings on a specific skill from different sources would be quite confusing for a feedback recipient. For example, what is a manager to do when organizational leaders indicate they have excellent interpersonal skills, but their peers and

subordinates indicate poor interpersonal skills? Similarly, if all sources disagree as to the nature of many different performance dimensions, the feedback recipient might be overwhelmed and left wondering who to believe and where to focus effort toward improvement.

To this end, work performance researchers argue that an explication of the meaning of source effects is crucial to improving the interpretation of MSF ratings (Borman, 1997; Farr, 2006). These researchers advocate an evaluation of the relationship between MSF source effects and external measures of managerial skills as a useful method of elucidating the meaning of MSF source effects. Similarly, other researchers have proposed that the nomological network of MSF source effects should be investigated (e.g., Conway, 2000; Woehr et al., 2005). A clear understanding of the relationship between source effects and external measures of managerial skills would give an indication of which set of ratings feedback recipients should focus on for certain managerial skills. For example, if a source effect representing subordinate ratings was more strongly related to leadership skills assessed using external measures than other source effects; this would signal that feedback recipients should focus on subordinate ratings when interpreting MSF leadership skills results. This type of information would be invaluable to feedback recipients and executive coaches by indicating which sources' ratings should be the focus for particular dimensions of managerial skills. Finally, if source effects represent error as opposed to SSPRV, MSF researchers must rethink the underlying assumptions made when interpreting MSF data. Such a result would signal the need to consider methods to enhance the differential validity of the ratings provided by

different sources and possibly whether collecting performance ratings form different sources is an efficacious practice at all.

As previously discussed, conceptually similar constructs are expected to converge, whether they are assessed using MSF, ACs, or paper and pencil measures. In contrast, a similar assumption cannot be made with respect to the nomological network surrounding source effects. That is, there is neither empirical nor theoretical precedence with which to classify another construct as conceptually similar or conceptually dissimilar to source effects. Consequently, analyses examining the relationship between source effects and external constructs will be exploratory.

In general, three broad patterns of results may emerge with respect to the relationship between source effects and external measures of managerial skills. First, MSF source effects may be unrelated to constructs assessed using external measurement methodologies. This would provide evidence that source effects do not represent substantively meaningful performance variance and that it is appropriate to interpret variance attributable to the source providing the rating as error (e.g., Schmidt et al., 2000). Or, at the least, that source effects do not represent performance relevant variance with respect to the external constructs typically used to evaluate the performance-relevant individual differences of managers.

Second, MSF source effects may be differentially related to external measures of managerial skills. For example, a latent factor representing peer ratings may be more strongly related to external measures of interpersonal skills than are other source effects (e.g., supervisor and subordinates), whereas a latent factor representing supervisor ratings may be more strongly related to external measures of conceptual/administrative skills

than the other source effects. Again, this pattern of findings would be quite helpful in interpreting source effects and could be quite useful to MSF practitioners in helping feedback recipients to determine which sources' ratings should be focused on for which performance dimension. This pattern of results would provide support for researchers who argue that interpreting all variance not shared across raters as error is inappropriate (Murphy & DeShon, 2000).

Finally, source effects may display equitable relationships with external measures of managerial skills. To illustrate, external measures of interpersonal skills may relate equitably to each of the latent source factors. This type of result would indicate that although source effects are substantively meaningful reflections of performance relevant variance, they provide the same type of information across sources. In other words, although source effects represent performance-related variance, this variance is equally related to constructs measured using external methods.

Research Question 9: What will the pattern of relationships be between MSF source effects and external constructs?

Summary

MSF is a commonly employed method of assessing managers' strengths and weaknesses. To this end, the utility of MSF systems is predicated on the accuracy of the feedback provided by MSF tools. The preponderance of research investigating the construct validity of MSF has incorporated an internal approach to construct validation. This research indicates that both dimension and source effects are important components of MSF ratings and that most MSF instruments are equivalent across sources. Importantly, some suggest that studies using the internal approaches are more

appropriately viewed in the context of the reliability of MSF scales (Arthur & Villado, in preparation) while others believe that interpreting the results from the internal approaches in the context of MSF reliability is inappropriate (Murphy & DeShon, 2000). In either case, determining what is being measured by MSF instruments is not possible if the results from research incorporating the internal approach are considered in isolation (Borman, 1997; Farr, 2006). Consequently, despite consistent agreement as to the latent structure of MSF, the construct validity of source and dimension effects has not yet been resolved.

One potential method of examining the construct validity of MSF dimension and source effects is to examine the relationships between these effects and constructs measured using external measurement methodologies. This approach would allow for a determination of the construct validity of dimension effects and the extent to which source effects represent performance relevant variance or error. However, existing research examining the relationship among MSF ratings and external constructs has confounded source and dimension effects. Thus, the present study will combine both internal and external approaches to construct validity so that the relative impact of MSF dimension and source effects can be partialled before examining the nomological network of MSF.

In this study, I will incorporate multiple approaches to evaluating the psychometric properties of MSF. First, previous internal construct validation research will be replicated with respect to agreement across sources, the equivalence of performance ratings by raters from different organizational levels, and the factor structure of multisource performance ratings. Next, the relationship between latent factors

representing MSF dimension effects and external constructs will be examined to determine the degree to which the dimensions measured by MSF are related to external measures of conceptually similar constructs and not related to conceptually dissimilar constructs. The results of this set of analyses will be used evaluate the construct validity of MSF dimensions. Additionally, the degree to which source effects represent SSPRV will be investigated by examining the relationship among source effects and external measures of performance-relevant individual differences. This set of analyses will speak directly to the assumption that raters from different organizational levels offer unique perspectives on their coworkers' performance. Together, the findings of this study will be discussed in the framework of the construct validity of MSF with a particular focus on implications for organizational scientists interested in using MSF for developmental purposes.

CHAPTER 3

Methodology

Participants

The sample for this study consisted of 404 managers enrolled in different divisions of an Executive Masters of Business Administration (EMBA) program at a large southeastern university. That is, managers enrolled in Senior EMBA (SEMBA), Aerospace EMBA (AEMBA), and Physicians EMBA (PEMBA) programs served as participants in this study. While enrolled in the EMBA program, the research participants concurrently worked as managers in a diverse range of organizations and industries. The majority of the participants were Caucasian (82%) males (68%) with a mean age of 44, an average of 11.3 years of managerial experience, and supervised 10 direct reports on average.

For the MSF ratings, coworkers of the 404 participants completed ratings. Specifically, 404 higher level managers, 1,236 peers, and 1,255 subordinates completed performance ratings of the 404 target managers. The majority of the higher level manager sample consisted of white (86%) males (85%) with an average age of 49 years and had worked with the target for an average of 5.4 years. The peer sample was largely white (84%) males (76%) with an average age of 44 years, and had worked with the target for an average of 5.3 years. Finally, the subordinate sample was 82% white, 50% male, with an average age of 42 years, and had worked with the target manager for 4.2 years on average.

Procedure

Before beginning the EMBA program, participants completed the Watson-Glaser Critical Thinking Appraisal (CTA; Watson & Glaser, 1980), the California Psychological Inventory (CPI; Gough & Bradley, 1996), and participated in an AC. For the MSF performance ratings, participants were mailed multi-source feedback forms to be completed by themselves as well as their supervisors, subordinates, and peers prior beginning the MBA program. The participants were allowed to choose which of their coworkers would complete the performance ratings. Those completing the surveys were instructed to mail the MSF forms directly back to the university upon completion and ensured that any information provided would be confidential and that personally identifying information would be used for statistical purposes only.

Approximately two months after completing all the measures, each participant was mailed a comprehensive feedback report containing: 1) a summary of their coworkers' ratings separated by rating source and dimension including a brief description of the construct definition of each rating scale, 2) a summary of their performance in the AC including a summary definition of each exercise and dimension and a DVD containing a recording of each exercise, and 3) graphical and written descriptions of their personality profile and CTA results. After receiving this feedback, each participant was paired with an executive coach who helped the participant draft a developmental action plan based on the feedback from these instruments. The participants were then responsible for meeting with their executive coach once a month to discuss approaches to meeting their performance improvement goals and progress on their developmental action plan.

Measures

Assessment Center Method

The AC was used for developmental purposes only. Exercises included two simulation exercises (e.g., 1-on-1 role plays), a leaderless group discussion, and an in-basket exercise. At least two experienced assessors made ratings for each participant on each exercise. These assessors participated in approximately twenty hours of frame of reference training (FORT) prior to serving as raters. In addition, all raters were given a two hour FORT "refresher" course each year prior to the administration of the ACs. AC dimensions rated included: analysis, judgment, oral communication, written communication, planning and organizing, decisiveness, initiative, leadership, sensitivity, persuasiveness/confrontation, team-building, delegation, stress tolerance, customer-service orientation, and coaching. Dimensions were rated on a 5-point behaviorally anchored rating scale ranging from 1 = “Unsatisfactory” to 5 = “Outstanding.” A staff of senior assessors participated in a consensus meeting to obtain the final ratings for each participant's performance on each skill dimension. The ratings generated on the basis of the consensus, or post consensus dimension ratings (PCDR) were used to operationalize AC dimension performance.

Personality

Personality dimensions were assessed using the California Psychological Inventory (Gough & Bradley, 1996; CPI). The CPI is a widely used and accepted measure of normal personality that measures twenty folk scales and several special-purpose scales using 435 items presented in the form of true-false statements. Primary scales assessed by the CPI include: dominance, sociability, social presence,

responsibility, self-acceptance, independence, empathy, responsibility, socialization, communality, well-being, tolerance, flexibility, achievement via independence, achievement via conformity, good impression, capacity for status, self-control, intellectual efficiency, masculinity, and psychological mindedness. The CPI manual reports acceptable reliabilities for each of these subscales and correlations of each subscale with conceptually similar constructs, supporting the construct validity of the scales of the CPI (Gough & Bradley, 1996).

Cognitive Ability

Cognitive ability was assessed using the Watson Glaser Critical Thinking Appraisal. This 80-item instrument is designed to measure critical thinking skills and has frequently been used in research as a measure of general mental ability (GMA). Watson and Glaser (1980) report split-half reliabilities ranging from .69 to .85, a parallel form reliability coefficient of .75, and strong correlations among other measures of GMA, supporting the construct validity of the CTA as a measure of critical thinking skills.

Multisource Feedback

A multisource feedback instrument designed to assess a variety of managerial competencies was used in the present study. The managerial competencies were assessed with 117 items and included: idealized influence, inspirational motivation, individualized consideration, intellectual stimulation, analysis, judgment and decision making, planning and organizing, team-building, sensitivity, tactfulness during confrontation, communication skills, stress tolerance, performance management, initiative, organizational acumen, integrity, and receiving feedback. Respondents were asked to make ratings on a five-point scale with 1 = "Strongly Disagree" and 5 = "Strongly

Agree". There is also a "Don't Know" response option. Supervisor, peer, and subordinate responses were used in the present study.

Preliminary Data Analyses and Procedures

Preparation for Data Analysis

Missing data was estimated using a pair-wise deletion strategy. In addition, previous experience with the MSF instrument has indicated that it is sometimes subject to random response bias. To eliminate the potential attenuating effect of response bias on substantive hypothesis testing, any respondent whose responses were exactly the same across all relevant items on the MSF form were deleted. Specifically, any rater who gave exactly the same numerical response (e.g., all "5" or all "1") for all items was eliminated from statistical analyses. This decision rule resulted in the deletion of 18 raters.

Item Composites

For both statistical and conceptual reasons, analysis of the MSF instrument was based on item composites, wherein individual items are combined to form scales prior to analyses. In this study, items measuring the same subscale on the MSF tool were averaged to form a single manifest indicator of that construct. For example, the four items designed to assess "inspirational motivation" were combined to form a single manifest indicator of "inspirational motivation." This procedure has been recommended by Lance, Woehr, and Fisicaro (1991) and West, Finch, & Curran (1991) for a variety of statistical and conceptual reasons.

Conceptually, this approach represents a "latent construct" approach to measurement that has been incorporated and recommended by other work performance researchers (cf., Hoffman, Blair, Meriac, & Woehr, in press; Lance et al., 1992; LePine et

al., 2002). Briefly, the latent construct approach entails viewing conceptually similar subscales as imperfect indicators of an underlying performance factor. That is, instead of loading items on a single factor, conceptually similar groups of subscales are specified to load on an overarching factor. For example, analysis and planning and organizing may each be seen as indicators of conceptual skills. Thus, this approach specifies that the items from these two subscales be aggregated to form two separate manifest indicators of analysis and planning and organizing. The resulting scale scores for each subdimension are subsequently loaded on a latent conceptual/administrative skills factor.

Similarly, Bagozzi and Edwards (1998) distinguished between aggregation and disaggregation approaches to construct models. Briefly, disaggregation approaches entail basing construct models on individual items. In contrast, the aggregation approach to construct modeling involves aggregating items to form more abridged representations of a construct. Bagozzi and Edwards (1998) recommended the aggregation approach to construct modeling when the purpose of the study is to examine broadly defined constructs as opposed to the nuances of particular items. Because the focus of this study is on broad factors of performance and is less concerned with the operation of specific items, the aggregation approach used here is the appropriate methodology. To this end, a similar approach has been incorporated by other researchers examining MSF using structural modeling (e.g., Scullen et al., 2003).

Statistically, item composites significantly reduce the number of parameter estimates required in a given model and as a result, decrease the probability of obtaining an improper solution (West, Finch, & Curran, 1995). Moreover, the item composite approach results in increased reliability of indicators and decreases the potential for item-

specific variance biasing parameter estimates. Finally, results based on item composites are more likely to approach a normal distribution and generalize across samples than are results based on individual items (West et al., 1995). For the aforementioned conceptual and statistical reasons, the item composite approach will be used in this study.

Structure of Performance Ratings

As previously discussed, the present study conceptualizes managerial performance with a model consisting of conceptual/administrative, interpersonal, and leadership skills. The 18 subscales measured by the MSF instrument were conceptually classified as indicative of conceptual/administrative, interpersonal, or leadership skills using Borman and Brush's taxonomy of managerial performance. Two subject matter experts (SMEs) independently compared the list of subscales provided by Borman and Brush with the subscales assessed by the MSF instrument and assigned each of the MSF subscale into one of the three performance mega-dimensions provided by Borman and Brush. Table 1 presents the results of the classification of the MSF subscales. Of the 18 performance subscales measured, 11 were classified into the same dimension by both SMEs. Three subscales were classified as indicative of conceptual/administrative skills (analysis, judgment and decision making, and planning and organizing), four were classified as interpersonal skills (communicating with others, tactfulness when confronting, sensitivity, and team building), and four were classified as leadership skills (inspirational motivation, idealized influence, intellectual stimulation, and performance management). Six subscales were not classified by either SME as these subscales were not represented in Borman and Brush's three factor taxonomy (participation, stress tolerance, organizational acumen, integrity, initiative, and receiving feedback). Finally,

Table 1. Classification of Managerial Skills and External Constructs

Conceptual Skills	Interpersonal Skills	Leadership Skills
Judgment & Decision Making[1]	Communication Skills[1]	Idealized Influence[1]
Decisiveness[2]	Oral Communication[2]	Inspirational Motivation[1]
Analysis[3]	Confrontation Tact[2]	Intellectual Stimulation[1]
Judgment[3]	Team Building[3]	Performance Management[1]
Planning and Organizing[3]	Sensitivity[3]	Influencing Others[2]
Responsibility[4]	Social Presence[4]	Persuasiveness[2]
Intelligence[5]	Empathy[4]	Coaching[2]
	Responsibility[4]	Dominance[4]
	Good Impression[4]	Social Presence[4]
	Tolerance[4]	Independence[4]
	Flexibility[4]	

Note. [1] denotes measured using MSF. [2] denotes measured by the assessment center. [3] denotes measured using both AC and MSF. [4] denotes measured by the California Psychological Inventory. [5] Denotes measured by the Watson Glaser Critical Thinking Appraisal.

one subscale, individualized consideration, was classified into two different mega-dimensions by each SME. One SME classified this dimensions as indicative of interpersonal skills, whereas the other classified it as leadership skills. Consequently, this scale was omitted from subsequent analyses.

The AC dimensions were also rationally classified using Borman and Brush's three factor taxonomy. Each SME was presented with the classification system outlined by Borman and Brush and a description of each of the 16 assessment center dimensions. Each SME was then asked to independently classify the AC dimensions into the three performance factors by comparing the content of the AC dimensions to the constructs provided in Borman and Brush's taxonomy. Table 1 also presents the results of the SME's classification of the AC dimensions. Of the 16 assessment center dimensions, 10 were classified into the same dimension by each SME. Specifically, 4 dimensions were classified as indicative of conceptual/administrative skills (analysis, judgment, planning and organizing, and decisiveness), three were classified as indicative of interpersonal skills (oral communication, sensitivity, and team building), and three were classified as indicative of leadership skills (leadership, persuasiveness, and coaching). Six of the dimensions could not be classified into Borman and Brush's three mega-dimensions (initiative, delegation, customer orientation, and stress tolerance). Although not a primary purpose of this study, it is important to examine the empirical construct validity of the SMEs' classification with respect to the AC dimensions. Thus, a CFA was conducted with each PCDR set to load on the agreed upon factor of managerial skills. This procedure of conceptual classification of PCDRs and subsequent empirical evaluation is consistent with previous research using a nomological network approach

to explore the construct validity of AC dimensions (Craik et al., 2002; Hoffman & Kudisch, 2002; Shore et al., 1990).

Finally, constructs measured using the paper and pencil measures (CPI and CTA) were also classified into the three performance dimensions by two SMEs familiar with these instruments. Again, the SMEs were provided with a description of each of the three performance categories that included a list of each of the MSF subscales and AC dimensions previously classified into each respective performance dimension and a description of each construct measured with the paper and pencil instruments. The SMEs were then asked to evaluate the degree of conceptual similarity between the constructs assessed by the paper and pencil measures and each of the three categories of performance on a 5-point scale ranging from 1 "very dissimilar" to 5 "very similar". For a subscale to be deemed conceptually similar to one of the performance domains, the average of the two SMEs' ratings had to have been 4.0 or above. Of the 21 constructs measured with the paper and pencil instruments, nine were deemed to be conceptually similar to one or more of the three managerial performance dimensions. Specifically, responsibility and GMA were classified as conceptually similar to conceptual/administrative skills; social presence, empathy, flexibility, good impression, tolerance, and responsibility were classified as conceptually similar to interpersonal skills; and dominance, independence, and social presence were classified as conceptually similar to leadership skills. Definitions of each of the subscales assessed using the paper and pencil instruments relevant to this study are presented in Appendix B.

In contrast to the AC dimensions and MSF instrument, the factor structure of the paper and pencil instruments was not empirically assessed. Because the paper and pencil

instruments (CPI and CTA) used in the present study are established measures, the SMEs' conceptual classification was accepted. Finally, the relationship between the distinct subscales measured by the paper and pencil instruments and the other constructs in the study were examined in all remaining analyses. That is, the SMEs' conceptual similarity ratings were used to determine whether the constructs assessed using the paper and pencil instruments were conceptually similar or conceptually dissimilar to the three categories of managerial performance, not to examine the factor structure of personality and cognitive ability. Such an analysis would be far beyond the scope of this study.

Within Source Agreement

In order to ensure adequate agreement within subordinate and peer rater groups on the MSF instrument, within rater agreement (r_{wg}) was calculated (James, Demaree, & Wolf, 1984). The r_{wg} statistic is used to assess interrater agreement based on a comparison of observed within group agreement to the agreement one would expect by chance (James et al., 1984). James and his colleagues argued that traditional interrater reliability indices underestimate agreement because the variance in responses to psychological measures is often restricted. In contrast, r_{wg} essentially corrects for this restriction in variance by comparing the observed variance across responses to the amount of possible variance in a scale. For these reasons, James and colleagues argue that r_{wg} is a more accurate, realistic method of assessing interrater agreement than traditional interrater reliability coefficients.

