PATTERNS OF MULTISOURCE PERFORMANCE RATINGS: AN INTEGRATED APPROACH FOR EXAMINING (DIS)AGREEEMNT ACROSS SOURCES

by

BETHANY HOFFMAN BYNUM

(Under the Direction of Gary Lautenschlager)

ABSTRACT

The present study contributes to the Multisource performance ratings literature by applying a person-centered approach to the investigation of cross-source agreement. Using latent profile analyses, ratees were grouped based on their pattern of ratings received by their supervisors, peers, and subordinates. The majority of cross-source performance profiles that emerged were associated with performance level differences and not differences in source ratings. Nevertheless, for a small number of individuals systematic patterns emerged that reveal homogeneous subsamples with ratings that diverge across sources. Additionally, dominance, well-being and social presence were significant predictors, and derailment and overall performance effectiveness were significant outcomes of cross-source performance profiles. However, the majority of findings suggest a linear relationship among the predictors and outcomes, which were accounted for by performance level and not rater disagreement. This study has key implications for researchers and practitioners faced with interpreting and using the results of multisource feedback.

INDEX WORDS: Multisource performance ratings, Latent Profile Analyses, cross-source agreement

PATTERNS OF MULTISOURCE PERFORMANCE RATINGS: AN INTEGRATED APPROACH FOR EXAMINING (DIS)AGREEEMNT ACROSS SOURCES

by

BETHANY HOFFMAN BYNUM

B.S., University of Kentucky, 2005

M.S., The University of Georgia, 2007

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

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2010

PATTERNS OF MULTISOURCE PERFORMANCE RATINGS: AN INTEGRATED APPROACH FOR EXAMINING (DIS)AGREEEMNT ACROSS SOURCES

by

BETHANY HOFFMAN BYNUM

Major Professor:	Gary Lautenschlager
Committee:	Brian Hoffman
	Charles Lance

Electronic Version Approved:

Maureen Grasso
Dean of the Graduate School
The University of Georgia
August 2010

DEDICATION

This project is dedicated to my husband. Matt, thank you for supporting me, encouraging me and making me stop, laugh and have fun. This journey is more exciting and special with you by my side.

ACKNOWLEDGEMENTS

I would like to thank Brian Hoffman for the extraordinary mentorship and support he has given me throughout this project. I would also like to thank my parents, Gene and Peggy, for reading, editing and commenting on various drafts of this paper.

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

Page

CHAPTER 1

INTRODUCTION

Multisource performance ratings (MSPR) are commonplace in organizations as a means for providing development feedback to individuals. In contrast to more traditional downward ratings, MSPRs are provided by raters from different organizational levels (e.g., supervisors, peers, and subordinates). Conceptually, the rationale is that raters from different levels provide unique performance information that would not be captured by relying on downward ratings alone (Borman, 1974). The feedback provided by raters from each source allows individuals to gain insight on a wide range of performance behaviors that can subsequently be used to identify strengths and areas in need of improvement.

A multitude of studies have been devoted to estimating the magnitude of agreement from raters at different levels (Conway & Huffcutt, 1997; Conway, Lombardo, & Sanders, 2001; Mabe & West, 1982; Harris & Schaubroeck, 1988; Viswesvaran, Schmidt, & Ones, 2002) and identifying the etiology of cross-source disagreement (Bynum, Hoffman, Meade, & Gentry, under review; Cheung, 1999; Diefendorff, Silverman, & Greguras 2005; Facteau & Craig, 2001; Gillespie, 2005; Greguras, 2005; Hannum, 2007; Hoffman, Lance, Bynum, & Gentry, 2010; Lance & Bennett, 1997; Maurer, Raju, & Collins, 1998; Mount, Judge, Scullen, Sytsma, & Hezlett, 1998; Scullen, Mount, & Goff; 2000; Viswesvaran, Schmidt, & Ones, 2005; Woehr, Sheehan, & Bennett, 2005). Research has shown moderate agreement between peers and supervisors, supervisors and subordinates, and peers and subordinates; and relatively low agreement between self and others. Borman (1974; 1997) argues that high interrater agreement

between sources should be neither desired nor expected. In support of these suggestions, recent research indicates that observed disagreement may be due to raters viewing different aspects of job performance or perceiving job behaviors differently (Borman, 1974; Campbell & Lee, 1988; Facteau & Craig, 2001) and that the addition of alternative raters provides incremental validity over supervisor's ratings (Conway & Huffcutt, 1997; Conway, et. al., 2001).