An important issue when calculating r_{wg} is the amount of variance expected in a given sample. Typically, r_{wg} has been calculated using the uniform distribution. The use of this distribution assumes that ratings will be randomly distributed between one and

five on a five point scale. Work performance researchers incorporate the uniform distribution because it is reflective of performance in the general population. Implicit in this statement is that the skewness typical of performance ratings is reflective of actual performance. As such, the uniform distribution will be used in the present study. According to James and his colleagues, r_{wg} vales above .80 are considered "acceptable", whereas r_{wg} vales below .80 indicate unacceptable levels of interrater agreement. For the purposes of the present study, the entire case for which a target's coworkers' ratings fell below .80 were omitted from analyses. Alternatively, acceptable levels of agreement provided the justification for aggregating subordinate and peer responses.

A special form of r_{wg}, $r_{wg(j)}$, was calculated to ensure adequate agreement among subordinate rater groups and peer rater groups on the dimensions of the MSF instrument (James et al., 1984). In essence, $r_{wg(j)}$ provides an agreement estimate based on a multi-item scale. In the context of this study, $r_{wg(j)}$ was calculated to estimate agreement among peer raters and among subordinate raters for each of the three mega-dimensions of performance. Consistent with the incorporation of scale-level data as manifest indicators of broad managerial skills taken in this study, each subscale indicative of a certain performance domain was treated as an item when calculating $r_{wg(j)}$. For example, each peer rater's ratings of analysis, judgment, and planning and organizing were treated as items measuring conceptual/administrative skills in these analyses. The results indicated an acceptable level of agreement for peer ratings of: conceptual/administrative skills (median $r_{wg(j)} = .89$), interpersonal skills (median $r_{wg(j)} = .89$), and leadership skills (median $r_{wg(j)} = .89$). Similarly, levels of agreement among subordinate raters was in acceptable range for conceptual/administrative skills (median $r_{wg(j)} = .89$), interpersonal

skills (median $r_{wg(j)}$ = .90), and leadership skills (median $r_{wg(j)}$ = .87). Although agreement was typically in acceptable range to justify aggregation, several of the individual rating groups did not evidence adequate levels of agreement. Specifically, peer ratings for 18 of the targets and subordinate ratings for 14 of the targets did not reach the recommended $r_{wg(j)}$ value of .80 to justify aggregation. Six of the cases of unacceptable $r_{wg(j)}$ values were for the same target for both rating groups, resulting in 26 total cases being deleted from analyses. The remaining raters within each source for each target were subsequently aggregated.

Evaluation of Model Fit

Research Questions one, five, six, seven, eight, and nine were analyzed using structural modeling. The evaluation of the appropriateness of the structural models specified in these research questions focused on an evaluation of relevant fit indices. Specifically, model evaluation incorporated five overall fit indices including: χ^2 test, Steiger's (1990) Root Mean Square Error of Approximation (RMSEA), Browne and Cudek's (1989) Expected Cross Validation Index (ECVI), the Tucker Lewis Nonnormed Fit Index (NNFI; Tucker & Lewis, 1973) and the Comparative Fit Index (CFI; Bentler, 1990). Although the χ^2 test is the most common method of examining the fit of measurement models, χ^2 tests are overly sensitive to large sample sizes. Specifically, χ^2 tests tend to produce significant results even with a relatively small degree of misfit. Thus, model evaluation largely focused on the four additional fit indices. Browne and Cudek (1993) suggest that RMSEA represents a measure of lack of fit per degree of freedom and that a value of .05 or less represents close fit whereas values up to .10 represent reasonable fit. The ECVI is an indication of model fit that incorporates both

model fit and the number of parameters used. Consequently, it is particularly useful to compare alternative models by ranking the models according to their ECVI value and choosing the model with the smallest value as providing the best representation of the data. Both NNFI and CFI are relative fit indices that 1) evaluate model fit relative to a null model, and 2) take into account the overall number of model parameters estimated. Both the NNFI and CFI range from zero to one with values closer to 1.0 indicating better model fit. General rules of thumb suggest that CFI and NNFI values between .90 and .95 indicate acceptable model fit, and values above .95 indicate good fit.

CHAPTER 4

Results

Research Questions 2-4: Agreement in MSF

Research Questions 2-4 were concerned with the level of agreement in performance ratings taken from raters occupying similar and different organizational levels. Research Question 2 sought to determine the correspondence between ratings provided by different sources. Evaluation of this research question was based on an examination of the correlation among each source's ratings. The MTMS correlation matrix, means, standard deviations, and coefficient alpha reliabilities among all study variables is presented in Appendix C. The average correlations between 1) peer and supervisor ratings, 2) peer and subordinate ratings, and 3) supervisor and subordinate ratings were compared. Before averaging the correlations, each correlation was converted to its corresponding z' value using Fisher's r to z' transformation. This approach is necessary because Pearson's r is not normally distributed. Fisher's z' transformation converts Pearson's r to a normally distributed variable, which allows for the averaging of correlations (Howell, 1992). The average correlation between peer and supervisor ratings was .25, the average correlation between supervisor and subordinate ratings was .19, and the average correlation between peer and subordinate ratings was .27. The average correlation between ratings provided by different sources was .23. Together, these results suggest that consistent with previous research, the correlation between different sources' ratings is somewhat weak.

Research Question 3 sought to determine the extent to which the correspondence among ratings obtained from the same source is greater than the correspondence in ratings taken from different sources. This question was evaluated on the basis of a comparison of the correlation within individual rating sources (e.g., the correlation between each target's subordinates' ratings) and the correlation across ratings sources (e.g., the correlation between peer and supervisor ratings). To the degree that the within source correlations are greater than correlations obtained across sources, preliminary support would be demonstrated for the impact of rating source on performance ratings. The results indicated that the mean within source correlation (mean r = .37) was greater than the average correlation across sources (mean r = .23), providing preliminary support for the presence of MSF source effects.

Finally, Research Question 4 was concerned with the possibility of differential levels of agreement across ratings sources, depending on the performance dimension being rated. A determination of the relevant dimensions was based on the SMEs' rational classification of performance dimensions into Borman and Brush's three categories of managerial performance. The average correlations between peers and subordinates on subscales classified as interpersonal skills (mean r = .26), leadership skills (mean r = .28), and conceptual/administrative skills (mean r = .28) were equitable. Similarly, the average correlations between supervisors and peers on subscales classified as interpersonal skills (mean r = .25), leadership skills (mean r = .27), and conceptual/administrative skills (mean r = .26) were also highly similar. Finally, the average correlations between supervisors and subordinates on subscales classified as interpersonal skills (mean r = .20), leadership skills (mean r = .17), and conceptual/administrative skills (mean r = .21)

were very similar. Together, these results show that the performance dimension being rated does not have an impact on the correlation between ratings provided by raters from different organizational levels, indicating a negative for Research Question 4.

To summarize, research questions 2-4 pertained to the level of correspondence within and across performance ratings made by raters from different organizational levels. Consistent with prior research (e.g., Conway & Huffcutt, 1997), the results regarding the relative degree of correspondence within and across sources suggest higher within source correspondence compared to across source correspondence. In addition, performance dimension being rated did not have an impact on across source agreement. However, ratings provided by peers and supervisors and peers and subordinates covaried to a greater extent than ratings provided by subordinates and supervisors.

Research Questions 1, 5-7: Modeling MSF

Confirmatory factor analysis (CFA) was used to investigate the research questions concerning the structure of performance ratings. The classification agreed upon by the SMEs was used to determine which MSF subscales were specified to load on each of the three performance factors. The variance-covariance matrix among subscales measured by each source providing ratings on the MSF instrument served as the input into LISREL version 8.5 (Joreskog & Sorbom, 1996). Research Question 1 focused on the dimensionality of managerial performance. In order to examine this question, the various models of MSF ratings were examined by specifying a three performance dimension framework in each relevant MSF model. In other words, the dimensionality of managerial performance was examined in conjunction with the five MSF models. Evidence concerning the dimensionality of managerial performance in the present sample is

contingent on examination of each of the proposed MSF models and as such, will be discussed in concert with analyses examining the latent structure of MSF.

Research Question 6 focused on models underlying performance ratings obtained form multiple sources. Based on the results of previous research, five models potentially characterizing MSF ratings were proposed. The results of the confirmatory factor analyses testing the structure of MSF are presented in Table 2. The previously presented Figures 1-5 graphically depict each of the proposed MSF models. Due to space constraints, the manifest indicators and disturbance terms were omitted from these graphical depictions. The first model specified a general performance factor across sources where all ratings provided by all raters are specified as manifest indicators of a single performance factor (Model 1). Results of the CFA indicated that this model did not provide an adequate representation of the data (χ^2 = 13,749.32; ECVI = 38.79; RMSEA = .27; NNFI = .35; CFI = .39).

The second model tested a MSF structure composed of three dimension factors (Model 2). In this model, each source's ratings of a given underlying performance dimension loads on the same latent dimension factor. For example, supervisor, peer, and subordinate ratings on the three conceptual/administrative skills subscales were set to load on a single conceptual/administrative skills factor. The second and third factors included each source's ratings on the subscales classified as interpersonal and leadership skills. Based on the results of the CFA, a model consisting of three dimensions does not fit the MSF data well (χ^2 = 13,802.32; ECVI = 38.83; RMSEA = .27; NNFI = .36; CFI = .41).

Table 2: Model Fit Statistics for Structural Models

	df	χ^2	ECVI	RMSEA	NNFI	CFI
Models of MSF						
Model 1: 1 General Factor	495	13794.32	38.79	.27	.35	.39
Model 2: 3 Dimension Factors	492	13802.51	38.83	.27	.36	.41
Model 3: 3 Source Factors	492	3495.24	10.12	.13	.77	.78
Model 4: 3 Sources and 1 Dimension	459	2149.29	6.56	.10	.83	.85
Model 5: 3 Source and 3 Dimension Factors	456	1548.40	4.90	.08	.87	.89
Measurement Equivalence						
Model 1: Congeneric	456	1548.40	4.90	.08	.87	.89
Model 2: Tau-equivelant	478	1646.07	5.05	.08	.87	.88
Model 3: Parallel[a]	--	--	--	--	--	--
Models of AC Dimensions						
One factor	35	268.65	.86	.14	.75	.80
Two Factor	34	55.05	.27	.042	.97	.98
Three Factor	32	49.79	.27	.039	.97	.98
Full Model	1106	2392.36	8.18	.06	.87	.89

Note. [a] this model failed to converge.

Third, a model with a latent factor representing each of the three sources was proposed (Model 3). This model specifies that all ratings made by a single source load on a single latent factor. Specifically, all supervisor ratings of all subscales load on a supervisor latent factor, all peer ratings of all subscales load on a peer latent factor, and all subordinate ratings on all subscales load on a subordinate latent factor. Again, the results of the CFA indicate that the MSF data is not adequately explained by this model (χ^2 = 3,495.24; ECVI = 10.12; RMSEA = .13; NNFI = .77; CFI = .78).

Next, a three source factor, one performance dimension factor model was tested (Model 4). This model indicates that all ratings made by a single source load on one of three source factors, and all ratings provided by all sources load on a general performance factor. Consistent with previous research, the source and dimension factors are not allowed to correlate in this model (the three elements of the phi matrix representing the correlation between the source and dimension latent factors were set to 0). This constraint was set because MTMM models using the correlated trait-correlated method approach often suffer from identification/convergence problems (Lance, Noble, & Scullen, 2002). CFA results suggested that this model adequately explained the MSF data (χ^2 = 2,149.29; ECVI = 6.56; RMSEA = .10; NNFI = .83; CFI = .85).

Finally, a model composed of three source and three performance dimension latent factors was examined (Model 5). This model proposes that the three sources' ratings on the subscales of a given performance dimension will load on a single dimension latent factor, resulting in three dimension latent factors. And, three latent source factors with all subscales assessed by a single source set to load on a single latent factor were modeled for each of the three sources. Similar to the previous model

specifying both source and dimension factors, the elements of the phi matrix representing the correlations between latent source and latent dimension factors were set to 0 in this model. The fit indices associated with this model suggest that a model specifying three source and three dimension factors adequately represents the MSF data (χ^2 = 1548.4; ECVI = 4.9; RMSEA = .08; NNFI = .87; CFI = .89).

An inspection of the model fit indices presented in Table 2 reveals that a MSF model consisting of three performance dimension latent factors and three source latent factors provides the best fit of the five models tested in the present study (Model 5). Each of the fit indices associated with this model is consistent with a greater degree of model fit when compared to the other four models tested. As previously discussed, the ECVI is a particularly useful index for comparing competing models such that models with the lowest ECVI provide the best representation of the data. Here again, the ECVI of Model 5 is the lowest of the models tested. Together, the fit indices support the six factor MSF model as the model that model best characterizes the MSF data. In addition, the support of the three performance dimension factor model with manifest indicator (e.g., subscales) loading consistent with those forwarded in Borman and Brush's taxonomy suggests that managerial performance can appropriately be conceptualized with three dimension factors including: conceptual/administrative, interpersonal, and leadership performance. The completely standardized parameter estimates for the six factor MSF model are presented in Table 3.

In sum, the CFA results examining the factor structure of the MSF instrument suggest that the MSF data in the present study is 1) best characterized by three performance dimensions including conceptual/administrative, interpersonal,

Table 3: Standardized Parameter Estimates for the Six Factor Model.

Performance Dimension	Parameter Loadings						
	Concp [a].	Intp [a]	Lead [a]	Manager [a]	Peer [a]	Subord [a]	Unique Variance [b]
II Manager			.41*	.73*			.32
II Peer			.34*		.82*		.22
II Subordinate			.31*			.81*	.26
IM Manager			.47*	.60*			.42
IM Peer			.49*		.73*		.22
IM Subordinate			.41*			.77*	.25
IS Manager			.10*	.76*			.42
IS Peer			.20*		.75*		.39
IS Subordinate			.16*			.79*	.35
PM Manager			.02	.63*			.61
PM Peer			.16*		.77*		.38
PM Subordinate			.09*			.76*	.41
TB Manager		.56*		.63*			.29
TB Peer		.51*			.72*		.23
TB Subordinate		.41*				.77*	.25
SEN Manager		.63*		.61*			.23
SEN Peer		.57*			.55*		.37
SEN Subordinate		.45*				.69*	.33
CO Manager		.14*		.67*			.53
CO Peer		.14*			.72*		.46
CO Subordinate		.03				.72*	.48
CS Manager		.41*		.69*			.37
CS Peer		.41*			.71*		.33
CS Subordinate		.28*				.77*	.33
AN Manager	.02			.85*			.27
AN Peer	.20*				.83*		.28
AN Subordinate	.02					.89*	.21
JD Manager	.15*			.83*			.29
JD Peer	.63*				.81*		.05
JD Subordinate	.20*					.86*	.23
PO Manager	.06			.74*			.45
PO Peer	.10*				.82*		.32
PO Subordinate	.02					.82*	.32
Mean	**.16**	**.38**	**.26**	**.70**	**.75**	**.79**	**.33**

Note. * denotes a significant path loading. II = Idealized Influence; IM = Inspirational Motivation; PM = Performance Management; TB = Team Building; SENS = Sensitivity; CO= Communicating with Others; CS = Confrontation Skill; AN = Analysis; JD = Judgment; PO = Planning and Organizing; Concp = Conceptual Skill; Intp = Interpersonal Skills; Lead = Leadership Skills; Subord = Subordinate. [a] LISREL lambda x completely standardized parameter estimates. [b] LISREL diagonal theta delta completely standardized parameter values.

and leadership performance, and 2) best characterized by a model consisting of three source and three performance dimension factors. Together, these results indicate an affirmative to Research Question 1 and Research Question 6.

Research Question 7 pertained to the relative proportion of variance accounted for by performance dimension, rating source, and uniqueness components of the MSF ratings. This research question was examined with the respective parameter estimates provided by LISREL version 8.5 for dimension, source, and uniqueness components for Model 5. To calculate the proportion of variance accounted for by performance dimension and rating source, the completely standardized parameter estimates were squared and averaged for each manifest indicator's loading on the source and dimension latent factors. To estimate the variance explained by the uniqueness component, the diagonal theta delta completely standardized parameter values provided by LISREL version 8.5 were averaged for each observed rating. Table 4 presents the proportion of variance accounted for by each rating component (performance dimension, ratings source, and uniqueness) in each observed rating. On average, the source factor accounted for the largest proportion of the variance (56%), followed by the uniqueness component (33%), while performance dimension explained the smallest amount of variance in performance ratings (11%). For manager ratings, source effects accounted for an average of 50% of the variance, uniqueness for 38%, and performance dimension for 11%. For peer ratings, source effects accounted for 57% of the variance, uniqueness for 29%, and performance dimension for 15% of the variance. For subordinate ratings, source effects accounted for 62% of the variance, uniqueness for 31%, and performance dimension for 7% of the variance. In short, these results suggest that source effects account for the most

Table 4: Proportion of Variance Attributable to Dimension, Source, and Uniqueness Components from the Six Factor Model

Performance Dimension	Variance Source						
	Concp [a]	Intp [a]	Lead [a]	Manager [a]	Peer [a]	Subord [a]	Unique Variance
Manager Rating							
II			0.17	0.53			0.32
IM			0.22	0.36			0.42
IS			0.01	0.58			0.42
PM			0.00	0.40			0.61
TB		0.31		0.40			0.29
SEN		0.40		0.37			0.23
COM		0.02		0.45			0.53
CSK		0.17		0.48			0.37
AN	0.00			0.72			0.27
JUD	0.02			0.69			0.29
PO	0.00			0.55			0.45
Mean	**0.01**	**0.22**	**0.10**	**0.50**			**0.38**
Peer Rating							
II			0.12		0.67		0.22
IM			0.24		0.53		0.22
IS			0.04		0.56		0.39
PM			0.03		0.59		0.38
TB		0.26			0.52		0.23
SEN		0.32			0.30		0.37
COM		0.02			0.52		0.46
CSK		0.17			0.50		0.33
AN	0.04				0.69		0.28
JUD	0.40				0.66		0.05
PO	0.01				0.67		0.32
Mean	**0.15**	**0.19**	**0.11**		**0.57**		**0.29**
Subordinate Rating							
II			0.10			0.66	0.26
IM			0.17			0.59	0.25
IS			0.03			0.62	0.35
PM			0.01			0.58	0.41
TB		0.17				0.59	0.25
SEN		0.20				0.48	0.33
COM		0.00				0.52	0.48
CSK		0.08				0.59	0.33
AN	0.00					0.79	0.21
JUD	0.04					0.74	0.23
PO	0.00					0.67	0.32
Mean	**0.01**	**0.11**	**0.08**			**0.62**	**0.33**

Note. II = Idealized Influence; IM = Inspirational Motivation; PM = Performance Management; TB = Team Building; SENS = Sensitivity; COM = Communicating with Others; CF = Confrontation Skill; AN = Analysis; JUD = Judgment; PO = Planning and Organizing; Concp = Conceptual Skill; Intp = Interpersonal Skills; Lead = Leadership Skills; Subord = Subordinate. [a] Proportion of variance accounted for by performance dimensions and rating source is represented as the squared parameter loadings.

variance in performance ratings followed by error variance and performance dimension effects.

To examine the equivalence of the multisource ratings (Research Question 5), the approach suggested by Woehr and his colleagues (2005) was used. These authors suggest that the appropriate method to examine the equivalence of performance ratings across sources is to specify a MSF model composed of both source and dimension latent factors. The previously discussed three performance dimension and three source latent factor model (Model 5) formed the base model for the assessment of measurement equivalence.

The assessment of equivalence proceeded in three hierarchical steps following the recommendations of Vandenberghe and Lance (2000). That is, the assessment of equivalence represents a parameter-nested sequence in which models are hierarchically nested from the most restricted (error variance invariance or parallel), to the next most restricted (metric invariance or tau-equivalence), to the least restricted (configural invariance or congeneric). For a measure to be considered equivalent across sources, both configural and metric invariance must be demonstrated (Cheung & Rensvold, 1999; Vandenberg & Lance, 2000). Importantly, error variance invariance is not necessary to conclude that a measure is equivalent across populations. To determine whether ratings made on the MSF instrument are equivalent across sources, a difference in chi-squared test is examined in conjunction with other fit indices. In such analyses, it is preferable to accept the most restricted model (the model with the largest degrees of freedom) that does not result in a significant reduction in fit over less restricted models (Bollen, 1989). The results of the measurement equivalence analyses are presented in Table 2.

First, a model specifying configural invariance (e.g., the congeneric model) was tested. This model may be interpreted as 1) ratings of the same performance dimensions (across sources) measure the same performance dimension, 2) the loadings of these scales may be of different magnitudes, and 3) the unique variance of the ratings may also be of different magnitudes (Woehr et al., 2005). Support for this model would indicate that the ratings provided from different sources represent the same factor structure. In effect, this model is exactly the same as the previously examined three source/three performance dimension model. Again, this model provided an acceptable fit with the data (χ^2 = 1548.4; ECVI = 4.9; RMSEA = .08; NNFI = .87; CFI = .89). Thus, the MSF instrument is configurally invariant, or congeneric, across raters from different organizational levels.

In that the configural invariance model was supported, the metric invariance model was subsequently tested. In the metric invariance model (e.g., tau-equivalent model), loadings of the scales on the same factor are constrained to be equal across sources. This model specifies that 1) ratings of the same dimension across sources measure the same underlying performance dimension, 2) the loadings of the manifest indicators of each of the three latent dimension factors are of equal magnitudes across measurement sources, and 3) the uniqueness components are of potentially equal magnitudes (Woehr et al., 2005). To test this model, the factor loading of the same subscale across rating sources were set equal to one another (e.g., idealized influence for managers, peers, and subordinates). Results of these analyses indicated that this model provides an acceptable fit with the data (χ^2 = 1646.07; ECVI = 5.05; RMSEA = .08; NNFI = .87; CFI = .88). The χ^2 test for a significant difference in χ^2 between this model

and the configural invariance model was significant ($\Delta\chi^2 = 97.67$; $p < .01$). Because the difference in χ^2 test can indicate significant results with large sample sizes, despite a relatively small degree of misfit, the remaining fit indices were examined. An inspection of the remaining fit indices presented in Table 2 evidences practically no change in model fit between this model and the configural invariance model. Taken together, the results of these analyses indicate that the MSF instrument is tau-equivalent across rating sources.

Because both the configural and metric invariance models were supported (e.g., the performance model is tau-equivelant), a third model specifying equal error variance across ratings sources was tested (the parallel model). This model specifies that 1) ratings of the same dimension across rating sources measure a common performance dimension, 2) the loadings of manifest indicators of ratings provided by different sources are of equal magnitude with respect to the performance dimensions being rated, and 3) the unique variance of the ratings are of equal magnitudes (Woehr et al., 2005). This model is the most restrictive in that it restricts all of the factor loadings of a MSF subscale to be equal to one another, and the unique variance components associated with a given subscale are also constrained to be equal. Support for this model indicates that the latent factor loading of performance ratings from different sources load on the same latent factor (configural invariance), do so to an equal degree (metric invariance), and contain the same amount of error variance (error variance invariance). If this model fits, the ratings provided by different sources are considered to be parallel. Unfortunately, this model did not converge when input into LISREL version 8.5, which is usually interpreted as indicating that the model does not adequately represent a given data set. Therefore, meaningful comparisons could not be drawn between this model and the configural and metric invariance models.

To summarize, the examination of measurement equivalence indicates that the MSF instrument incorporated in the present study is tau-equivalent across sources.

In sum, a set of confirmatory factor analyses of a MTMS matrix representing the correlations between three different sources' ratings of eleven different performance dimensions were used to investigate the structure of the MSF instrument. Again, these results were consistent with previous research inasmuch as a model composed of performance dimension and rating source latent factors provided the best fit with the MSF data. and the MSF instrument was tau-equivalent across sources. In addition, source effects explained the majority of the variance in MSF, followed by uniqueness (e.g., error), and performance dimension explained the least amount of variance in MSF. Finally, a three dimension model of managerial performance consisting of conceptual/administrative, leadership, and interpersonal performance informed by Borman and Brush's taxonomy characterized the MSF data.

The Nomological Network of MSF

The final research questions seek to examine the relationship among MSF source and performance dimension latent factors and external constructs. Again, these questions will be addressed using LISREL 8.5. This model adds latent factors representing each of the external constructs (AC dimensions, personality constructs, and intelligence) to the previously supported six factor model (Model 5). First, to examine whether the AC dimensions conform to the proposed three factor model of managerial skills, the variance-covariance matrix containing the PCDRs for each of the 9 dimensions classified as similar to one of the three broad categories of managerial skills by the SMEs served as input into LISREL 8.5.

Structure of AC Dimensions

Three models potentially characterizing the AC dimensions were tested. The first model specified a one factor model where each of the AC PCDRs served as manifest indicators of a general performance dimension. Analyses indicated that this model did not provide an acceptable fit with the AC data (χ^2 = 268.65; ECVI = .86; RMSEA = .14; NNFI = .75; CFI = .80). Next, a two factor model with the PCDRs classified as conceptual/administrative skills (analysis, judgment, planning and organizing, and decisiveness) served as manifest indicators of conceptual/administrative skills and where the PCDRs classified as interpersonal (oral communication, sensitivity, and team building) and leadership skills (confrontation, persuasiveness, and leadership) served as manifest indicators of an interpersonal skills latent factor. The two dimension model specifies that managerial skills can best be described by an interpersonal skills and a conceptual/administrative skills factor. Of the three performance dimensions proposed by Borman and Brush, leadership skills and interpersonal skills are expected to overlap to the greatest extent (Conway, 1999). Thus, for the two dimension model, the manifest indicators of leadership skills and the manifest indicators of interpersonal skills are specified to load on an interpersonal skills latent factor. Results of the CFA indicate that this model provided a close fit to the AC data (χ^2 = 55.05; ECVI = .27; RMSEA = .042; NNFI = .97; CFI = .98).

Finally, the proposed three factor model of managerial skills was tested with each of the PCDRs set to load on one of the three broad dimensions of managerial skills as classified by the SMEs. Results of the CFA indicate that this model also fit the AC data relatively well (χ^2 = 49.79; ECVI = .27; RMSEA = .039; NNFI = .97; CFI = .98). In

comparing the fit of the three models tested, both the two factor and three factor AC models provide a solid fit with the AC data, with the three factor model showing slightly better fit. An examination of the correlations among latent factors indicated that AC interpersonal and AC leadership skills were strongly intercorrelated (r = .85). Still, these factors share 72% of the variance, leaving the possibility open that these are distinct constructs. Because a secondary purpose of this paper was to examine the construct validity of AC dimensions using external measures, the three factor model will be used to conceptualize AC performance in the remaining analyses. The results of the analyses evaluating the nomological network of the three latent AC dimension factors will provide evidence to further evaluate the distinctness of the interpersonal and leadership skills latent factors.

Finally, the AC dimensions and the nine paper and pencil assessed constructs were added to the previously supported six factor MSF model (Model 5). For the paper and pencil instruments, each of the constructs judged to be conceptually similar to at least one of the three performance domains served as a single manifest indicator of a latent factor in this model. Consequently, each of the single manifest indicator factors was constrained such that the factor loading (lambda x) was set to the square root of the reliability of each of the relevant scales reported by their test manuals (Gough & Bradley, 1996; Watson & Glaser, 1980). The internal consistency reliabilities for each of the relevant subscales of the paper and pencil instruments are presented in Appendix C. Descriptions of each of the CPI folk scales used in this study can be found in Appendix B. The fit indices for the full model (a model composed of latent factors representing three MSF source latent factors, three MSF performance dimension latent factors, three

AC dimension latent factors, and the nine subscales measured by the paper and pencil instruments) suggest that this model provides an adequate fit with the data (χ^2 = 2392.36; ECVI = 8.18; RMSEA = .06; NNFI = .87; CFI = .89).

Research Question 8: The Construct Validity of MSF Dimensions

To determine the relationship between each of the latent factors, the phi matrix of the full model, which represents the correlation among latent factors, was examined. The phi matrix representing the correlation matrix among latent constructs is presented in Table 5. For the research question concerned with the relationship among external constructs and latent MSF dimension factors (Research Question 8), the pattern of relationships between the three MSF latent dimension factors and the latent factors representing the external measures of conceptually similar and dissimilar individual differences was examined. For the MSF conceptual/administrative skills latent factor, none of the correlations with conceptually similar externally measured constructs (AC conceptual/administrative skills, GMA, and responsibility) was significant, indicating a lack of convergent validity evidence for the MSF conceptual/administrative skills factor. The MSF interpersonal skills latent factor was significantly related with two of the seven conceptually similar externally measured constructs including: AC interpersonal skills ($r = .21$; $p < .01$) and empathy ($r = .13$; $p < .05$). These results provide some evidence for the convergent validity of the MSF interpersonal skills factor. Finally, two of the four externally measured constructs rated as conceptually similar to leadership skills were significantly related to MSF leadership skills including: AC leadership skills ($r = .16$; $p < .05$) and social presence ($r = .15$; $p < .05$).

Table 5: Correlations Among Latent Factors

		1	2	3	4	5	6	7	8	9	10	11	12
1	MSF Leadership	1.0											
2	MSF Interpersonal	.58**	1.0										
3	MSF Conceptual	.12	.43**	1.0									
4	Manager	--	--	--	1.0								
5	Peer	--	--	--	.38**	1.0							
6	Subordinate	--	--	--	.26**	.37**	1.0						
7	AC Conceptual	.09	.02	-.08	.18**	.14*	.17**	1.0					
8	AC Interpersonal	.20*	.21**	-.08	.13*	.23**	.06	.53**	1.0				
9	AC Leadership	.16*	.06	-.10	.13*	.17**	.29**	.62**	.85**	1.0			
10	Dominance	.06	-.17**	-.19**	-.02	.00	.09	.12*	.19**	.32**	1.0		
11	Sociability	.15*	.02	-.03	-.13*	-.06	.02	.06	.17*	.17**	.65**	1.0	
12	Independence	-.06	-.24**	-.17*	-.04	.03	.04	.08	.11	.23**	.69**	.53**	1.0
13	Empathy	.12	.13*	-.08	-.04	.05	.05	.15*	.29**	.18**	.42**	.67**	33**
14	Responsibility	.03	.06	.02	.03	-.07	-.06	.15*	.20**	.15*	.37**	.24**	.23**
15	Good Impression	-.05	.05	-.06	-.01	.01	.05	-.05	.15*	.06	.09	.10	.17**
16	Tolerance	-.08	.02	.03	.07	-.02	.01	.10	.24**	.15*	.19*	.24**	.26**
17	Flexibility	.09	-.02	-.06	-.04	.02	.03	.13*	.19*	.12	.08	.30**	.33**
18	Intelligence	-.16*	-.06	.08	.05	.03	.01	.34**	.18*	.19**	.11	.09	.13*

Note. * Denotes significant at the .05 level; ** denotes significant at the .01 level.

		13	14	15	16	17	18
13	Empathy	1.0					
14	Responsibility	.28**	1.0				
15	Good Impression	.15*	.49**	1.0			
16	Tolerance	.45**	.69**	.43**	1.0		
17	Flexibility	.55**	.16*	-.04	.43**	1.0	
18	Intelligence	.20**	.23**	-.14*	.33**	.19**	1.0

To evaluate the discriminant validity evidence for the MSF performance dimension factors, the correlations between the MSF performance dimension factors and conceptually dissimilar externally measured constructs were examined. To the extent that an MSF dimension is uncorrelated with a construct judged to be conceptually dissimilar, discriminant validity evidence is provided for the MSF dimension. The MSF conceptual/administrative performance dimension was significantly related to two of the nine externally measured conceptually dissimilar constructs including: dominance ($r = -.19$; $p < .01$) and independence ($r = -.17$; $p < .05$). The MSF interpersonal performance dimension was also significantly related to two of the nine externally measured conceptually dissimilar constructs including: dominance ($r = -.17$; $p < .01$) and independence ($r = -.24$; $p < .01$). Importantly, although the MSF conceptual/administrative and MSF interpersonal factors were significantly related to conceptually dissimilar constructs, the relationships between the external constructs and these MSF dimensions were negative. Implications of these findings will be further elucidated in the discussion section. Finally, discriminant validity evidence was provided for the MSF leadership performance dimension in that it was significantly related to two of the nine conceptually dissimilar constructs. Specifically, of the nine conceptually dissimilar constructs, AC interpersonal skills ($r = .20$; $p < .01$) and intelligence ($r = -.16$; $p < .05$) were significantly related to the MSF leadership latent factor. In general, analyses examining the discriminant validity of the three MSF dimensions provided support for the discriminant validity of the MSF interpersonal, conceptual/administrative, and leadership performance dimension factors by demonstrating that each were generally unrelated to conceptually dissimilar externally measured constructs.

To further evaluate the construct validity evidence for the MSF performance dimension factors, the mean correlations between the MSF dimensions and conceptually similar external constructs was compared to the mean correlation between MSF dimensions and conceptually dissimilar external constructs. The approach recommended by Meng, Rosenthal, and Rubin (1992) for testing significant differences in correlated correlations was used to determine if the correlations between conceptually similar constructs differed from the correlation between conceptually dissimilar constructs. In essence, the method recommended by Meng and his colleagues corrects the weaknesses of traditional methods. Specifically, Hotelling's t (Hotelling, 1940) was explicitly designed to compare correlations with fixed regressors, as in an experimental design. According to Meng et al., when using random regressors, Hotelling's method specifies the null hypothesis inappropriately and as such, provides biased significance results (Meng et al., 1992). Meng et al.'s approach corrects for this limitation of previous methods of comparing correlated correlations. Because the regressors in this study are random, Meng and his colleagues' approach of testing for significant differences in correlated correlations was used.

To determine whether the correlations differ significantly, the correlations between each of the MSF dimensions and conceptually similar constructs and each of the MSF dimensions and conceptually dissimilar constructs were averaged using Fishers r to z transformation (r'). Then, the average correlations between each MSF dimension and conceptually similar and dissimilar constructs were compared to determine whether they differed significantly using Meng et al.'s approach. Table 6 presents the results of tests for significant differences in the correlations between MSF performance dimensions

Table 6: Correlations Among MSF Dimensions and External Constructs

External Constructs	MSF Latent Performance Dimension Factor		
	MSF Conceptual	MSF Interpersonal	MSF Leadership
AC Conceptual	**-.08**	.02	.09
AC Interpersonal	-.08	**.21****	.20*
AC Leadership	-.10	.06	**.16***
Dominance	-.19**	-.17**	**.06**
Independence	-.17*	-.24**	**-.06**
Sociability	-.03	**.02**	**.15***
Empathy	-.08	**.13***	.12
Flexibility	-.06	**-.02**	.09
Good Impression	-.06	**.05**	-.05
Tolerance	.03	**.02**	-.08
Responsibility	**.02**	**.06**	.03
Intelligence	**.08**	-.06	-.16*
All constructs			
Mean r among all conceptually similar	.01	.07	.08
Mean r among all conceptually dissimilar	-.08	-.08	.03
z-value for significant difference between similar and dissimilar	1.35	2.33**	.76
AC Dimensions			
Mean r among conceptually similar	-.08	.21	.16
Mean r among conceptually dissimilar	-.09	.04	.15
z-value for significant difference between similar and dissimilar	.21	4.41**	.39

Note. * denotes $p < .05$; ** denotes $p < .01$, one-tailed; Correlations in bold denote constructs judged to be conceptually similar.

with conceptually similar and dissimilar externally measured constructs. Because it is expected that the relationship between MSF dimensions and conceptually similar constructs will be greater than that between MSF dimensions and dissimilar constructs, a one-tailed significance test ($z = 1.65$ as a critical value for significance at the .05 level) was used in this set of analyses.

The correlation between the MSF conceptual/administrative performance dimension and conceptually similar constructs (mean $r = .01$) was not significantly different than the correlation between the MSF conceptual/administrative performance dimension and conceptually dissimilar constructs (mean $r = -.08$; $z = 1.35$), indicating weak evidence for the construct validity of this latent dimension factor. Next, the mean correlation between MSF interpersonal performance and conceptually similar (mean $r = .07$) externally measured constructs was significantly greater than the mean correlation between MSF interpersonal performance and conceptually dissimilar constructs (mean $r = -.08$; $z = 2.33$). However, the average correlation between MSF interpersonal skills and conceptually similar constructs was nonsignificant, making the significant difference between correlations with conceptually similar and dissimilar constructs meaningless (Cohen et al., 2002). Finally, the correlation between MSF leadership performance and conceptually similar (mean $r = .08$) constructs was not significantly different from the correlation between MSF leadership performance and conceptually dissimilar constructs (mean $r = .03$; $z = .76$), indicating weak evidence for the construct validity of the MSF leadership performance dimension factor. By demonstrating that the mean correlation between MSF latent dimension factors and conceptually similar constructs was not significantly different from the mean correlation between MSF latent dimension factors

and conceptually dissimilar constructs, this set of analyses indicates poor construct related validity evidence for the MSF performance dimension latent factors.

One of the strengths of this study was the incorporation of multiple measurement methods to assess the external constructs. Of the eleven paper and pencil constructs, only two were significantly correlated with the corresponding MSF dimension. However, two of the three AC dimensions were significantly correlated with the corresponding MSF performance dimension factor. Because stronger support was provided for the construct validity of the MSF dimensions by the AC dimension factors, the relationships between the MSF and AC dimensions were examined separately. Just as before, the correlation between each MSF dimension and the corresponding AC dimension factor was compared to the mean correlation between each MSF dimension and the other two AC dimension factors. The previously discussed method of comparing correlated correlations forwarded by Meng and his colleagues will also be used in this set of analyses. The results of these analyses are also presented in Table 6. The mean correlations between neither MSF conceptual/administrative skills nor MSF leadership skills and conceptually similar AC latent factors (mean r = -.08 and .16 respectively) was significantly different from their correlation with conceptually dissimilar AC dimensions (mean r = -.09, z = .21 and .15, z = .39, respectively). However, the correlation between MSF interpersonal performance and AC interpersonal skills (r = .21) was significantly greater than the mean correlation between MSF interpersonal performance and conceptually dissimilar AC skill dimension factors (mean r = .04; z = 4.41), providing support for the construct validity of the MSF interpersonal performance dimension factor.

In sum, the correlations between MSF performance dimension factors and conceptually similar and dissimilar externally measured constructs were examined in order to evaluate the convergent and discriminant validity evidence for the MSF dimensions. Results indicated that MSF dimensions were significantly correlated with 4 of the 14 conceptually similar externally measured constructs and 6 of the 14 conceptually dissimilar externally measured constructs. Finally, the correlation between MSF dimensions and conceptually similar constructs was greater than that with conceptually dissimilar constructs for the MSF interpersonal factor but not for the MSF conceptual/administrative or MSF leadership latent factors. Together, these results provide relatively weak evidence for the construct validity of the MSF dimensions.