On a separate front, with large discrepancies between self and other ratings, concern has increased over the substantive and practical implications of self-other disagreement (Ashford, 1989; Atwater & Brett, 2005; Bass & Yammario, 1991; Brutus, Fleenor, & McCauley, 1999; Church, 1997; Fletcher, 1997; Johnson, & Ferstl, 1999; London & Smither, 1995; McCall & Lombardo, 1983; Ostroff, Atwater, & Feinberg, 2004; Roush & Atwater, 1992; Van Velsor, Taylor, Leslie, 1993; Yammarino & Atwater, 1993). Disagreement between self and others has been linked to negative performance outcomes such as derailment (McCall & Lombardo, 1983), decreased likelihood of promotion (Bass & Yammario, 1991), career failure (London & Smither, 1995), difficulty navigating organizational politics (Ashford, 1989; London & Smither, 1995), and difficulty managing work relationships (Fletcher, 1997).

In conjunction, the evidence for the unique performance perspectives provided by different rater sources and the substantive implications of rater disagreement from self-other agreement studies underscores the importance of investigating the substantive implications of cross source disagreement. However, limited research has investigated the organizational and individual outcomes resulting from disagreement between raters from different organizational levels. Prior research has considered bivariate relationships between each source's ratings and outcomes (i.e. Barrick, Parks, & Mount, 2005; Blair, Hoffman, & Helland, 2008; Conway, et al., 2001; Hoffman & Woehr, 2009; Vecchio & Anderson, 2009). However, using a bivariate

approach does not allow for investigation of the pattern of disagreement between each source and subsequent outcomes of multiple, and potentially divergent perspectives of performance. In the last decade organizational researchers have called for using an integrated person-centered pattern approach for understand dynamic relationships when multiple processes are involved (Bergman, Magnusson, & El-Khouri, 2003; Foti & Hauenstein, 2007; Katzell, 1994; Yukl, 2006; Zaccaro, 2007; Mumford, Zaccaro, Johnson, Diana, Gilbert, & Threlfall, 2000). A person-centered approach would cluster individuals based on a pattern of responses on multiple variables. In a MSPR context, ratees would be clustered based on the pattern of ratings received from their supervisors, subordinates, and peers. To date, no research has used a person-centered approach to investigate the systematic patterns of agreement and disagreement among supervisor, peer, and subordinates. Moving beyond bivariate relationships to pattern-based approaches can give an indication of the antecedents (e.g., personality) and consequences (e.g., derailment, unit effectiveness) of particular patterns of ratings beyond which can be understood by investigating bivariate relationships alone. This information can potentially be instrumental in helping individuals interpret the frequently discrepant ratings and helping organizations identify at risk or high potential employees.

The lack of research devoted to examining the interrelationship of rater source disagreement has severely limited understanding of the practical implications of using MSPRs, representing a critical gap in the literature. The current study is the first to integrate a person-centered approach to understand distinct patterns of ratings across sources. In doing so, the current study has three aims: (a) identify whether there are systematic patterns of cross source ratings using a latent clustering pattern approach, (b) examine ratee personality as antecedents of

rating patterns, and (c) investigate potential outcomes of rater patterns by determining if the likelihood of derailment is more prevalent when rater sources disagree.