Research Question 9: The Meaning of MSF Source Factors

An examination of the pattern of relationships between external constructs and each of the three latent source factors formed the basis for evaluating Research Question 9. Specifically, the phi matrix of latent factor correlations was used to draw inferences regarding the extent to which source effects represent SSPRV or error. The manager latent source factor was significantly related to one of the externally measured constructs indicative of conceptual/administrative skills (for AC conceptual/administrative skills, $r = .18$; $p < .01$), two of the constructs indicative of interpersonal skills (for AC interpersonal skills, $r = .13$, $p < .05$; for social presence, $r = -.13$, $p < .05$), and one of the constructs indicative of leadership skills (for AC leadership skills, $r = .13$, $p < .05$). The peer latent source factor was significantly related to one of the constructs indicative of conceptual/administrative skills (for AC conceptual/administrative skills, $r = .14$; $p < .05$), one construct indicative of interpersonal skills (for AC interpersonal skills, $r = .23$; p

< .01), and one construct indicative of leadership skills (for AC leadership skills, r =.17, p < .01). Finally, the subordinate latent source factor was significantly related to one of the constructs indicative of conceptual skills (for AC conceptual skills, r = .17, p < .01), none of the constructs indicative of interpersonal skills, and one of the constructs indicative of leadership skills (for AC leadership skills, r = .29, p < .01). These significant correlations between MSF source factors and externally measured constructs indicate that MSF source factors do not simply reflect error. However, these results do not provide an indication of the extent to which the performance relevant variance represented in source factors is *source specific*.

To determine whether the performance relevant variance represented in the source factors is source specific, the next set of analyses examined the extent to which a given externally measured construct was more strongly related to one source factor than the other two latent source factors. For these analyses, only externally measured constructs that were significantly related to at least one of the three source factors were included. If a construct (e.g., dominance) is not correlated with any of the three latent source factors, it makes little sense to ask whether the correlation between that construct and the source factors differ significantly (Cohen et al., 2002). The only four externally measured constructs that exhibited significant correlations with the latent source factors were: AC conceptual/administrative skills, AC interpersonal skills, AC leadership skills, and social presence. The correlations between each of these constructs and each of the three MSF latent source factors were compared using the approach for comparing correlated correlations recommended by Meng and his colleagues (1992). Because there is little reason to believe that source effects will be more strongly related to one set of external

constructs than the others, a two-tailed significance test (z = 1.96 as a critical value for significance at the .05 level) was used in this set of analyses.

Results indicated that the correlation between AC conceptual/administrative skills and the manager and peer source factors (r = .18 and .14, respectively; z = .67), manager and subordinate source factors (r = .18 and .17, respectively; z = .15), and peer and subordinate source factors (.14 and .17, respectively; z = .50) did not differ significantly. In contrast, the AC interpersonal skills factor was significantly more strongly related to the peer source factor (r = .23) than to either the manager (r = .13; z = 1.76) or the subordinate (r = .06; z = 2.90) source factors. However, the correlation between AC interpersonal skills and manager (r = .13) and subordinate (r = .06) source factors did not differ significantly (z = 1.10). In sum, the peer source factor was more strongly related to AC interpersonal skills than were the subordinate and supervisor source factors. The AC leadership skills factor was more strongly related to the subordinate (r = .29) source factor than to the peer (r = .17; z = 2.10) or manager (r = .13; z = 2.57) source factors. However, the correlations between the AC leadership factor and peer and manager latent source factors did not differ significantly (z = .69). These results suggest that the subordinate source factor represents leadership skills to a greater extent than do peer and supervisor source factors. Finally, social presence was more strongly related to the manager (r = -.13) source factor than to the subordinate source factor (r = .02; z = 2.52). However, the difference in correlation between social presence and the manager (r = -.13) and peer (r = -.06) source factors, and the correlation between social presence and the peer (-.06) and subordinate (r = .02) source factors was not significant (z = 1.10 and .69, respectively).

This set of analyses attempted to evaluate the extent to which MSF source effects represent SSPRV versus error. These inferences were facilitated by examining the pattern of correlations between MSF latent source factors and externally measured constructs. First, each of the three source factors was significantly correlated with externally measured constructs, providing evidence that source effects represent performance relevant variance. Interestingly, the AC dimension factors were each related to the latent source factors, while the paper and pencil constructs were relatively unrelated to the source factors. Finally, the AC interpersonal skills factor was more strongly related to the peer latent source factor, and the AC leadership skills factor was significantly more strongly related to the subordinate source factor than to the other two source factors. These findings indicate that the meaning of MSF source effects differs across sources.

Construct Validity of AC Dimensions

Finally, in order to "triangulate" construct-related validity evidence, the correlation (based on the phi matrix) between the three AC skill dimension factors and conceptually similar and dissimilar constructs measured using paper and pencil instruments was examined. These results were used to evaluate the construct-related validity evidence of the three AC skill dimensions factors. Again, the t-value provided by Lisrel 8.5 was used to determine whether the relationships among AC skill dimension factors and external constructs were significant, and the approach recommended by Meng and his colleagues for comparing correlated correlations was used to determine whether the relationship among the latent factors differ significantly.

For the two constructs conceptually similar to AC conceptual/administrative skills, both responsibility ($r = .15$; $p < .05$) and GMA ($r = .34$; $p < .01$) were significantly

related to the AC conceptual/administrative skills dimension factor. All six of the constructs conceptually similar to AC interpersonal skills were significantly correlated with the AC interpersonal skills latent factor including: social presence ($r = .17$; $p < .05$), empathy ($r = .29$; $p < .01$), responsibility ($r = .20$; $p < .01$), good impression ($r = .15$; $p < .01$), tolerance ($r = .24$; $p < .01$), and flexibility ($r = .19$; $p < .05$). Finally, all three of the correlations between the constructs conceptually similar to AC leadership skills were significant. Specifically, dominance ($r = .32$; $p < .01$), independence ($r = .23$; $p < .01$), and social presence ($r = .17$; $p < .01$) were significantly related to the AC leadership skill dimension factor. Together, these results provide solid evidence for the convergent validity of the AC conceptual/administrative, interpersonal, and leadership dimension factors.

The discriminant validity evidence for the AC skill dimension factors was evaluated by examining the correlations between the three AC skill dimension factors and conceptually dissimilar constructs. Of the seven external constructs conceptually dissimilar to AC conceptual/administrative skills, two were significantly correlated with AC conceptual/administrative skills including: empathy ($r = .15$; $p < .05$) and flexibility ($r = .13$; $p < .05$). Of the three constructs conceptually dissimilar to AC interpersonal skills, two were significantly related to the AC interpersonal skills dimension factor including: dominance ($r = 19$; $p < .01$) and GMA ($r = .18$; $p < .05$). Finally, of the six external constructs conceptually dissimilar to AC leadership skills, four were significantly correlated with the AC leadership skills latent factor including: empathy ($r = .18$; $p < .01$), responsibility ($.15$; $p < .05$), tolerance ($r = .15$; $p < .05$), and GMA ($r = .19$;

$p < .01$). Together, these results provide partial support for the discriminant validity of AC interpersonal, leadership, and conceptual/administrative skill dimension factors.

Next, the mean correlation between each AC dimension factor and conceptually similar externally measured constructs was compared to the average correlation between that AC dimension factor and conceptually dissimilar constructs using the previously discussed method of comparing correlated correlations (Meng et al., 1992). Because it is expected that the relationship between AC dimensions and conceptually similar constructs will be greater than that between AC dimensions and dissimilar constructs, a one-tailed significance test ($z = 1.65$ as a critical value for significance at the .05 level) was used in this set of analyses. The results of this set of analyses are presented in Table 7. The mean correlation between AC conceptual/administrative skills and conceptually similar constructs (mean $r = .25$) was significantly greater than the mean correlation between AC conceptual/administrative skills and conceptually dissimilar constructs (mean $r = .08$; $z = 2.67$). However, the mean correlation between AC interpersonal skills and conceptually similar constructs (mean $r = .21$) did not differ significantly from the mean correlation between AC interpersonal skills and conceptually dissimilar constructs (mean $r = .16$; $z = .80$). Finally, the mean correlation between AC leadership skills and conceptually similar constructs (mean $r = .25$) was significantly greater than the mean correlation between AC leadership skills and conceptually dissimilar constructs (mean $r = .14$; $z = 1.75$). Together, these results provide solid construct validity evidence for AC conceptual/administrative and AC leadership skills factors and moderate evidence for the AC interpersonal skills factor.

Table 7: Correlations Among AC Dimensions and External Constructs

External Construct	Assessment Center Dimension		
	AC Conceptual	AC Interpersonal	AC Leadership
Dominance	.12*	.19**	**.32****
Independence	.08	.11	**.23****
Social Presence	.06	**.18***	**.17****
Empathy	.15*	**.29****	.18**
Flexibility	.13*	**.19***	.12
Good Impression	-.05	**.16***	.06
Tolerance	.10	**.24****	.15*
Responsibility	**.15***	**.20****	.15*
Intelligence	**.34****	.18*	.19**
Mean r Among Conceptually Similar	.25**	.21**	.25**
Mean r among Conceptually Dissimilar	.08	.16**	.14**
z value for significant difference in r	2.67**	.80	1.75*

Note. * denotes $p < .05$; ** denotes $p < .01$; Correlations in bold denote constructs judged to be conceptually similar.

In an effort to "triangulate" construct validity evidence, the covariance between AC dimensions and conceptually similar and dissimilar paper and pencil constructs was examined. The results provided strong support for the convergent validity of the AC dimensions (all 11 of the correlations between AC dimensions and conceptually similar constructs were significantly) and moderate support for the discriminant validity of the AC dimensions (9 of the 16 correlations with conceptually dissimilar constructs were significant). Finally, AC leadership skills and the AC conceptual/administrative skills were significantly more strongly related to conceptually similar constructs than to conceptually dissimilar constructs, providing solid evidence for the construct validity for these two AC dimension factors. As will be discussed in greater detail in the following section, the AC dimensions evidenced stronger construct validity than did the MSF dimensions. A summary of the construct validity evidence found for the MSF and AC dimensions is presented in Table 8.

Table 8. Summary of Construct Validity Evidence for MSF and AC Dimensions.

	MSF Latent Dimensions			AC Latent Dimensions		
Summary Question	MSF Concep.	MSF Interp.	MSF Lead.	AC Concep.	AC Interp.	AC Lead.
How many significant r's with similar?	0/3	2/7	2/4	2/2	6/6	3/3
How many significant r's with dissimilar?	2/9	2/9	2/9	2/7	2/4	4/6
Significant difference in r with similar and dissimilar?	No	No	No	Yes	No	Yes

Chapter 5

Discussion

Summary of Results

Over the past two decades, MSF has been an increasingly popular method of diagnosing the strengths and weaknesses of managers. Typically, the feedback derived from MSF instruments provides information to managers that is subsequently used to inform skill development. Despite the prevalence of MSF as a developmental tool, there is a dearth of empirical research examining the meaning of MSF source and dimension effects. Given the prominence of MSF as a developmental tool, an accurate assessment of the quality of the skill diagnoses derived from MSF is crucial to the appropriate application of MSF.

To better understand what is actually being measured by MSF tools, this study incorporated both internal and external approaches to construct validation. First, this study sought to replicate previous MSF research focusing on the internal approaches to construct validation. Consistent with previous MSF research, results indicated that: 1) the agreement of raters from the same organizational level is greater than agreement between raters from different organizational levels, 2) MSF ratings are best described with a model consisting of rating source and performance dimension latent factors, and 3) MSF dimension ratings are tau-equivalent across sources.

Next, this study sought to extend previous research be exploring the relationship between MSF source and dimension latent factors and externally measured constructs.

Based on the correlations between MSF performance dimension factors and conceptually similar and dissimilar external constructs, somewhat weak support was provided for the construct validity of MSF dimensions. In addition, MSF latent source factors differentially correlated with some of the externally measured constructs, providing evidence that MSF source factors represent source specific, systematic variance relating to external measures of managerial skills. Finally, the results of this study provided support for the construct validity of AC dimension by demonstrating that two of the three AC skill dimension factors correlated more strongly with conceptually similar externally measured constructs than with conceptually dissimilar externally measured constructs. In the following sections, these results will be discussed in the context of three questions: 1) Does the construct validity evidence provided by the internal approaches replicate previous research? 2) What is the construct validity evidence for the MSF dimensions? and 3) What do MSF latent source factors represent? In answering these questions, a particular focus will be placed on implications of this study and avenues for future research. Finally, limitations of this study will be discussed.

Internal Approaches

Rater Agreement

The preponderance of MSF research to date has used internal approaches to construct validity to draw inferences with the respect to the psychometric properties of MSF instruments. That is, previous MSF research has typically examined the pattern of covariance among different sources' ratings to ascertain the construct validity of MSF. One of the objectives of this study was to replicate previous MSF research using the internal approaches by incorporating rater agreement, confirmatory factor analysis, and

measurement equivalence as methods of evaluating the psychometric properties of the present MSF instrument.

The level of agreement within and across sources has been one of the primary methods of examining the psychometric properties of MSF. Previous meta-analytic research suggests that the correlation between ratings made by raters from the same organizational level is moderate (mean r = .40) and is stronger than the correlation between ratings by raters from different organizational levels (mean r = .22; Conway & Huffcutt, 1997). The results of the present study were quite consistent with previous agreement research with respect to both the magnitude and the pattern of agreement estimates and provides initial evidence for the presence of source effects in the present sample. As previously discussed, some previous research has indicated that across source agreement is impacted by the performance dimension being rated (Conway & Huffcutt, 1997) while other research has not (Viswesvaran et al., 2002). The present results indicate that the performance dimension being rated does not have a significant impact on rating covariation for any of the possible rating pairs (e.g., peer-subordinate, peer-supervisor, and supervisor-subordinate). Stated differently, regardless of the performance dimension being rated, the level of correspondence in ratings from different sources was relatively consistent.

However, the level of agreement did differ depending on the rating pair. Specifically, peer-subordinate and peer-supervisor rating agreement was greater than subordinate-supervisor agreement. Interestingly, these results diverge from previous research. In contrast to the results of this study, the meta-analysis by Conway and Huffcutt demonstrated that although peer and supervisors agreed to a greater extent than

other rating pairs, the agreement between peer-subordinate ratings and subordinate-supervisor ratings was equitable. Despite the divergence from prior research, the findings of this study are intuitively appealing. In terms of rater hierarchical level, a target's peers are directly between subordinates and supervisors. As such, peer raters may have the opportunity to observe a broader range of behavior or have performance expectations that are congruent with both supervisors and subordinates. In contrast, a target's supervisors and subordinates are further removed in terms of hierarchical level and as a result, may have qualitatively different expectations of or interactions with the target. In fact, it is plausible that a target manager's subordinates and supervisors have never been in the same room together. To summarize, these results of analyses investigating the correspondence of ratings provided by different raters closely matched the findings of previous agreement research with respect to both the magnitude and the pattern of within and across source correlations.

Factor Structure of MSF

Examination of the factor structure of performance ratings from multiple sources supported a performance model consisting of three source and three dimension factors. These results provide empirical support for the three factor model of managerial performance consisting of conceptual/administrative, interpersonal, and leadership skills proposed by Borman and Brush. In addition, these findings are consistent with previous MSF research which has indicated that both source and dimension effects characterize MSF data. In essence, support for this model suggests that ratings of the same performance dimension taken from different sources converge to some degree (dimension

effects) and that each source's ratings can be characterized by an overall or general impression factor (source effect).

It should be noted that the results of the confirmatory factor analyses for the structure of multisource ratings (MSR) indicated that the six factor model provided only an adequate fit with the data. To this end, the loadings of manifest indicators on the conceptual/administrative skills latent factor were quite small (mean = .15). In fact, subordinate ratings of both analysis and planning and organizing did not significantly load on the conceptual/administrative skills factor. This suggests that the conceptual/administrative skills factor is not well defined, particularly by subordinate raters. Interestingly, previous meta-analytic research suggests that cross source ratings of administrative competence converge to a lesser extent across sources than other performance domains such as leadership (Viswesvaran, Schmidt, & Ones, 2002). Typically, administrative/conceptual skills are more difficult to observe than are leadership or interpersonal skills. To illustrate, it is far more difficult to observe an individual integrating information (characteristic of administrative/conceptual skills) than building rapport with others (characteristic of interpersonal skills). In fact, many of the behaviors associated with conceptual/administrative skills are either done mentally (e.g., integrating information) or can be performed in the privacy of one's office (e.g., planning for upcoming events). Together, the degree of convergence for the indicators of the conceptual/administrative skills latent factor is consistent with the conclusion of prior conceptual and empirical research.

Despite the six factor model achieving only an adequate level of fit with the data and problematic loadings on the conceptual/administrative skills factor, this model was

retained for subsequent analyses. The loading of manifest indicators on each of the six latent factors was based solely on conceptual grounds and this model provided the best fit with the data of the models tested. Moreover, previous MSF research consistently supports a model consisting of both source and dimension latent factors. Consequently, the six factor model was viewed as the most appropriate of those tested and incorporated into subsequent analyses. Still, future research should focus on replicating the six factor model of managerial MSF supported here.

The relative proportion of variance explained by performance dimension (11%), rating source (56%), and uniqueness (33%) differed from that reported in previous research using a similar methodology. Woehr et al. (2005) demonstrated that performance dimension latent factors explained 30% of the variance in performance ratings, source latent factors explained 25% of the variance in MSF ratings, and uniqueness explained 41% of the variance in MSF ratings. The most pronounced difference in the study by Woehr and his colleagues and the present study is the proportion of variance explained by MSF dimension and source effects. Again, dimensions effects explained much more variance in the study by Woehr and his colleagues than in the present study (11% and 30%, respectively). In contrast, source effects explained more variance in performance ratings in the present study (56%) than in the study by Woehr and his colleagues (25%).

An inspection of the performance rating methodologies/contexts across the two studies may shed some light on the reason for this difference. First, the sample in the study by Woehr and his colleagues consisted of U.S. Air Force technicians (e.g., mechanics), a set of jobs with relatively concrete requirements in comparison to the

requirements of managerial work. Previous work performance research has suggested that performance dimensions are more accurately rated in situations with concrete or technical job requirements (such as mechanics) as compared to jobs with more nebulous or interpersonal requirements (such as managers; Viswesvaran et al., 2002). Thus, it is not surprising that the performance dimension effect accounted for a greater amount of the variance in ratings in the study by Woehr and his colleagues as compared to the present study.

In addition to differences in job characteristics across the two studies, the rating context varied in one important respect. That is, all raters in the study by Woehr and his colleagues participated in frame of reference training (FORT) prior to providing performance ratings, whereas raters in the present study did not participate in rater training. FORT is designed to provide raters with the same frame of reference of behaviors indicative of effective and ineffective performance on a given performance dimension. To this end, previous research supports the use of FORT as a method to increase the differentiation among performance dimensions and subsequent rating accuracy (Woehr & Huffcutt, 1994). Given the difference in rating methods across the two studies, the difference in relative proportion of variance explained by dimension and source variance is not surprising. That is, because raters in the study by Woehr and his colleagues received FORT, it would be expected that dimension effects would explain more variance in ratings than in a rating context where raters did not receive training. In fact, Woehr and his colleagues asserted that the rating context in their study represented a "best case scenario" for performance rating and that their results might not be replicated by other research conducted under more typical rating conditions. Still, a direct

examination of the impact of FORT on the magnitude of MSF source and dimension effects would be an interesting area for future research.