CHAPTER 2

PATTERNS OF MULTISOURCE PERFORMANCE RATINGS

Cross Source Agreement

Substantial research has examined rater agreement across different organizational levels and rater agreement within each organizational level (i.e. Borman, 1974; Cheung, 1999; Conway & Huffcutt, 1997; Diefendorff, Silverman, & Greguras 2005; Facteau & Craig, 2001; Harris & Schaubroeck, 1988; Maurer, et al., 1998; Mabe & West, 1982; Woehr et al., 2005; Viswesvaran, et al., 2002). Early researchers suggested low to moderate agreement among peer and supervisor raters (Borman, 1974) and low correlations between self and others' ratings (Mabe & West, 1982). More recent meta-analytical reviews, have shown stronger relationship between supervisor and peer ratings (r=.34-.64) and relatively weak relationships across other organizational levels with correlations ranging from .14 (self-subordinate ratings) to .22 (peer-subordinate ratings and supervisor-subordinate ratings; Conway & Huffcutt, 1997; Harris & Schaubroeck, 1988; Viswesvaran, et al., 2002).

Overall, there is strong consensus that raters tend to disagree. In accordance, there have been several theories proposed for why raters disagree. First, job performance is multidimensional, and as such, supervisors may not have the opportunity to view all aspects and behaviors of an individual's job performance. Raters from different organizational level may view different dimensions of performance and, therefore, may be more appropriate to evaluate certain performance dimensions (Borman, 1974). Second, even if each rater is able to view all aspects of job performance, raters may perceive, interpret, and evaluate behaviors differently

(Facteau & Craig, 2001; Campbell & Lee, 1988). Lastly, from an ecological perspective, raters from different organizational level may differentially perceive and evaluate behaviors based on their different behavioral affordances and subsequent interaction goals (Lance, Baxter, & Mahan, 2006; Lance, Hoffman, Gentry, & Baranik, 2008). Despite differences in the causes and meaning of disagreement from cross level raters, the central assumption of each of these theories is that raters from different levels provide unique performance information.

Methodological approaches to understand disagreement. A variety of methodological approaches have been utilized to delineate the underlying cause of the observed disagreement between sources. One line of research proposes that raters disagree because raters from different organizational levels conceptualize job performance differently (Cheung, 1999; Diefendorff, Silverman, & Greguras 2005; Facteau & Craig, 2001; Maurer, et al., 1998; Woehr et al., 2005). Using measurement invariance models and item response theory differential item functioning analyses, researchers examined the equivalence of raters from different sources. The majority of this line of research suggests that raters from different organizational levels view job performance similarly (Cheung, 1999; Diefendorff, Silverman, & Greguras 2005; Facteau & Craig, 2001; Maurer, et al., 1998; Woehr et al., 2005). However, more recent work (Bynum, Hoffman, Meade, & Gentry, under review) has provided some evidence for the different conceptualization hypothesis by demonstrating that although cross level raters conceptualize performance similarly, in an absolute sense, raters from the same level conceptualize performance more similarly than do cross level raters.

On a different front, multitrait-multimethod (MTMM) analyses have been used to estimate the proportion of variance that is attributable to performance dimension factors, source effects, and idiosyncratic rater effects as a means to investigating the substantive meaning of

rater method effects (Mount, et al., 1998; Hoffman, et al., 2010). Overall, this line of research suggests that performance dimensions account for a relatively small proportion of variance in ratings and idiosyncratic rater effects represent a relatively large proportion of variance in ratings. However, controversy had risen over the magnitude and importance of rater source effects (Hoffman et al; 2010; Mount et al., 1998; Viswesvaran et al., 2002; Viswesvaran et al., 2005). What is clear from this line of research is that both source effects and performance dimensions account for meaningful proportions of variance in ratings, and source effects account for more variance than performance dimensions. Overall, these results supported systematic source effects, suggesting that at least a portion of observed disagreement between sources is due to source specific variance.

Finally, a growing body of research has examined the degree to which different sources' ratings explain unique variance in outcome variables. In a meta-analysis, Conway, et al. (2001) found that supervisor, peer and subordinate raters provided incremental validity in explaining objective performance, suggesting that raters from different sources provide valid and unique information about performance. Oh and Berry (2009) showed that the relationship between personality and performance increased by 50% when additional sources were included as indicators of performance over traditional downward ratings (supervisor only). They suggest that the "low validities of personality traits may be overcome when more diverse and comprehensive rating perspectives are used in measuring performance" (pg. 1507). Hoffman and Woehr (2009) found that rater sources were differentially correlated with individual differences, suggesting that raters from different sources capture unique perspectives on performance. Although somewhat limited, research in this area, provides initial support that different sources provide valid and incremental performance relevant information.

Although this literature has consistently been dogged by questions as to the value of collecting performance data from cross level raters, mounting evidence supports that different rater groups provided unique information and in doing so, underscores the value of MSPRs as an organizational tool. With this knowledge, it is critical to turn our attention to the meaning of cross source disagreement and the consequences of disagreement for individuals and organizations.

Implications of cross source disagreement. Stemming from consistently observed discrepancies between self and others' ratings, a wide body of research has focused on the substantive underpinnings and developmental implications of self-other disagreement. Research has shown that low agreement between self and others can lead to career failure (London & Smither, 1995), difficulty navigating political realties (Ashford, 1989; London & Smither, 1995), and difficulty managing work relationships (Fletcher, 1997). Additionally, inflated self-ratings have been linked to derailment (McCall & Lombardo, 1983), lower levels of performance, (Church, 1997; Van Velsor, Taylor, Leslie, 1993; Yammarino & Atwater, 1993), and decreased likelihood of promotion (Bass &Yammario, 1991). While it is well documented that divergent self and other ratings can have detrimental consequences on performance, there has been limited discussion on the impact of disagreement between raters from different sources on performance outcomes. The limited understanding of the appropriate interpretation and implications of divergent patterns of cross source ratings has been a consistent concern with respect to the interpretation of MSPRs (Vinson, 1996).

Multisource Performance Feedback and Cross Source Disagreement

MSPRs are used as a developmental feedback tool to provide information about strengths and weaknesses of performance. A successful MSPR system would lead to positive behavioral

changes based on feedback received. However, as a feedback process, there are multiple components that play into whether MSPRs successfully lead to positive changes. Landy and Farr (1980, 1983) outline five key components of the feedback process: giving feedback, processing feedback, acceptance of feedback, desire to respond to feedback and intentions to respond to feedback. In MSPR systems, each of these feedback processes plays a critical role in whether an individual uses the feedback to improve their performance. Landy and Farr (1980, 1983) suggest that feedback should clearly specify which behaviors should be changed. However, when performance feedback is given by multiple sources, inconsistent feedback is likely to frequently occur. When appropriate behavioral changes are unclear ratees may experience conflict and confusion on which behaviors to adopt, lowering their intentions to respond to feedback. Further, Ilgen et al. (1979) suggest that the more an individual sees performance feedback as an accurate portrayal of behaviors, the more likely they are to accept the feedback. In contrast, from an attribution framework (Weiner, 1974, 1986), individuals may view incongruent feedback as being inconsistent and thus, make an external, rather than an internal attribution. When ratees are presented with feedback that diverges, they may not know how to interpret the feedback, be dismissive of the incongruous feedback, become frustrated, and ultimately, dismiss the feedback process. More innocuous but no less detrimental to the usefulness of MSPRs, ratees may key on the wrong source's feedback when interpreting feedback and engaging in developmental activities.

Although MSPRs are built on the foundation that raters from different sources should provide diverse perspectives, if the inconsistencies in feedback characteristic of cross source ratings lead to an inability to make positive performance changes, then the usefulness of MSPRs is severely limited. Given the consistent evidence of divergent ratings from different sources,

coupled with the central role that different raters' perspectives play in the rationale behind the use of MSPRs, the limited research in this area represents a critical gap in the literature. A key first step in investigating inconsistencies in MSPRs is to examine the patterns of ratings given by raters from different organizational levels. In order to examine the interrelationship between rater disagreement and resulting outcomes, ratings from different organizational levels must be examined using an integrated approach where performance ratings from each rater source are examined together versus in isolation.