The finding that MSF ratings were equivalent across rating sources is consistent with previous MSF research. However, with one notable exception, previous research has examined equivalence by specifying separate models for each source providing ratings. The approach used here was consistent with the approach recommended by Woehr and his colleagues (2005) where the same dimension rated by different sources is set to load on the same latent factor. Again, the findings that performance ratings were metric invariant (e.g., tau equivalent) across sources are consistent with the findings of Woehr and his colleagues. Together, this suggests that performance information provided by raters from different levels can be compared. It is worthwhile to note that the χ^2 test suggested a significant difference between the configural invariance and metric invariance models. As previously mentioned the χ^2 test is sensitive to large sample sizes and as a result finds significant results with relatively small degree of misfit. The other fit indices were quite similar, indicating relatively minor differences between the two models. Thus, the results support the tau-equivalence of ratings provided by manager, peer, and subordinate raters.

In sum, the construct validity evidence yielded for the MSF instrument by the internal approaches generally replicated the results of prior research. The consistency between the present findings and the results of existing MSF research provides evidence that the present MSF instrument functions similarly to other MSF instruments. This issue is particularly important as the present MSF instrument was designed in-house for use in the LDP program. Thus, the question can be raised as to if the results of this study are

only pertinent to this specific MSF system. In that the results regarding rater agreement, MSF factor structure, and measurement equivalence yielded results consistent with prior MSF research, some confidence can be given to the similarity of the present MSF instrument to other MSF instruments and the subsequent generalizability of the present findings.

External Approaches

In order to evaluate the construct validity evidence of the source and dimension effects ubiquitous to MSF, this study examined the relationships between these respective factors and externally measured constructs. In addition, multiple methods of assessing external constructs were incorporated in order to "triangulate" construct validity evidence. In this section, the evidence provided for the construct validity of MSF dimensions, AC dimensions, and the meaning of MSF source effects will be discussed. A summary of this evidence was previously presented in Table 8.

Construct Validity Evidence for MSF Dimensions

As can be seen in Table 8, somewhat weak construct validity evidence was provided for the MSF dimensions. Four of the fourteen correlations between MSF dimensions and conceptually similar externally measured constructs (empathy and AC interpersonal skills with MSF interpersonal skills; social presence and AC leadership skills with MSF leadership skills) was significant. Of the paper and pencil assessed constructs, empathy appears to be the closet conceptual match to interpersonal skills and social presence is also closely linked to leadership skills (potentially only second to dominance). That these relationships were significant provides some evidence for the convergent validity of the MSF dimensions; however, that the other externally measured

conceptually similar constructs were unrelated to the specified MSF dimensions suggests problematic evidence for the convergent validity of the MSF dimensions.

Interestingly, the AC skill dimensions converged with conceptually similar MSF performance dimensions to a greater extent than did conceptually similar paper and pencil constructs. Specifically, two of the three AC dimension factors were significantly correlated with the corresponding MSF dimension, whereas only two of the eleven paper and pencil constructs were significantly related to the conceptually similar MSF dimension. The MSF conceptual/administrative skills latent factor had a particularly poor showing, as it did not significantly correlate with any of the externally measured conceptually similar constructs. Perhaps these findings should not be surprising given the small factor loadings associated with the MSF conceptual/administrative performance dimension factor. Together, these results provide somewhat weak convergent validity evidence for the interpersonal and leadership dimensions and no convergent validity evidence for the MSF conceptual/administrative skills dimension.

In terms of discriminant validity evidence, each of the three MSF dimensions was significantly correlated with two conceptually dissimilar constructs. However, five of the six correlations between MSF dimensions and conceptually dissimilar external constructs were negative. First, dominance and independence were negatively related to both MSF interpersonal skills and MSF conceptual/administrative skills. Of course, that these significant relationships were negative is not necessarily indicative of poor discriminant validity evidence. As it was defined here, conceptual dissimilarity can encapsulate either no relationship or possibly a negative relationship between constructs. To this end, although not anticipated, the inverse relationship between independence and dominance

and MSF interpersonal skills makes some sense. The CPI Manual (Gough & Bradley, 1996; p. 12) describes individuals *high* in independence as someone who is "....detached....persistent in seeking goals, whether others agree or not" and those low in independence as someone who "seeks support from others [and] tries to avoid conflict." Based on these descriptions, it is quite easy to see how highly independent managers (e.g., detached) may also be viewed as lacking in interpersonal skills by their coworkers, and those with low levels of independence (e.g., seeks support from others) as having high levels of interpersonal skills. Similarly, Gough and Bradley noted that someone who is very high in dominance can be seen a "domineering" and "overly controlling." Again, it is not unlikely that the coworkers of very dominant managers will be alienated by the individual's exertion of control, particularly in the present context where the participants are highly successful executives, as are their coworkers. Specifically, in a group of highly successful people, an overly controlling person may be disliked or seen as lacking in interpersonal skills.

Less clear is why dominance and independence would be negatively related to MSF conceptual/administrative skills. As previously discussed, these unexpected results are possibly attributable to problems with the measurement of conceptual/administrative skills. Similarly, there is little reason to expect that critical thinking skills would be negatively related to coworker ratings of leadership, given that previous research typically demonstrates a positive relationship between intelligence and leader effectiveness (Judge, Colbert, & Ilies, 2004). However, this research has almost exclusively conceptualized leader effectiveness using overall performance as a criteria. Consequently, observed correlations between intelligence and ratings of effectiveness

may be due to the correlation between intelligence and other dimension of managerial performance, such as conceptual/administrative skills that are implicitly included in ratings of overall manager effectiveness. To this end, Bass (2001) noted that counterintuitively, GMA is often unrelated to ratings of transformational leadership. Because limited research has examined the relationship between intelligence and actual ratings of leadership, it is difficult to ascertain the efficacy of these findings.

A note is in order regarding the discriminant validity evidence provided by examining correlations with conceptually dissimilar constructs. That is, in the absence of convergent validity evidence, discriminant validity evidence is meaningless (Cohen et al., 2002). To illustrate, that MSF conceptual/administrative skills did not correlate with any of the external constructs, cannot be interpreted as "positive results" for the construct validity of this MSF dimension. In other words, if a construct is related to neither what it should be nor to what it shouldn't be, no firm conclusions can be drawn with respect to the construct's meaning. Thus, because conceptual/administrative skills was unrelated to any of the conceptually similar paper and pencil constructs, the failure to correlate with conceptually dissimilar constructs does not provide interpretable discriminant validity evidence for the MSF conceptual/administrative dimension. Similarly, MSF leadership and interpersonal skills were each only significantly related to one of the paper and pencil measured constructs. Here again, that the MSF dimensions did not correlate with paper and pencil assessed external constructs should not be taken as a positive indication of the discriminant validity of the MSF dimensions.

It should be noted that the AC dimensions fared somewhat better than did the paper and pencil constructs in terms of convergence with conceptually similar constructs.

Because the AC dimensions correlated with the MSF dimensions to a greater extent (two of three correlations were significant) than did the paper and pencil constructs (two of eleven correlations were significant), the AC dimensions were examined separately. Although convergent validity evidence was provided for MSF interpersonal and leadership performance based on their correlations with the AC dimensions, the discriminant validity evidence for the MSF dimensions was still problematic. First, the correlation between the AC leadership factor and MSF leadership factor was not significantly greater than the correlation between the MSF leadership factor and conceptually dissimilar AC dimensions. These findings do not provide supporting discriminant validity evidence for the MSF leadership factor. Although the AC interpersonal skills factor was significantly related to the MSF leadership performance factor, AC interpersonal skills and MSF interpersonal performance were more strongly related than MSF interpersonal performance and conceptually dissimilar external constructs. Thus, in the context of correlations with the AC skill dimensions, convergent and discriminant validity evidence was provided for the MSF interpersonal performance dimension. However, little evidence was demonstrated for the construct validity of MSF conceptual/administrative performance based on correlations with AC skill dimensions. Together, although slightly more positive, the support for the construct validity of the MSF dimensions provided by the correlations with the AC dimensions is similarly weak to that based on the correlations with paper and pencil assessed constructs.

Although strong construct validity evidence was not provided for the MSF dimensions by the AC dimensions, that the AC dimensions did fare somewhat better than the paper and pencil assessed constructs is quite interesting. Similar to the MSF

performance dimensions, the AC skill dimensions were classified using Borman and Brush's taxonomy. In other words, the AC and MSF assessed functionally the same construct (at least conceptually), just with different methods. In contrast, the conceptual similarity of the paper and pencil constructs had to be judged by SMEs because these constructs were thought to be related to the managerial skills, not the same construct. Consequently, it can be argued that the AC skill dimensions represent a stronger conceptual linkage to the MSF performance dimensions than the paper and pencil constructs.

The AC dimensions measure managerial skills, whereas the paper and pencil constructs assess traits. The paper and pencil constructs represent a general tendency or preference toward a behavior, whereas the AC and MSF dimensions are composed of *ratings* of actual, observable behaviors. Traits are expected to act as antecedents to the display of a given behavior, whereas a correlation between a MSF and AC dimension would represent the consistency of behavior across situations. For example, if an individual is highly empathetic, one would certainly expect them to display behaviors indicative of this trait such as interpersonal sensitivity. However, the display of sensitivity in a simulation of the work environment such as an AC (e.g., a sample of behavior) is more closely linked to subsequent work behaviors than is a general tendency to behave in a certain way, such as a paper and pencil assessed personality construct (e.g., a sign; Wernimoont & Campbell, 1968). Together, because of the causal precedence of the AC skills on MSF performance dimensions, this pattern of results should not be surprising.

Still, previous research consistently indicates that personality constructs and cognitive ability correlates with performance ratings (e.g., Barrick & Mount, 1991; Schmidt & Hunter, 1998). Thus, the finding that personality and cognitive ability were relatively unrelated to performance ratings marks a clear divergence from the extant literature. However, differences in the present MSF process and the traditional performance appraisal context may account for the disparity in results.. For example, traditional PA systems are specific to a given job, based on the results of a job analysis, have fewer items than MSF systems, and assess fewer performance dimensions than typical MSF systems (Brutus & Derayeh, 2002; Church & Allen, 1997; Healy, Walsh, & Rose, 2003; Rogelberg & Waclawski, 2001). As is typical with MSF systems administered by external providers (e.g., EMBA programs), the MSF instrument used in the present study was based on generic managerial competencies, was not based on a job analysis, and included more items and dimensions than traditional PA instruments. As a result of these differences in rating process, it may be inappropriate to assume that the relationships between individual differences and performance will be equitable across such divergent performance ratings contexts.

In fact, the difference in performance rating methodologies across the present setting and traditional PA contexts may shed some light on the failure to support expected relationships between the MSF dimensions and external constructs. A well accepted tenant of designing a performance appraisal system is that it be based on an analysis of the specific requirements of a given job. Unfortunately, a formal job analysis was simply not feasible in the present study as the participants worked in a variety of professions, jobs, and organizations located in a variety of different geographic regions. Thus, one

potential explanation for the inconsistent findings is that the present MSF system was based on generic managerial competencies, as opposed to job specific competencies derived from a formal job analysis.

In addition, the present MSF instrument contained more items and dimensions than do traditional performance appraisal instruments. Because the goal setting literature suggests that specific goals are more likely to be attained than general goals (O'Leary, Harkcom, Jackson, & Tears 2003), MSF designers are possibly prone to include additional dimensions in an effort to capture as much specific information as possible. However, completing an overly long survey may actually have a deleterious impact on rater motivation and subsequent accuracy (Rogelberg & Waclawski, 2001). Ironically, it is possible that in an effort to maximize the utility of MSF by increasing the number of dimensions/items, MSF developers have actually decreased rater motivation and subsequent rating accuracy. In fact, the relatively large impact of rating source (representing source specific general impression) and uniqueness components (representing error) on performance rating variance, coupled with the relatively small amount of variance explained by performance dimensions is consistent with this possibility. As previously mentioned, the relatively small amount of variance explained by MSF dimensions may also be attributable to the omission of rater training. Together, it is possible that the MSF process in the present study attenuated the correlation between external constructs and the MSF dimensions.

When the externally measured constructs did correlate with conceptually similar MSF dimension factors, the magnitude of the correlations was quite small. Thus, what little evidence was provided by the pattern of correlations is further tempered by the

magnitude of the correlations. Differences in rating context are one potential explanation for the somewhat disappointing external construct validity evidence rendered for the MSF dimensions. To be certain, the limitations of the present rating context present a serious issue with the MSF instrument. However, the present MSF context is quite consistent with the methods under which MSF is typically administered (Rogelberg & Waclawski, 2001; Antonioni & Woehr, 2001), particularly when MSF is administered by an external source (e.g., consulting firms and executive education programs). Thus, although the present rating context is certainly not ideal, it represents typical practice in the field. As such, this study provides important information with respect to the construct validity of MSF dimensions, as they are currently assessed.

That performance dimension-level convergence was examined at all represents an important contribution to the MSF literature. In particular, the majority of previous research has operationalized performance using an overall performance model and rarely seeks to examine differential relationships with constructs conceptually similar and dissimilar to multi-dimensional criteria. This study represents one of few to attempt to match multiple methods of measuring external constructs to multi-dimensional criteria at a conceptual level and the first to do so using CFA so that the relative impact of external constructs on dimension and source factors could be examined. As with all primary studies, this study must be replicated to ascertain the extent to which these findings generalize to alternate rating contexts.

Clearly, the relatively weak construct validity evidence demonstrated for the MSF dimensions is quite troubling for organizational scientists interested in using MSF in developmental settings. Specifically, these results suggest that giving performance

dimension feedback to managers on the basis of this MSF tool may at best be unhelpful and at worst counterproductive to managerial development. MSF has become a staple of management development as it provides the foundation on which subsequent development plans and skill building effort is built. If the dimension level feedback presented to managers does not accurately reflect their performance in a given area, subsequent attempts to improve will be misguided.

Construct Validity Evidence for AC Dimensions

A summary of the results for the construct validity evidence of the three AC dimensions was also presented in Table 8. Each of the AC dimensions was significantly related to all of the conceptually similar external constructs, providing strong support for the convergent validity of the three AC dimensions. In contrast, the discriminant validity evidence for the AC dimensions was somewhat less pronounced. That is, both AC interpersonal and AC leadership skills were significantly related to many of the conceptually dissimilar external constructs. And, the correlation between AC interpersonal skills and conceptually similar constructs was not significantly different from the correlation between AC interpersonal skills and conceptually dissimilar constructs. Together, adequate convergent and discriminant validity evidence was provided for AC conceptual/administrative skills and AC leadership skills, yet the discriminant validity for AC interpersonal skills did not meet acceptable levels.

Interestingly, this pattern of results is consistent with the findings of previous AC research. First, previous AC research using both external and internal methodologies has indicated that moderate evidence exists for the convergent validity of AC dimensions, whereas empirical evidence for the discriminant validity of AC dimensions has been

weak (Bowler & Woehr, in press). In addition, previous research that has used an external approach to examine the construct validity evidence for AC conceptual, interpersonal, and leadership skill dimensions has found strong support for the construct validity evidence of the conceptual and leadership skill dimensions, but not for the interpersonal skill dimensions (Hoffman & Kudisch, 2002).

As previously mentioned, both the two (conceptual/administrative and a factor composed of both interpersonal and leadership skills) and three (conceptual administrative, interpersonal, and leadership skills) factor AC models fit the AC data similarly. When interpreting CFA results, in cases where two models fit the data equitably, the most parsimonious model is typically viewed as the most appropriate. In this case, the two factor model would be supported as the most appropriate of the models tested. However, an examination of the pattern of relationships between the AC dimension factors and external constructs was also used to determine the most appropriate AC model.

The evidence provided by the correlations with external constructs signals moderate evidence for the distinctness of AC interpersonal and leadership skill dimensions. Two of the three externally measured constructs judged to be conceptually similar to interpersonal skills were significantly related to leadership skills. However, previous leadership research has often included interpersonal skills as a critical component of effective leadership (Bass, 1985; Fleishman, 1953). As such, it is not surprising that the same personality constructs are correlated with each. Moreover, the correlation between conceptually similar external constructs and AC leadership was significantly greater than the correlation between these external constructs and AC

interpersonal skills. These findings suggest that AC leadership and AC interpersonal skills are strongly related yet distinct constructs. Together, although the CFA results were ambiguous, the three factor model appears to be the most appropriate based on the correlations with external constructs.

Interestingly, the construct validity evidence for the AC skill dimensions was stronger than the construct validity evidence for the MSF performance dimensions. These results were somewhat surprising given that substantial empirical attention has been directed toward examining the construct validity of the ACs, whereas the construct validity of MSF has rarely been examined. In fact, many AC researchers have deemed construct validity evidence for ACs dimensions inadequate and suggested that ACs no longer be interpreted using skill dimensions and instead, that ACs should be operationalized using an overall assessment rating or an overall exercise rating (e.g., Lance, Newbolt, Gatewood, Foster, French, & Smith, 2000). In contrast, despite the dearth of research examining the construct validity of MSF, it remains one of the most popular tools of managerial development (Carruthers, 2003; Church & Allen, 1997). However, the weak construct validity evidence shown here suggests that the typical procedure of presenting managers with performance dimension feedback derived from MSF ratings may not be useful. Based on the correlation with external constructs, AC dimensions represent a more accurate reflection of an individual's skills than do MSF dimensions. Clearly, future research must address this issue in an attempt to resolve the discrepancy between construct validity evidence and perceived usefulness of each tool.

The Meaning of MSF Source Factors

This study represents the first attempt to ascertain the meaning of MSF source effects through an examination of the relationship between MSF source effects and externally measured constructs. Analyses involving the MSF source factors demonstrated that each of the three MSF source factors significantly correlated with the externally measured constructs and that externally measured constructs differentially correlated with the peer and subordinate source factors. Together, these results suggest that 1) MSF source factors represent substantively meaningful variance, and 2) the meaning of the MSF source factor differs depending on the source.

Interestingly, similar to the MSF dimension factors, source factors typically did not correlate with the paper and pencil constructs. In fact, only one of the paper and pencil constructs correlated with a MSF source factor (social presence with the manager source factor) and this correlation was negative. However, of the nine possible correlations between AC dimension factors and MSF source factors, eight were significant (AC interpersonal skills and the subordinate source factor was not). These findings suggest that each source's overall impression is formed based on their coworker's conceptual/administrative skills, interpersonal skills, and leadership skills.

As previously mentioned, the meaning of MSF source effects has been the subject of a contentious debate in the work performance literature. Some argue that examining the correspondence between different raters' performance ratings is the appropriate method of examining the construct validity of MSF (Schmidt et al., 2000; Viswesvaran et al., 2002). In fact, these authors have suggested that MSF source effects are meaningless and that disagreement across sources can be accounted for by other issues (as opposed to

the organizational level of the target). For example, by treating within source correlations as a true score and comparing these values to across source correlations, Viswesvaran and his colleagues (2002) concluded that rating difficulty explains across source disagreement to a greater extent than rater organizational level. Based on these results, Viswesvaran and his colleagues concluded that MSF source effects are meaningless.

Others have argued that examining the covariance among different ratings of the same measure is ill-equipped to draw inferences regarding the construct validity of MSF (Arthur & Villado, in preparation; Farr, 2006; Murphy & DeShon, 2000). These authors have asserted that although examining the covariance of an instrument with itself provides useful information regarding a measure's reliability, such investigations do not provide information needed to make inferences regarding a construct's. Arthur and Villado noted that Campbell and Fiske's (1959) seminal work on the MTMM approach to examining construct validity has often been misinterpreted and misapplied. In essence these authors argue that a single construct must be measured using *maximally dissimilar* methods in order to interpret MTMM results in the context of construct validity. Stated differently, simply examining the correlation between different sources' ratings of the same construct does not constitute an adequate examination of construct validity. On a similar front, Murphy and DeShon argued that traditional interrater reliability indices collapse meaningful variance (information about a person's performance) into error and as a result underestimate reliability. Murphy and DeShon further note that collapsing substantively meaningful variance into the error term can result in inappropriate inferences as to the construct validity of performance ratings.

According to Murphy and DeShon and Arthur and Villado, examining the pattern of relationships between the constructs assessed with a given measurement instrument and externally measured constructs is a more appropriate method of evaluating construct validity. Accordingly, this study answers recent calls (Farr, 2006) for an adequate examination of MSF construct validity in an attempt to ascertain the extent to which MSF source effects are most accurately classified as error (Viswesvaran et al., 2002) or substantively meaningful variance (Arthur & Villado, in preparation; Murphy & DeShon, 2000). The MSF source factors each correlated with external measures of managerial skills. These results suggest that raters' overall impression of their coworkers reflects target skills as measured by the AC. And, previous correlations between AC skill dimensions and managerial performance ratings is attributable to the impact of managerial skills on the performance dimension being rated *as well as* raters' overall impression of their manager. In contrast to the assertions of Viswesvaran and his colleagues, these correlations signify that each rating source's overall impression represents substantively meaningful variance. Thus, by using external methods of construct validation to examine the meaning of MSF source effects, this study arrives at markedly different conclusions than those by research relying strictly on internal construct validation (Viswesvaran et al., 2002).

Although the correlations between external constructs and source factors support the performance relevance of source factors, an examination of the source specific nature of this performance relevant variance is also of importance. Differential correlations with externally measured constructs would suggest that the source factors are source specific. For example, if the AC interpersonal skills factor was significantly and equitably

correlated with all three source factors and the remaining externally measured constructs were uncorrelated with the source factors, this would indicate that although source factors reflect performance relevant variance, this shared variance is not source specific. In other words, this pattern of results would indicate that all three of the source factors reflect a general impression based on interpersonal skills.

Consequently, to determine the extent to which the source effects were truly source specific, the correlations between externally measured constructs and the three source factors were examined for significant differences. The AC conceptual/administrative skills factor was significantly correlated with each of the three latent source factors; however, none of these correlations differed significantly from one source to another. Thus, although target AC conceptual/administrative skills explain variance in each source's general impression factor, each source factor represents conceptual/administrative skills to an equitable degree. The AC interpersonal skills factor was significantly related to the supervisor and peer source factor, but not to the subordinate source factor. And, AC interpersonal skills were more strongly related to the peer factor than the supervisor or subordinate factors. Thus, peers' general impressions are contingent on an individual's interpersonal skills to a greater extent than are the general impressions of supervisors and subordinates. Finally, the AC leadership skills factor was significantly related to all three source factors but was more strongly related to the subordinate source factor than to either the peer or supervisor source factors. Again, these results indicate that subordinates' general impression reflect leadership skills to a greater extent than peer or supervisor general impressions.

Together, the finding that source effects were differentially correlated with externally measured constructs provides evidence that the performance relevant variance represented in MSF source effects is source specific. The pattern of correlations between the AC dimensions and the source effects is quite striking. Specifically, the AC interpersonal skills factor was the most strongly related to the peer source factor, and the AC leadership skills factor was the most strongly related to the subordinate source factor. Previous leadership research has most almost exclusively relied on subordinate raters when collecting questionnaire-based measures of leadership. Leadership researchers maintain that because of the hierarchical nature of organizations, managers have legitimate power over their subordinates and as such are most likely to attempt to exert influence on them (Yukl, 2005). In other words, influencing their subordinates is clearly an important component of any manager's job, whereas this type of behavior is not necessarily expected in managers' dealings with their supervisors or peers. Consequently, the results that AC leadership skills correlated more strongly with the subordinate source factor than the remaining source factors should not be surprising given the role requirements of managers and common practice in leadership research.

Next, the AC interpersonal skills factor was more strongly related to the peer source factor than the other source factors. At work, a successful relationship with peers is contingent upon successful interpersonal interactions and working collaboratively in order to meet organizational goals (Cardy & Dobbins, 1994). It is not surprising then that interpersonal skills were more strongly related to the peer factor than the remaining source factors. Again, because of the sheer amount of time peers spend with one another, having the necessary interpersonal skills to facilitate a cohesive working environment

may take precedence over other dimensions of performance in the general impression formed about a peer. The lack of a significant relationship between interpersonal skills and the subordinate source factor signifies that subordinates' overall impression of their manager is not contingent on the manager's interpersonal skills. In contrast to these results, previous leadership research stemming from early behavioral approaches and the more recent transformational model has consistently indicated that leader consideration behaviors are positively related to subordinate satisfaction (Fleishman, 1953; Hoffman, Oliver, & Woehr, 2005). However, the findings of the present study are not necessarily inconsistent with the results of previous leadership research. Again, these results simply indicate that subordinates' overall impression of their boss is not correlated with interpersonal skills, not that interpersonal skills do not impact subordinate ratings. In fact, the AC interpersonal skills factor was significantly related to both the MSF interpersonal performance factor and the MSF leadership performance factor. Thus, interpersonal skills do certainly play a role in other's performance ratings; however, interpersonal skills did not play a role in subordinates' general impression of their boss.

For many years, managers have occupied the role of primary performance evaluator in organizations. To this end, work performance researchers have advocated the use of managers as the primary source of performance rating information, arguing that managers are less biased evaluators than peers or subordinates, have a better gauge of what constitutes acceptable and unacceptable performance from the organization's perspective, and are at a vantage point such that they can observe a broad range of their employees' performance (Farr & Newman, 2001). Interestingly, none of the AC factors correlated with manager general impression more strongly than the other source factors.

However, the manager source factor was significantly related to each of the three AC skill factors. These results suggest that managers' general impression of subordinate performance is a function of all three of the broad AC skills and is not impacted to a greater extent by one domain of skills than the other skills. In fact, these findings support the assumptions of previous work performance researchers regarding the use of managers as evaluators of performance.

Interestingly, the conceptual/administrative skills factor correlated equitably with all three source factors. In other words, none of the raters considered conceptual/administrative skills to a greater extent than the others when forming general impressions of target performance. Thus, it appears that regardless of the organizational level of the rater, a general level of competence is equally essential to all organizational constituents. It should be noted that similar to the correlations with external constructs and MSF dimension factors, the correlations between MSF source factors and external constructs were not very large (e.g., the largest was AC leadership and subordinate source factor with $r = .29$). And, the paper and pencil constructs were generally uncorrelated with the source factors.

This suggests that in the present study, managerial skills, as opposed to traits, are the primary source of information raters use to form general impressions. Interestingly, previous leadership research has typically posited that certain traits (e.g., dominance and intelligence) are the signals that humans use to categorize others as a leader or a non-leader (Lord, Foti, & DeVader, 1984). However, the results of this study suggest that evaluators rely on target skills when forming a general impression of coworkers. Of course, these explanations for the formation of a prototype are not mutually exclusive.

Most likely, traits (e.g., intelligence) are necessary but not sufficient for the display of a skill (e.g., conceptual/administrative skills). Thus, actual behaviors (skills as measured in the AC) would be more closely causally linked to overall impressions than would traits. Stated differently, AC ratings represent behaviors that probably represent behavioral manifestations of traits. As such, AC performance is more closely linked to the display of a given skill set than is a given trait.

Since the rise in popularity of MSF, scholars have debated the meaning, if any, of MSF source effects. This study represents the first attempt to investigate this question by combining internal and external approaches to examining construct validity of MSF source effects. Although many arguments have been forwarded regarding the meaning of MSF source effects, the most basic question is whether source effects represent performance relevant variance or error. As previously discussed, the correlation between manager, peer, and subordinate effects and external measures of managerial skills suggests that source effects represent performance relevant variance. By first providing evidence that source effects represent performance relevant variance, this study justifies the further investigation of substantive explanations for the presence of source effects.

Unfortunately, no conclusion can be drawn as to *why* source effects are present in MSF data based on the findings of the present study. That is, the results of this study are consistent with multiple hypotheses regarding the substantive meaning of source effects. For example, one of the most popular explanations for the presence of source effects is that raters from different organizational levels have a greater opportunity to observe a given dimension of behavior than raters from other organizational levels. As previously mentioned, most would agree that subordinates likely have a greater opportunity to

observe leadership skills than peers or supervisors. Similarly, because of the amount of time spent together and the need for collaboration to meet organizational goals, peers likely have a greater opportunity to observe interpersonal skills than do raters from other organizational sources. Thus, the present results are consistent with the opportunity to observe hypothesis. Although consistent with the opportunity to observe hypothesis, firm support for the opportunity to observe hypothesis cannot be provided given the method taken in the present study.

For example, the results are also consistent with the ecological hypothesis (Lance & Woehr, 1989). The ecological hypothesis proposes that managers intentionally display different behaviors around different groups of raters. Although clearly akin to the opportunity to observe hypothesis, the ecological hypothesis implies manager activity in choosing which behaviors to display in the presence of different raters and according to the demand of different situations (Salam, Cox, & Sims, 1997). Again, the results of this study are consistent with this perspective as it is reasonable to suspect that a manager would actively portray a leadership role in the presence of subordinates and intentionally display interpersonal skills in the presence of peers.

Finally, another popular explanation for the presence of source effects is that different rating groups value certain skills and behaviors differently (Beaouvis & Dubois, 2004). The argument here is that raters form an overall impression of a coworker based on how valuable the coworker is to their well-being. Again, the present results are quite consistent with this hypothesis. Specifically, we would expect that if someone values a given set of behavior a great deal, the display of these behaviors would have a significant impact on one's general impression of their coworkers. That subordinates would find

leadership skills personally beneficial more so than peers and supervisors is certainly a plausible alternative. Equally plausible is that peers value interpersonal skills to a greater extent than do subordinates and supervisors.

This study contributes to the existing literature by demonstrating that source effects represent performance relevant variance and gives an indication of what aspects of managerial skills are represented in each sources' general impressions. Unfortunately, because the results of this study are consistent with multiple perspectives on the causes of MSF source effects, it is impossible to ascertain the reason for the presence of MSF source effects with the design of the present study. Clearly, additional research examining the cause of MSF source effects is warranted.

Limitations

Despite making important contributions to the MSF literature, this study is not without limitations. First, as with many performance appraisal instruments, the MSF instrument used in the present study is a "one of a kind" instrument that is specific to the university leadership development program from which the sample was drawn. Thus, the generalizability of the results is a cause for concern. However, the results of the agreement, factor structure, and equivalence analyses indicate that in general, the MSF instrument used in the present study performs in a similar manner as other MSF instruments. Thus, some confidence can be given to the similarity of this MSF instrument to other MSF instruments. Still, the replication of this study with other MSF instruments is clearly warranted.

As discussed above, the performance rating context in the present study was certainly not optimal. The MSF process in the present study diverged from the typical

performance appraisal context in that the present MSF instrument included more items, more dimensions, and was based on generic managerial competencies, as opposed to the specific qualifications of the participants' jobs (Rogelberg & Waclawski, 2001). Although these qualities of the rating context are not optimal, they represent rating characteristics typical to MSF contexts (Rogelberg & Waclawski, 2001).

In addition to the previously discussed rating context issues, the omission of rater training deserves additional attention. Although performance appraisal experts strongly recommend the incorporation of rater training whenever performance ratings are made, rater training is not commonplace in MSF systems (Antonioni & Woehr, 2001; Craig et al., 2006). Issues preventing the incorporation of rater training in typical performance appraisal settings (e.g., cost, trainee time, etc.) are compounded in the context of MSF. Specifically, whereas traditional performance appraisal systems incorporate only a worker's immediate supervisor, MSF systems incorporate many raters from many different organizational levels (an average of 8 raters for each target in this study). After factoring the time away from work for all the raters participating in MSF, the costs associated with MSF is much greater than in typical performance appraisal systems and as a result, formal rater training may be cost prohibitive. In addition, the sample for the present study was drawn from multiple international organizations. Clearly, having raters attend a formal rater training session would be impractical when the MSF tool is distributed from an external vendor, as in the present study. The incorporation of rater training into MSF systems may indeed be a case where practicality and utility collide. Future research should investigate pragmatic, yet effective methods of enhancing rating accuracy in MSF systems. Craig and his colleagues (2006) proposed that advances in

technology afford MSF designers the opportunity to develop on-line rater training and urged MSF researchers to explore the efficacy of on-line rater training.

One final limitation associated with the present MSF methodology is worthwhile to note. That it, the participants were asked to choose their own rater. Some argue that this practice may hinder the accuracy of performance ratings because feedback recipients may have the tendency to choose raters with whom they have close personal relationships. Because of their close relationship with the target manager, target chosen raters may be more likely to rate the target more positively, regardless of the individual's performance (Farr & Newman, 2001). To this end, although some research has indicated that allowing targets to chose raters decreases rater accuracy (O'Leary, Harkcom, Jackson, & Tears, 2003), other research does not (Nieman, Metlay, Kaplan, & Wolfe, 2006). Thus, although the use of target chosen raters in the present study may adversely impact the quality of the performance ratings, previous research is unclear as to this point.

One of the primary analytic tools in this study was the comparison of correlated correlations using the significance testing approach recommended by Meng and his colleagues. When testing multiple bivariate relationships for statistically significant differences, many researchers suggest correcting for the number of significant relationships that one would expect by chance (Cohen et al., 2002). This approach is recommended because as the number of significant tests conducted increases, the number of significant relationships attributable to chance will also increase. Clearly, research seeking to replicate the present findings is needed in order to ensure that the results are not attributable to chance.

Finally, the impact of individual raters was not explicitly modeled in the above structural models. However, different raters provided ratings for each target in the present sample. In other words, the same supervisor, peer, or subordinate never provided ratings for more than one target. Thus, the potentially confounding effect of rater level and specific individual rater was eliminated (Woehr et al., 2005). Essentially, the source effect components represent effects common across different supervisors, peers, or subordinates. Thus, common within source variance cannot be attributed to a common rater providing ratings. Consequently, confounding source by specific rater was not an issue in the present study.

Implications

This study has important implications for manager development. Although the ability to present feedback from multiple perspectives is ostensibly an advantage of MSF, presenting feedback recipients with contrasting feedback from multiple sources may be confusing for feedback recipients. For instance, in the case of directly contrasting feedback, a manager may have difficulty determining whose feedback to believe. Statistically compiling the feedback across sources prior to presenting it to managers may prove a useful remedy to this problem. That is, a model similar to the one supported here could be specified and factor scores generated based on the communality between all sources' ratings of a given construct (e.g., CFA derived performance dimension factor scores). Similarly, factor scores could be generated based on the source factors to present feedback recipients with multiple "overall performance" ratings, separated source. Doing so would certainly make the feedback more parsimonious and as a result, easier to interpret. Clearly, research would need to examine feedback recipient reactions to

receiving this type of feedback as opposed to the feedback separated by performance dimension and rating source typical to MSF. In addition, to altering the presentation of MSF summaries, MSF could be interpreted differently. For example, different sources' ratings could be emphasized when interpreting MSF feedback. Based on the results of this study, subordinate ratings could be emphasized when attempting to ascertain a manager's leadership skills and peer ratings could be emphasized when attempting to draw inferences regarding a manager's interpersonal skills.

Unfortunately, the convergent and discriminant validity evidence for the MSF dimensions was quite weak. These results are quite troublesome for those interesting in using MSF in developmental contexts. Clearly, presenting managers with inaccurate feedback can severely undermine the utility of MSF. Consequently, the results of this study call into question the value of presenting targets with dimension level feedback. Interestingly, the construct validity evidence accumulated for the AC dimensions was somewhat more supportive than that found for the MSF dimensions. Over the last two decades, the construct validity of AC dimensions has been the subject of substantial research attention and criticism. In fact, some have suggested that the AC dimension framework be abandoned in favor of presenting and interpreting AC feedback in the framework of exercises (Lance et al., 2000). In contrast, relatively little research has addressed the construct validity of MSF using the external approaches to construct validation (Borman, 1997; Farr, 2006). One possible reason for this disparity in research focusing on the construct validity of constructs derived from these two methods is the *perceived* respective cost of each. That is, ACs are quite expensive because of the manpower associated with having multiple trained raters observe simulation exercises

and compose narrative-style developmental summaries (Eurich, Krause, Cigularov, & Thornton, 2006). In contrast, the primary material cost associated with MSF is simply the cost of distributing multiple surveys. However, the hidden cost of rater time associated with having multiple raters take time away from work can be substantial. Thus, cost does not appear to be a credible reason for administering MSF instead of ACs as tools of managerial development. In short, organizational scientists should consider the relative utility of MSF and ACs before deciding on the preferred approach to providing managers with developmental feedback.

Finally, although a three factor model of managerial performance consisting of conceptual/administrative, interpersonal, and leadership skills was supported in the present study, it is unclear how useful a three dimension conceptualization of performance would be in developmental settings. The goal setting literature consistently demonstrates that specific goals give better information as to where to direct effort toward goal attainment and as such, are more likely to be attained (O'Leary et al., 2003). Although using a latent factor approach that combines existing subscales into broad domains of managerial performance is useful from a psychometric perspective, three broad performance dimensions may prove too broad to be useful in developmental contexts. For instance, telling a manager to improve their conceptual skills may be too broad to be useful, whereas suggesting the need to improve planning skills is much more specific and as such may be more useful at directing goal oriented behavior. However, the support for multiple subdimensions of performance is rarely evidenced with empirical research (Viswesvaran et al., 2005). This juxtaposition between the need for specific feedback and accurate measurement must be addressed to fully understand the best

practices in measuring managerial performance and the subsequent presentation of developmental feedback.

Summary and Conclusions

The present study answers the call by work performance researchers to evaluate the construct validity of MSF source and dimension factors using a nomological network approach (Borman, 1997; Farr, 2006). First, CFA was used to partial MSF source and dimension factors, and the relationships between these factors and externally measured constructs was subsequently examined. The results of agreement, equivalence, and CFA analyses indicated that in general, the present MSF instrument performed in accordance with previous research examining the psychometric properties of MSF. These findings were used to support the generalizability of the present MSF instrument to other MSF instruments.

Next, the construct validity of three MSF performance dimension factors was evaluated by examining their correlation with conceptually similar and dissimilar externally measured constructs. Surprisingly, limited construct validity evidence was provided for the MSF performance dimension factors. Clearly, these findings do not bode well for the incorporation of MSF into developmental contexts. In contrast to the troubling evidence for the construct validity of the MSF performance dimensions, somewhat stronger support was demonstrated for the construct validity of the AC skill dimensions. The stark contrast in these results is somewhat surprising given the relative amount of research attention focused on examining the construct validity of AC dimensions compared to that examining the construct validity of MSF dimensions. Based on the comparative construct validity evidence provided for MSF and AC dimensions,

organizational scientists are encouraged to further evaluate the relative efficacy of MSF and ACs as tools of manager development.

Finally, this study represents the first attempt at examining the meaning of MSF source effects through an examination of their correlation with external constructs. The meaning of source effects has been subject to considerable contention over the past twenty years, with some suggesting they are meaningless (Viswesvaran et al., 2002) and others arguing that they represent important performance relevant variance (Borman, 1974; Farr, 2006; Murphy & DeShon, 2000). The results presented here indicate that MSF source effects do indeed reflect performance relevant variance and that this performance relevant variance is source specific. Based on these results, future researchers are encouraged to further elucidate the causes of MSF source effects. It is my hope that this study will serve to stimulate awareness of issues with the constructs assessed by MSF as well as present an alternate method of examining the construct validity of MSF. Clearly specifying source and dimension effects as constructs characteristic of MSF is essential to further understanding the meaning of MSF and the subsequent appropriate application of these tools.