Pattern Approaches

Recently there has been an increasing interest in applying multivariate pattern approaches to understanding the influence of individual differences on behavior. In traditional variable or trait model approaches, each variable is examined as a discrete entity and the focus is on the contribution that each variable has on explaining the criterion variable of interest (Mumford, Zaccaro, Johnson, Diana, Gilbert, & Threlfall, 2000). This approach does not consider the interrelationship between traits or how traits interact to influence behaviors (Bergman, Magnusson, & El-Khouri, 2003; Yukl, 2006). A pattern approach takes into consideration the interaction among discrete constructs in order to gain a deeper understanding of the phenomenon under investigation. In contrast to traditional trait research, pattern approaches allow the complexity of the person and the complexity of the situation to be integrated and studied holistically (Foti & Hauenstein, 2007). Katzell (1994) suggests that using a pattern approach to comprehensively assess multiple behaviors may lead to better long-term predictions of performance. Additionally, Zaccaro (2007) urged leadership researchers to take an integrated approach to studying the combination of leadership characteristics and how they influence leadership behavior. Zaccaro asserts that, "The combination of traits and attributes, integrated in

conceptually meaningful ways, are more likely to predict leadership than additive or independent contributions of several single traits" (pg. 6).

For example, in traditional trait research, the relationship between intelligence, social skills, and performance would be examined by regressing performance on intelligence and social skills to determine if social skills predict performance over and beyond intelligence. However, Ferris, Witt, and Hochwater (2001) found that the relationship between social skills, intelligence and performance was better explained using an integrated approach. Their results suggest that higher performers exhibited high social skills and high intelligence. In contrast, individuals with high intelligence and low social skills performed similarly to individuals with low intelligence and low social skills. Interestingly, individuals with high social skills and low intelligence were the lowest performers. Here, greater insights were afforded by investigating the pattern of individual differences, rather than considering each variable in isolation.

There are several methodologies used to examine the interactive influence of multiple variables. Moderated regression is the most common method for understanding the interactive effects of variables by investigating the interaction among two or more variables in predicting an outcome variable. Ferris, Witt, and Hochwater's (2001) examination of performance, intelligence and social skill is illustrative of the application of moderated regression to understand the simultaneous influence of multiple constructs on performance. Multivariate regression is an alternative regression based approach to examine the pattern of multiple dependent variables. Ostroff et al. (2004) used this approach to examine self-other agreement. Multivariate regression was used to determine if background characteristics predicted self and other ratings jointly. Although interaction and multivariate approaches are informative, there are several limitations to these methods. First, only a limited number of variables can be included in the analyses. Using

moderated regression, the interactive effects of four traits would include each trait, each two-way interaction, each three-way interaction, and the four-way interaction. The interactions are likely to be collinear, leading to difficulties delineating any meaningful results (Hair, Anderson, Tatham, & Black, 1998; Jaccard, Turrisi, & Wan, 1990) or finding statistically significant results (Stone-Romero & Anderson, 1994). Similarly, multivariate regression is limited in the number of dependent variables that can be simultaneously investigated.

Polynomial regression is a third technique that has been used to examine patterns of multiple variables. Polynomial regression is used as an alternative to difference scores to examine the congruence between two variables of interest. Polynomial regression has been used to investigate person-environment fit (Yang, Levine, Smith, Ispas, & Rossi, 2008), value congruence (Edwards & Cable, 2009), job complexity (Edwards & Parry, 1993) and selection (Hom, Grifeth, Palich, & Bracker, 1999). Vecchio and Anderson (2009) used polynomial regression to investigate the influence of demographic variables and personality on self-other agreement. Polynomial regression is an important technique for examining congruence and predictors of congruence but only two predictors can be investigated simultaneously. Separate polynomial regression analyses must be run if one is interested in examining the congruence of more than two variables.

These methods are variable-centered approaches for understanding the relationship among multiple variables of interest. However, Bergman, et al. (2003) argues that since "[an] individual develops as an active participant in an integrated person-environment system" (p. 7) person-centered pattern approaches are necessary to understand the dynamic interrelationships among behaviors. Cluster analysis is a person-centered approach where people are classified into homogenous groups based on patterns of scores on multiple variables of interest (Foti &

Hauenstein, 2007). Once individuals are classified, the groups become the variable of interest. The systematic groups that are identified by the unique interactions among variables can be examined and predictors and outcomes of the pattern of behaviors can be studied.

Few areas of organizational psychology have applied person-centered approaches to understanding trait patterns. However, most organizational constructs are associated with a combination of behaviors and interactive environmental influences. Using a person-centered pattern approach to understand how individuals respond to and interact with environmental situations is critical to understanding behaviors. Pattern approaches have been used to investigate person-environment fit (Schneider, 1987; Schneider, Goldstein & Smith, 1995; Schneider & Schneider, 1994; Borman, 1991; Katzell, 1994), how managers spend their time (Stewart, 1984,1988; Oshagbemi, 1988), teacher job satisfaction (Oshagbemi, 1997), leadership philosophies (Korac-Kakabadse, Korc-Kakabadse, & Myers, 1998) and biodata (i.e. Owens & Schoenfeldt, 1979). Contemporary leadership researchers have used cluster analyses to better understand leadership effectiveness, leadership behaviors, and leadership development (Foti & Hauenstein, 2007; Mumford, et al., 2000; Hautaluoma, Dickinson, & Inada, 2001; Oshagbemi & Ocholi, 2005; Smith & Foti, 1998, Zaccaro, 2007). Mumford et al. (2000) used cluster analyses to investigate the relationship among leadership characteristics and advancement. They examined managers' scores on ability, personality, and motivation measures and whether discrete profiles of related variables emerged in high versus lower level managers. They found that individuals who were high on extroversion, responsibility, achievement, dominance, verbal reasoning, generation, and revision, but low on intuition, feeling and perception were more likely to be in high ranking positions. When each variable was examined discretely, no strong relationships emerged. In contrast, when the variables were examined interactively, strong

relationships between individual differences and performance were found. Hautaluoma, Dickinson, and Inada (2001) used a cluster analyses to classify managers with similar trait patterns on the California Psychological Inventory (CPI). They used the patterns to investigate the relationship among personality traits and biographical information, job function and managerial style. They identified six personality types based on mean responses on six CPI trait measures. They found differences among personality types and family background, educational background, physical health, social comparisons, managerial styles, and time spent on job functions. These results substantiate the value of person-centered approaches to understanding organizational phenomenon.

The Present Study

Although a great deal of MSPR research has focused on understanding rater disagreement, there have been no studies that integrate and examine ratings from different sources using a person-centered approach. Despite widespread acknowledgement of divergent ratings across sources and growing recognition of the validity of these unique perspectives, research has not considered whether systematic patterns characterize MSPRs and if so, the implication of these patterns on individuals and organizations. A person-centered pattern approach is an ideal methodology for understanding the interrelationships between raters from different sources and will provide key practical insights to the appropriate use and interpretation of MSPRs.

Using a person-centered pattern approach, each ratee would be classified into a homogenous group based on the pattern of ratings from their supervisors, peers, and subordinates. One group may be composed of high ratings from all sources, while another group may be composed of high supervisor and peer ratings, but low subordinate ratings. The initial

classification of ratees would provide information as to whether there are systematic patterns that characterize subgroups of ratees. If subgroups of ratees emerge with similar cross source performance profiles, then the antecedents and consequences of particular rating patterns can be examined.