.

List of References

Antonioni, D. & Woehr, D.J. (2001). Improving the Quality of Multisource Rater Performance. In Bracken, D.W., Timmreck, C.W., & Church, A.H. (Eds.), *The Handbook of Multisource Feedback*. Jossey-Bass, Inc.

Arthur, W. & Villado, A.J. (in preparation). A comment on the importance of distinguishing constructs and methods in personnel selection research and practice. *Unpublished manuscript from Texas A & M University*.

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

Atkins, P.W.B. & Wood, R.E. (2002). Self-versus others' ratings as predictors of assessment center ratings: Validation evidence for 360-degree feedback programs. *Personnel Psychology, 55*, 871-904.

Atwater, L.E. & Yammarino, F.J. (1993). Personal attributes as predictors of superiors' and subordinates' perceptions of military academy leadership. *Human Relations, 46(5)*, 645-668.

Bagozzi, R.P. & Edwards, J.R. (1998). A general approach for representing constructs in organizational research. *Organizational Research Methods, 1*, 45-87.

Bales, R.F. (1970). Personality and interpersonal behavior. Holt, Rinehart, & Winston, Oxford, England.

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.

Bass, B.M. (2001). Cognitive, Social, and Emotional Intelligence of Transformational Leaders. In R.E. Riggio, S.E. Murphy, & F.J. Priozzolo (Eds.), Multiple Intelligences and Leadership (pp 54-77). Mahwah, NJ: Lawrence Erlbaum and Associates.

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

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

Borman, W.C. (1974). The rating of individuals in organizations: An alternative approach. *Organizational Behavior and Human Performance, 12*, 105-124.

Borman, W.C. (1997). 360 ratings: An analysis of assumptions and research agenda for evaluating their validity. *Human Resource Management Review, 7*, 299-315.

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

Borman, W.C. & Motowidlo, S.J. (1993). Expanding the criterion domain to include elements of contextual performance. In N. Schmitt & W. Borman (Eds.), *Personnel selection in organizations* (pp. 71-98). San Francisco: Jossey-Bass.

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

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

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. Bollen & J.S. Long (Eds.), *Testing Structural Models*, 136-162. Newbury Park,

CA: Sage.

Brutus, S., & Derayeh, M. (2002) Current practices in multi-source assessment. *Human Resources Development Quarterly, 13*, 187-202.

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

Campbell, D.J. & Lee, C. (1988). Self-appraisal in performance evaluation: Development versus evaluation. *Academy of Management Review, 13*, 302-314.

Campbell, J.P., McHenry J.J., & Wise, L.L. (1990). Modeling job performance in a population of jobs. *Personnel Psychology, 43*, 313-333.

Cardy, R.L. & Dobbins, G.H. (1994). *Performance appraisal: Alternative Perspectives*. Cincinnati, OH: South-Western Publishing Co.

Carruthers, F. (2003, November 14). Nothing but the truth. *Australian Financial Review*, 78-84.

Cheung, G.W. (1999). Multifaceted conceptions of self-other disagreement. *Personnel Psychology, 52*, 1-36.

Cheung, G.W. & Rensvold (1999). Testing factorial invariance across groups: A reconceptualization and proposed new method. *Journal of Management, 25(1)*, 1-27.

Church, A.H. & Allen, D.W. (1997). Advancing the state of the art of 360-degree feedback. *Group and Organization Management, 22*, 149-161.

Cohen, J., West, S.G., Aiken, L., & Cohen, P. (2002). *Multiple Regression/Correlation Analysis for the Behavioral Sciences* (3rd Edition). Lawrence Erlbaum Associates. Hillsdale, NJ.

Conway, J.M. (1996). Analysis and design of multitrait-multirater performance appraisal studies. *Journal of Management, 22*, 139-162.

Conway, J.M. (1999). Distinguishing contextual performance from task performance for managerial jobs. *Journal of Applied Psychology, 84*, 3-13.

Conway, J.M. (2000). Managerial performance development constructs and personality correlates. *Human Performance, 13*, 23-46.

Conway, J. M. (1996). Analysis and design of multitrait-multirater performance appraisal studies. *Journal of Management, 22*, 139-162.

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

Conway, J.M., Lombardo, K., & Sanders, K.C. (2001). A meta-analysis of incremental validity and nomological networks for subordinate and peer ratings. *Human Performance, 14 (4),* 267-303.

Cooper, W. H. (1981). Ubiquitous halo. *Psychological Bulletin, 90*, 218-24.

Coovert, M.D., Craiger, J.P., & Teachout, M.S. (1997). Effectiveness of the direct product versus confirmatory factor model for reflecting the structure of multimethod-multirater job performance data. *Journal of Applied Psychology, 82*, 271-280.

Cote, J.A. & Buckley, R.A. (1987). Estimating trait, method, and error variance: Generalizing across 70 construct validation studies. *Journal of Marketing Research, 24*, 315-318.

Craig, S.B., Hayes, B.C., Preston, L.A., Lebow, R. (2006). Rater training for 360

assessment: Creating beta and gamma change. In B.C. Hayes (Chair), *The Four "Rs" of 360 feedback: Determinants of its effectiveness.*. Symposium conducted at the 21st annual conference of the Society for Industrial and Organizational Psychology, Dallas, TX.

Craik, K.H., Ware, A.P., Kamp, J., O-Reilly, C., Staw, B., & Zedeck, S. (2002). Explorations of construct validity in a combined managerial and personality assessment programme. Journal *of Occupational and Organizational Psychology, 75(2)*, 171-193.

Cronbach, L.J. & Meehl, P.E. (1955). Construct validity in psychological tests. *Psychological Bulletin, 62*, 281-302.

Diefendorff, J.M., Silverman, S.B., & Greguras, G.J. (2005). Measurement equivalence and multisource ratings for non-managerial positions: Recommendations for research and practice. *Journal of Business and Psychology, 19 (3)*, 399-425.

Eurich, T.A., Krause, D.E., Cigularov, K., & Thornton, G.C. (2006). Assessment Centers: Current Practices in the United States. Poster presented at the 21st annual conference of the Society for Industrial and Organizational Psychology, Dallas, TX.

Facteau, J.D. & Craig, S.B. (2001). Are performance ratings from different ratings sources comparable? *Journal of Applied Psychology, 86 (2),* 215-227.

Farr, J.L. (2006). In D.W. Bracken & C.W. Timmreck (Co-Chairs), *The Evolution of 360 Feedback: What Hath We Wrought?*. Symposium conducted at the 21st annual conference of the Society for Industrial and Organizational Psychology, Dallas, TX.

Farr, J.L. & Newman, D.A. (2001). Rater Selection: Sources of Feedback. In Bracken, D.W., Timmreck, C.W., & Church, A.H. (Eds.), *The Handbook of Multisource Feedback*. Jossey-Bass, Inc.

Fiske, D.W. (1982). Convergent-discriminant validation in measurements and research strategies. In D. Brinberg & L. Kidder (Eds.), *New directions for methodology of social and behavioral science: Forms of validity in research.* (Vol. 12, pp. 72-92). San Francisco: Jossey Bass.

Fleishman, E.A. (1953). The description of supervisory behavior. *Personnel Psychology, 37*, 1-6.

Fleishman, E.A., Mumford, M.D., Zaccaro, S.J., Levin, K.Y., Korotkin, A.L., & Hein, M.B. (1991). Taxonomic efforts in the description of leader behavior: A synthesis and functional interpretation. *Leadership Quarterly*, *2*, 245-287.

Goffin, R.D. & Jackson, D.N. (1992). Analysis of multitrait-multirater performance appraisal data: Composite direct product method versus confirmatory factor analysis. *Multivariate Behavioral Research, 27 (3)*, 363-385.

Gough, H.G., & Bradley, P. (1996). *California Psychological Inventory Manual*. Palo Alto CA: Counseling Psychologist's Press.

Harris, M.M. & Schaubroeck, J. (1988). A meta-analysis of self-supervisor, self-peer, and peer-supervisor ratings. *Personnel Psychology, 41*, 43-62.

Hazucha, J.F., Hezlett, S.A., & Schneider, R.J. (1993). The impact of multi-source feedback on management skills development. *Human Resource Management, 32*, 325-351.

Healy, M. C., Walsh, A. B., & Rose, D. S. (2003). A benchmarking study of North

American 360-degree feedback practices. Paper presented at the 18th Annual Conference of the Society for Industrial and Organizational Psychology, Orlando, FL.

Hoffman, B.J., Blair, C., Meriac, J., & Woehr, D.J. (in-press). Expanding the criterion domain? A quantitative review of the OCB literature. *Journal of Applied Psychology.*

Hoffman, B.J. & Frost, B.C. (2006). Multiple intelligences of transformational leaders: An empirical examination. *International Journal of Manpower, 27(1),* 37-51.

Hoffman, B.J. & Kudisch, J. & (2002). Expanding the nomological network of assessment center dimensions using external measures of cognitive ability and personality. In J.D. Kudisch, *Assessment center construct validity research: New methods for examining an old problem*. Symposium conducted at the 18th annual conference of the Society for Industrial and Organizational Psychology, Toronto, Canada.

Hoffman, B.J., Oliver, J.T., & Woehr, D.J. (2005). A Quantitative Review of the Structural Validity of the MLQ. Poster presented at the 20th annual meeting of the Society for Industrial and Organizational Psychology, Los Angeles, California.

Hogan, J. & Holland, B. (2003). Using theory to evaluate personality and job performance relations: A socioanalytic perspective. *Journal of Applied Psychology, 88(1)*, 100-112.

Holzbach, R.L. (1978). Rater bias in performance ratings: Superior, self, and peer ratings. *Journal of Applied Psychology, 63*, 579-588.

Hotelling, H. (1940). The selection of variates for use in prediction, with some comments

on the general problem of nuisance parameters. *Annals of Mathematical Statistics, 11*, 271-283.

Howell, D.C. (1992). *Statistical methods for psychology* (3rd ed.). Belmont, CA: Duxbury.

Hulin, C.L. (1982). Some reflection on general performance dimensions and halo rating error. *Journal of Applied Psychology, 67*, 165-170.

James, L.R., Demaree, R.G., & Wolf, G. (1984). Estimating with-in group interrater reliability with and without response bias. *Journal of Applied Psychology, 69*, 85-98.

Joreskog, K.G. & Sorbom, D. (1996). *LISREL 8.5: User's reference guide*. Chicago: Scientific Software.

Judge, T.A., Colbert, A.E., & Ilies, R. (2004). Intelligence and leadership: A quantitative review and test of theoretical propositions. *Journal of Applied Psychology, 89(3)*, 542-552.

Katz, D., Maccoby, N., & Morse, N. (1950). *Productivity, supervision, and morale in an office situation*. Ann Arbor, MI: Institute for Social Research.

Katz, D. & Kahn, R.L. (1952). Some recent findings in human-relations research in industry. In E. Swanson, T. Newcomb, & E. Hartley (Eds.), *Readings in social psychology* (pp.650-665). New York: Holt.

Katz, R.L. (1974). Skills of an effective administrator. *Harvard Business Review, 52*, 90-102.

King, L.M., Hunter, J.E., & Schmidt, F.L. (1980). Halo in a multidimensional forced-

choice performance evaluation scale. *Journal of Applied Psychology, 65*, 507-516.

Klimoski, R.J. & London, M. (1972). Role of the rater in performance appraisal. *Journal of Applied Psychology, 59 (4)*, 445-451.

Lance, C.E. & Bennett, W. (1997). *Rater source differences in cognitive representations of performance dimensions*. Paper presented at the annual meeting of the Society for Industrial and Organizational Psychology, St. Louis, MO.

Lance, C.E., Newbolt, W.H., Gatewood, R.D., Foster, M.R., French, N.R., & Smith, D.E. (2000). Assessment center exercise factors represent cross-situational specificity, not method bias. *Human Performance, 13(4)*, 323-353.

Lance, C.E., Noble, C.L., & Scullen, S.E. (2002). A critique of the correlated trait-correlated method and correlated uniqueness models for multitrait-multimethod data. *Psychological Methods*, 228-274.

Lance, C. E., Teachout, M. S., & Donnelly, T. M. (1992). Specification of the criterion construct space: An application of hierarchical confirmatory factor analysis. *Journal of Applied Psychology, 77*, 437-452.

Lance C.E., & Woehr, D.J. (1986). Statistical control of halo: A clarification of two models from the performance appraisal process. *Journal of Applied Psychology, 71*, 679-685.

Lance C.E., & Woehr, D.J. (1989). The validity of performance judgments: Normative accuracy model versus ecological perspectives. In D.F. Ray (Ed.), *Southern Management Association Proceedings* (pp. 115-117). Oxford, MS: Southern Management Association.

Lance, C. E., Woehr, D. J., & Fisicaro, S. A. (1990). Cognitive categorization processes in performance evaluation: Confirmatory tests of two models. *Journal of Organizational Behavior*, 12, 1-20.

Lawler, E.E. III. (1967). The Multitrait-Multirater approach to measuring managerial job performance. *Journal of Applied Psychology, 51*, 369-381.

LePine, J. A., Erez, A., & Johnson, D. E. (2002). The nature and dimensionality of organizational citizenship behavior: A critical review and meta-analysis. *Journal of Applied Psychology, 87,* 52-65.

London, M. & Smither, J.W. (1995). Can multi-source feedback change perceptions of goal accomplishment, self-evaluations, and performance -related outcomes? Theory-based applications and directions for future research. *Personnel Psychology, 48*, 803-839.

Lord, R.G., Foti, R.J., & DeVader, C.L. (1984). A test of leadership categorization theory: Internal structure, information processing, and leadership perceptions. *Organizational Behavior and Human Performance, 34*, 343-378.

Mann, F.C. (1965). Toward an understanding of the leadership role in formal organizations. In R. Dubin, G.C. Homans, F.C. Mann, & D.C. Miller (Eds.), *Leadership and Productivity* (pp. 68-77). San Francisco: Chandler.

Maurer, T., Raju, N.S., & Collins, W.C. (1998). Peer and subordinate performance measurement equivalence. *Journal of Applied Psychology, 83*, 693-702.

Meng, X.L., Rosenthal, R., & Rubin, D.B. Comparing Correlated Correlation Coefficients. *Psychological Bulletin, 111 (1)*, 172-175.

Mento, A.J., Steel, R.P., Karren, R., J. (1987). A meta-analytic study of the effects of

goal setting on task performance: 1966-1984. *Organizational Behavior and Human Decision Processes, 39 (1)*, 52-83.

Mintzberg, H. (1975). The manager's job: Folklore and fact. *Harvard business review, 53*, 49-61.

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

Mulaik, S.A., James, L.R., Van Alstine, J., Bennett, N., Lind, S., & Stilwel, C.D. (1989). Evaluation of goodness-of-fit indices for structural equation models. *Psychological Bulletin, 105*, 430-445.

Murphy, K.R. (1989). Dimensions of job performance. In R.F. Dillon & J.W. Pellegrino (Eds.), *Testing: Theoretical and applied perspectives* (pp. 218-247). New York: *Praeger*.

Murphy, K.R. & Cleveland, J.N. (1995). *Understanding performance appraisal: Social, organizational, and goal based approaches*. Thousand Oaks, CA: Sage.

Murphy, K.R., Cleveland, J.N., & Mohler, C.J. (2001). Reliability, validity, and meaningfulness of multisource ratings. In D.W. Bracken, C.W. Timmreck, & A.H. Church (Eds.), *The Handbook of multisource feedback* (pp. 130-148). San Francisco: Jossey-Bass.

Murphy, K.R. & DeShon, R. (2000). Interrater correlations do not estimate the reliability of job performance ratings. *Personnel Psychology, 53*, 873-900.

Nieman, J. Metlay, W., Kaplan, I.T., & Wolfe, K. (2006). The effect of rater selection on

rating accuracy. Poster presented at the 21st annual conference of the Society for Industrial and Organizational Psychology, Dallas, TX.

O'Leary, R. S., Harkcom, K. R., Jackson, K. A., & Tears, R. S. (2003). Rater selection in performance evaluation: Does it make a difference? Paper presented at the 18th Annual Conference for the Society of Industrial and Organizational Psychology, Orlando, Fl.

Rogelberg, S.G. & Waclawski, J. (2001). Instrumentation Design. In Bracken, D.W., Timmreck, C.W., & Church, A.H. (Eds.), *The Handbook of Multisource Feedback*. Jossey-Bass, Inc.

Salam, S., Cox, J.F., & Sims Jr., H.P. (1997). In the eye of the beholder: How leadership relates to 360-degree performance ratings. *Group and Organizational Management, 22(2),* 185-209.

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.

Schmidt, F.L., Viswesvaran, C., & Ones, D.S. (2000). Reliability is not validity and validity is not reliability. *Personnel Psychology, 53*, 901-912.

Schmitt, N. & Stults, D.M. (1986). Method review: Analysis of multitrait-multimethod matrices. *Applied Psychological Measurement, 10*, 1-22.

Scullen, S.E., Mount, M.K., & Goff (2000). Understanding the latent structure of job performance ratings. *Journal of Applied Psychology, 85*, 956-970.

Scullen, S.E., Mount, M.K., & Judge, T.A. (2003). Evidence of the construct validity of

developmental ratings of managerial performance. *Journal of Applied Psychology, 88(1)*, 50-66.

Seifert, C.F., Yukl, G., & McDonald, R.A. (2003). Effects of multisource feedback and a feedback facilitator on the influence behavior of managers toward subordinates. *Journal of Applied Psychology, 88*, 561-569.

Shore, T.H., Thornton, G.C., & Shore, L.M. (1990). Construct validity of two categories of assessment center dimension ratings. *Personnel Psychology, 43(1)*, 101-116.

Smither, J.W., London, M., & Reilly, R.R. (2005). Does performance improve following multisource feedback? A theoretical model, meta-analysis, and review of empirical findings. *Personnel Psychology, 58*, 33-66.

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

Tett, R.P., Jackson, D.N., & Rothstein, M. (1991). Personality measures as predictors of job performance: A meta-analytic review. *Personnel Psychology, 47,* 157-172.

Thompson, H.S. (1971). *Fear and Loathing in Las Vegas: A Savage Journey to the Heart of the American Dream*. Vintage, Los Angeles, CA.

Timmereck, C.W. & Bracken, D.W. (1997). Multisource feedback: A study of its use in decision-making. *Employment Relations Today, 24*, 21-27.

Tornow, W.W. & Pinto, P.R. (1976). The development of a managerial job taxonomy: A system for describing, classifying, and evaluating executive positions. *Journal of Applied Psychology, 61*, 410-418.

Tucker, L.R & Lewis, C. (1973). A reliability coefficient for maximum likelihood factor analysis. *Psychometrika*, *38*, 1–10.

Vance, R. J., MacCallum, R. C., Coovert, M. D., & Hedge, J. W. (1988). Construct validity of multiple job performance measures using confirmatory factor analysis. *Journal of Applied Psychology, 73*, 74-80.

Vandenberg, 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. (1993). *Modeling job performance: Is there a general factor?* Unpublished doctoral dissertation, University of Iowa, Iowa City.

Viswesvaran, C., Ones, D.S., & Schmidt, F.L. (1996). Comparative analysis of the reliability of job performance ratings. *Journal of Applied Psychology, 81*, 557-574.

Viswesvaran, C., Schmidt, F.L., & Ones, D.S. (2002). The Moderating Influence of Job Performance Dimension on Convergence of Supervisory and Peer Ratings of Job Performance: Unconfounding Construct-Level Convergence and Rating Difficulty. *Journal of Applied Psychology, 87 (2)*, 245-354.

Viswesvaran, C., Schmidt, F.L., & Ones, D.S. (2005). Is there a general factor in ratings of performance? A meta-analytic framework for disentangling substantive and error influences. *Journal of Applied Psychology, 90(1)*, 108-131.

Watson, G., & Glaser, E. (1980). *Manual for the Watson Glaser Critical Thinking Appraisal*. New York: Psychological Corporation.

Wernimoont, P.F. & Campbell, J.P. (1968). Signs, samples, and criteria. *Journal of Applied Psychology, 52(5)*, 372-376.

West, S.G., Finch, J.F., & Curran, P.J. (1995). Structural equation models with

nonnormal variables: Problems and remedies. In R.H. Hoyle (Eds.), *Structural equation modeling: Concepts, issues, and applications* (pp. 56-75), Thousand Oaks, CA: Sage.

Widman, K.F. (1985). Hierarchically nested covariance structure models for multitrait-multimethod data. *Applied Psychological Measurement, 9*, 1-26.

Woehr, D.J., Hoffman, B.J. & Maldegan, R. (2004). A meta-analytic investigation of the relationship between leader traits and leader effectiveness. *15th Annual International Society for the Study of Work Values Proceedings.*

Woehr, D.J. & Huffcutt, A.I. (1994). Rater training for performance appraisal: A quantitative review. *Journal of Occupational and Organizational Psychology, 67*, 189-206.

Woehr, D.J., Sheehan, M.K., & Bennett, W. (2005). Assessing measurement equivalence across ratings sources: A multitrait-multirater approach. *Journal of Applied Psychology, 90 (3),* 592-600.

Yammarino, F.J. & Bass, B.M. (1990). Long-term forecasting of transformational leadership and its effects among naval officers: Some preliminary findings. In Clark, K.E. & Clark, M.B. (Eds.). *Measures of leadership*, (pp. 151-169). West Orange, NJ, US: Leadership Library of America, Inc.

Yukl, G.A. (1989). Managerial leadership: A review of theory and research. *Journal of Management, 15*, 251-289.

Yukl, G.A. (2005). *Leadership in Organizations* (6th edition). New Jersey: Prentice-Hall, Inc.

Zedeck, S. & Baker, H.T. (1972). Nursing performance as measured by behavioral

expectation sales: A multitrait-multirater analysis. *Organizational Behavior and Human Performance, 7 (3)*, 457-466.

Appendices

Appendix A

Definition of Assessment Center Dimensions and California Psychological Inventory Folk Scales

Analysis	Identifying key issues, securing relevant information, and assimilating data from different sources to ascertain possible causes of problems.
Coaching	Providing advice, guidance, and/or specific suggestions directed at helping others improve their performance.
Decisiveness	Rendering firm decisions, committing to clear courses of action, and doing so in an emphatic tone.
Judgment	Providing suggestions based on logical assumptions based on factual/accurate information, generating alternative courses of action, and considering the implications of one's actions.
Oral Communication	Communicating ones message clearly with appropriate tone, inflection, enthusiasm, and expressiveness.
Planning & Organizing	Planning for upcoming events, approaching issues in an organized format, and using a strategic focus.
Persuasion	Effectively convincing others to follow one's chosen course of action by asserting dissenting opinions, defending one's perspective when challenged, and doing so tactfully.
Sensitivity	Concern for the feelings and needs of others.
Team building	Working to build consensus among team members, working effectively as part of a larger group.

Appendix B

Definition of California Psychological Inventory Folk Scales

Dominance	Confident, assertive, task-oriented.
Empathy	Understands the feelings of others, perceptive of social nuances, optimistic.
Flexibility	Likes change and variety; easily bored by routine and everyday experience.
Good Impression	Wants to make a good impression; tries to please others.
Independence	Self-sufficient and resourceful; persistent in seeking goals whether others agree or not.
Responsibility	Responsible, reliable, serious about obligations.
Sociability	Sociable, likes to be with people, friendly.
Tolerance	Tolerant of others' beliefs and values, fair-minded and tactful.

Appendix C

Means, Standard Deviations, Coefficient Alpha Reliabilities, and Intercorrelations Among Study Variables

Means, Standard Deviations, Coefficient Alpha Reliabilities, and Intercorrelations Among All Study Variables

Variable		Mean	SD	1	2	3	4	5	6	7	8	9	10	11	12
1	IIM	4.02	.65	.76											
2	IIP	4.05	.52	.36	.81										
3	IIS	4.09	.48	.26	.36	.83									
4	IMM	3.99	.55	.67	.30	.15	.80								
5	IMP	3.96	.48	.39	.77	.36	.38	.87							
6	IMS	4.03	.42	.24	.37	.75	.23	.42	.84						
7	ISM	3.94	.56	.62	.26	.18	.51	.25	.13	.72					
8	ISP	4.02	.44	.23	.69	.24	.21	.64	.26	.25	.70				
9	ISS	4.09	.41	.22	.32	.67	.17	.27	.69	.18	.34	.73			
10	TBM	4.02	.44	.59	.28	.19	.59	.35	.19	.48	.15	.12	.81		
11	TBP	4.08	.46	.27	.69	.28	.25	.66	.26	.17	.58	.23	.38	.88	
12	TBS	4.15	.46	.26	.32	.76	.21	.32	.69	.16	.21	.66	.32	.37	.84
13	CSKM	3.87	.58	.53	.18	.10	.47	.24	.13	.48	.09	.08	.68	.26	.20
14	CSKP	3.92	.42	.24	.59	.23	.20	.60	.19	.21	.56	.21	.30	.68	.28
15	CSKS	3.97	.40	.17	.25	.63	.18	.24	.63	.14	.17	.66	.20	.23	.70
16	SENM	4.12	.60	.56	.31	.18	.51	.31	.17	.43	.20	.18	.75	.42	.33
17	SENP	4.19	.46	.16	.51	.20	.13	.44	.22	.13	.39	.17	.30	.74	.31
18	SENE	4.22	.48	.30	.28	.59	.11	.21	.57	.09	.14	.54	.23	.32	.70
19	PMM	3.94	.64	.49	.18	.11	.35	.18	.04	.48	.15	.14	.37	.16	.15
20	PMP	4.00	.44	.28	.71	.22	.20	.64	.23	.19	.58	.18	.24	.60	.24

Note. Correlations .11 or greater are significant at $p < .05$; Correlations .14 or greater are significant at $p < .01$; N = 360; Coefficient alphas are listed along the diagonal; Final letter of M = manager ratings; P = peer ratings; S = subordinate ratings; A = assessment center ratings; II = idealized influence; IM = inspirational motivation; IS = intellectual stimulation; PM = performance management; TB = team building; CSK = confrontation skill; SEN = sensitivity; COM = communication skills; AN = analysis; JD = judgment; PO = planning and organizing; OC = oral communication; AN = analysis; JD = judgment; DC = decisiveness; PO planning and organizing; LD = leadership; CO = coaching; CN = confrontation; TB = team building; SN = sensitivity;; DO = dominance; SP = social presence; IN = independence; EM = empathy; RE = responsibility; GI = good impression; TO = tolerance; FX = flexibility; CTA = critical thinking ability

Variable		Mean	SD	1	2	3	4	5	6	7	8	9	10	11	12
21	PMS	4.03	.44	.14	.26	.65	.04	.28	.61	.07	.21	.64	.10	.24	.64
22	ANM	4.02	.54	.54	.27	.14	.41	.20	.11	.66	.20	.19	.45	.15	.16
23	ANP	4.08	.37	.23	.66	.26	.15	.60	.27	.27	.66	.28	.20	.59	.22
24	ANS	4.12	.40	.17	.24	.70	.11	.18	.64	.15	.16	.69	.08	.12	.63
25	JDM	4.07	.51	.55	.23	.15	.44	.21	.11	.62	.17	.18	.53	.21	.23
26	JDP	4.12	.39	.28	.65	.25	.19	.58	.24	.30	.62	.26	.20	.63	.29
27	JDS	4.13	.42	.21	.30	.68	.14	.24	.64	.21	.18	.65	.12	.19	.63
28	POM	3.98	.64	.49	.25	.14	.43	.23	.09	.56	.17	.19	.39	.17	.17
29	POP	4.02	.45	.27	.64	.24	.21	.60	.28	.30	.61	.29	.16	.56	.24
30	POS	4.02	.48	.17	.29	.62	.11	.26	.63	.18	.20	.65	.08	.18	.58
31	COMM	4.10	.59	.53	.37	.21	.49	.37	.16	.47	.24	.17	.46	.31	.20
32	COMP	4.17	.43	.29	.66	.28	.24	.60	.26	.23	.51	.23	.21	.55	.21
33	COMS	4.24	.42	.16	.30	.59	.12	.24	.58	.12	.18	.54	.05	.19	.51
34	OCA	3.08	.22	.12	.14	.14	.11	.18	.12	.03	.10	.06	.02	.07	.11
35	ANA	3.40	.36	.14	.12	.11	.13	.14	.12	.17	.13	.16	.09	.06	.12
36	JDA	3.29	.40	.15	.10	.12	.14	.16	.15	.14	.13	.16	.12	.06	.11
37	POA	3.20	.39	.06	.10	.05	.00	.09	.03	-.01	.10	.06	-.05	.06	.05
38	DCA	3.14	.44	.02	.08	.07	.02	.09	.08	.02	.03	.03	.01	.03	.04
39	LDA	3.27	.51	.11	.15	.25	.06	.17	.23	.08	.12	.26	.06	.10	.25
40	COA	3.04	.43	.11	.11	.15	.08	.15	.12	.00	.03	.04	.07	.05	.09
41	TBA	3.27	.39	.07	.13	.12	.12	.12	.08	.07	.14	.08	.08	.17	.12
42	CNA	3.26	.42	.12	.12	.17	.05	.15	.15	.11	.08	.14	.12	.07	.12
43	SNA	3.23	.39	.05	.14	.07	.07	.16	.02	.03	.11	-.05	.16	.16	.02
44	DO	65.03	8.58	.07	.03	.08	.03	.07	.07	.01	.04	.06	-.10	-.04	.04
45	SY	53.17	7.89	.01	-.02	.05	.01	.06	.06	-.10	-.01	.06	-.05	.02	.02
46	IN	58.69	6.86	-.05	.04	.04	-.04	.03	.00	-.01	.05	.07	-.14	-.07	-.04
47	EM	53.34	8.56	.04	.07	.06	.00	.10	.04	-.03	.02	.06	.03	.11	.06
48	RE	55.39	6.82	.05	-.08	-.04	.03	-.01	-.07	.05	-.06	-.06	.05	-.05	-.04
49	GI	56.62	8.55	-.05	-.03	.02	-.06	.01	.02	-.01	-.02	.02	.00	-.02	.04
50	TO	55.07	7.11	.03	-.08	.04	-.07	-.02	-.07	.05	-.02	.03	.07	-.01	-.05
51	FX	47.70	9.03	.00	.04	.10	-.05	.07	.06	.00	.10	.08	.00	.01	-.02
52	CTA	63.73	7.42	.00	-.03	.00	-.04	-.04	-.12	.02	.02	.02	.01	.01	-.05

Variable		13	14	15	16	17	18	19	20	21	22	23	24	25	26
13	CSKM	.85													
14	CSKP	.33	.76												
15	CSKS	.20	.31	.77											
16	SENM	.66	.34	.20	.73										
17	SENP	.25	.61	.19	.46	.75									
18	SENE	.21	.28	.66	.33	.46	.81								
19	PMM	.43	.13	.08	.31	.06	.04	.82							
20	PMP	.20	.57	.15	.20	.37	.11	.32	.83						
21	PMS	.05	.13	.60	.05	.12	.45	.24	.33	.85					
22	ANM	.57	.16	.14	.44	.09	.09	.52	.20	.15	.89				
23	ANP	.20	.63	.18	.23	.50	.25	.15	.61	.26	.28	.90			
24	ANS	.11	.10	.66	.10	.10	.58	.10	.14	.66	.23	.25	.88		
25	JDM	.61	.20	.17	.53	.19	.18	.51	.23	.11	.75	.24	.19	.86	
26	JDP	.25	.68	.24	.31	.59	.29	.15	.64	.24	.28	.79	.23	.32	.83
27	JDS	.11	.17	.68	.14	.22	.62	.09	.24	.64	.21	.26	.80	.26	.37
28	POM	.43	.16	.17	.40	.16	.12	.49	.21	.13	.65	.19	.17	.63	.22
29	POP	.18	.56	.18	.21	.42	.18	.22	.67	.24	.27	.71	.24	.22	.72
30	POS	.10	.16	.60	.09	.15	.50	.10	.27	.65	.20	.30	.73	.19	.31
31	COMM	.46	.29	.12	.48	.21	.14	.36	.24	.04	.57	.28	.20	.58	.28
32	COMP	.19	.53	.15	.26	.45	.28	.19	.52	.16	.23	.64	.19	.21	.60
33	COMS	.05	.13	.52	.08	.11	.57	.06	.15	.44	.17	.23	.70	.13	.20
34	OCA	.09	.13	.08	.05	.02	.06	.10	.13	.00	.03	.03	.06	.07	.07
35	ANA	.08	.15	.17	.06	.00	.10	.07	.02	.04	.15	.05	.11	.11	.08
36	JDA	.10	.11	.15	.11	.06	.13	.08	.02	.04	.13	.04	.10	.14	.05
37	POA	-.05	.08	.13	-.08	-.04	.03	.00	.11	.11	.03	-.01	.12	-.03	-.02
38	DCA	.11	.17	.68	.14	.22	.62	.09	.24	.64	.21	.26	.80	.26	.37

Variable		27	28	29	30	21	32	33	34	35	36	37	38	39	40
27	JDS	1.0													
28	POM	.20	1.0												
29	POP	.26	.34	1.0											
30	POS	.74	.26	.42	1.0										
31	COMM	.15	.50	.30	.14	1.0									
32	COMP	.24	.20	.60	.28	.46	1.0								
33	COMS	.63	.13	.19	.57	.33	.39	1.0							
34	OCA	.06	.05	.00	.03	.17	.18	.17	1.0						
35	ANA	.11	.11	.06	.06	.17	.14	.18	.27	1.0					
36	JDA	.12	.14	.06	.07	.18	.13	.18	.23	.68	1.0				
37	DCA	.07	.10	.08	.06	.06	.03	.17	.15	.41	.35	1.0			
38	POA	.05	.04	.03	.02	.10	.11	.14	.18	.45	.44	.30	1.0		
39	LDA	.21	.05	.09	.16	.17	.19	.28	.39	.44	.38	.20	.34	1.0	
40	COA	.10	.02	.09	.07	.12	.10	.16	.29	.22	.26	.06	.25	.42	1.0
41	TBA	.06	.03	.12	.05	.14	.13	.13	.19	.19	.22	.18	.13	.22	.16
42	CNA	.06	.08	.06	.07	.16	.10	.14	.23	.31	.30	.06	.24	.47	.34
43	SNA	-.03	.05	.12	-.03	.13	.16	.04	.32	.23	.28	.18	.18	.43	.34
44	DO	.06	.01	-.06	.06	.05	.04	.11	.13	.08	.08	.10	.14	.29	.09
45	SY	.00	-.06	-.11	.03	-.01	.03	.02	.12	.03	.06	.01	.07	.15	.03
46	IN	-.01	-.03	-.03	.02	.03	.04	.09	.09	.06	.06	.08	.10	.20	.07
47	EM	-.01	-.06	-.02	-.01	.09	.09	.08	.13	.09	.10	.14	.08	.17	.00
48	RE	-.04	.02	-.09	-.09	.08	-.04	-.03	.05	.10	.11	.09	.10	.13	.03
49	GI	-.02	.07	.01	.00	.02	-.03	-.01	.00	-.04	-.04	-.01	-.02	.05	-.02
50	TO	.01	.03	-.04	-.06	.08	-.03	.05	.06	.06	.08	.10	.07	.14	.06
51	FX	.00	-.10	-.05	-.02	.02	.04	.10	.10	.08	.11	.09	.07	.09	.07
52	CTA	.02	-.01	-.04	-.03	.07	.03	.06	.08	.24	.24	.25	.23	.18	.09

Variable		13	14	15	16	17	18	19	20	21	22	23	24	25	26
39	LDA	.06	.15	.24	.02	.02	.20	.08	.02	.14	.05	.04	.19	.03	.08
40	COA	.08	.10	.03	.03	.04	.10	.08	.13	.16	.08	.06	.09	.09	.06
41	TBA	.10	.19	.03	.08	.14	.10	-.04	.06	-.03	.01	.06	.06	.04	.15
42	CNA	.06	.13	.09	.05	.04	.11	.08	.07	.06	.12	.06	.11	.08	-.01
43	SNA	.14	.16	-.02	.13	.14	.08	.08	.13	.02	.04	.11	-.05	.00	.06
44	DO	-.05	.01	.09	-.16	-.15	-.04	.04	.01	.14	-.03	-.06	.08	-.06	-.07
45	SY	-.13	.00	.07	-.09	-.05	.01	-.08	-.11	.05	-.15	.00	.04	-.12	-.06
46	IN	-.07	-.05	.04	-.16	-.19	-.09	-.02	-.02	-.01	-.01	.00	.07	-.09	-.04
47	EM	.01	.10	.09	.03	.03	.06	-.06	-.03	.00	-.09	.06	.04	-.05	-.02
48	RE	-.01	-.01	-.03	.05	.03	.01	.03	-.08	-.04	.00	-.02	-.01	.00	-.05
49	GI	.03	.02	.05	.01	.06	.09	.05	-.04	.01	-.03	.01	.05	-.05	-.02
50	TO	.03	-.02	.00	.08	.00	-.01	.08	-.06	-.01	.02	.04	.06	.03	-.01
51	FX	.00	.01	.05	-.02	-.08	.02	-.07	-.04	-.05	-.02	.09	.02	-.02	-.02
52	CTA	.04	.02	.04	-.03	-.06	-.05	-.03	-.05	.00	.08	.06	.07	.05	.07

Variable		41	42	43	44	45	46	47	48	49	50	51	52
41	TBA	1.0											
42	CNA	.12	1.0										
43	SNA	.28	.40	1.0									
44	DO	.04	.14	.13	1.0								
45	SY	.05	.08	.10	.58	1.0							
46	IN	.02	.12	.06	.61	.46	1.0						
47	EM	.07	.04	.18	.36	.56	.27	1.0					
48	RE	.10	.07	.14	.33	.20	.20	.24	1.0				
49	GI	.12	.05	.11	.08	.09	.15	.12	.43	1.0			
50	TO	.10	.06	.17	.17	.21	.23	.37	.61	.38	1.0		
51	FX	.11	.05	.10	.07	.25	.27	.44	.13	-.03	.37	1.0	
52	CTA	.13	.03	.10	.10	.08	.11	.16	.20	-.12	.29	.16	1.0

Brian J. Hoffman received his Associate of Arts degree from Hinds Community College in 1999. Then, he graduated Summa Cum Laude with a Bachelor of Science in Psychology and a minor in Management from the University of Southern Mississippi in 2001. His doctoral degree in Industrial/Psychology was conferred from the University of Tennessee in 2006.

Brian's research has been published in a variety of outlets including the Journal of Applied Psychology and the Journal of Vocational Behavior and he has authored some 25 professional presentations. In fact, Brian's research has received recognition in the form of awards and award nominations from the Society of Industrial and Organizational Psychology, the Academy of Management -HR Division, and the Society for Human Resource Management and has been the subject of national media attention. He is currently employed as an Assistant Professor in the Applied Psychology Program at the University of Georgia.