Expected patterns of MSPRs. Person-centered patterns approaches are exploratory in nature; as a result, no a priori hypothesis will be put forth. In addition to not knowing the specific profile patterns that will emerge, the number of groups with distinct cross source performance profiles is unknown as well. Multiple profile solutions will be compared to determine the number of homogenous groups that best represent the population. However, there are two cross-source performance profiles that are likely to emerge. First, raters could all agree and in-agreement raters could further be subdivided into two distinct groups. The first group would include ratees that are rated high by all raters and the second would include ratees that are rated low by all raters. Ratees that receive high ratings across rater groups are expected to be the outstanding performers with high potential and those that receive consistently low ratings would be deemed an at risk group. On the flip side, there are several combinations of rater disagreement that could emerge as distinct profiles. For example, rater patterns with high supervisor and peer agreement accompanied by divergent subordinate ratings could emerge, manifesting in either subordinates providing higher ratings than peers and supervisors or subordinates providing lower ratings than peers and supervisors. Table 2.1 presents eight distinct cross source performance profiles with agreement between at least two sources. In addition to the two in-agreement pattern profiles, a subset of the profiles presented in Table 2.1 are likely to emerge as distinct groups.

Research Question 1: Do systematic patterns of ratings characterize MSPRs?

Self ratings. A plethora of research has examined self ratings in relationship to others' ratings. The specific focus of this paper and the primary contribution is the examination of the pattern of agreement between supervisors, peer, and subordinate raters. However, once a profile solution is determine, average self-ratings will be examined for each distinct cross source performance profile to determine whether self-ratings play an informative role in the pattern of others' ratings. Although no specific hypothesis will be made in regards to self ratings, a research question is proposed as to the relationship between self ratings and cross source performance profiles.

Research Question 2: Are there distinct patterns of self-ratings that correspond to patterns of agreement between supervisors, peers, and subordinates?

Performance dimensions. Performance is typically examined using multiple dimensions. Consistent with past theoretical (Borman & Brush, 1993) and empirical work (Carty, 2003; Hoffman & Woehr, 2009, Lombardo & McCauley, 1994; Lombardo, McCauley, McDonald-Mann, & Leslie, 1999; McCauley & Lombardo, 1990; McCauley, Lombardo, & Usher, 1989; Spangler, 2003; Zedeck, 1995) on the nature of managerial performance, the MSPRs used in this study is composed of three performance dimensions: Leadership, Interpersonal Dealings, and Technical Activities. The first set of analysis, corresponding to *Research Question 1*, will examine cross source performance profiles for each performance dimension separately. While some ratees may have the same cross source performance profile for each performance dimension, other ratees may have different profile pattern for each performance dimension. Prior research has shown that rater sources more readily attend to different performance dimensions. Specifically, research has shown that supervisors are more equipped to rate task performance related behaviors (Hoffman & Woehr, 2009; Viswesvaran, et al., 2005**).** Bynum, et al. (under

review) found that supervisors' general impressions were more influenced by task activities, while subordinates' general impressions were more influenced by leadership dimensions. As a result, the differences in cross-source performance profiles across performance dimensions may be meaningful and systematic cross-source performance profiles may emerge when ratings for each performance dimensions are considered together. A second set of analyses will be conducted where each ratee will be classified based on ratings provided by each rater source on each performance dimension.

Research Question 3: What cross source performance profiles emerge when ratees are classified using MSPRs on three performance dimensions?

Nomological Network of Cross Source Performance Profiles

Once ratees are classified, the groups become the focus of the analyses. A second advantage of using a person-centered pattern approach and the second purpose of this study is to investigate antecedents and consequences of the systematic rater pattern groups. Past research has examined the relationship between individual characteristics and self-other disagreement. However, research has been limited on examining predictors and outcomes of disagreement across all sources. Using a pattern approach, this paper will examine the relationship between individual characteristics and patterns of performance as measured by MSPRs; specifically, whether ratees personality characteristic predict rater patterns. Additionally, the extent to which rater patterns predict potential to derail will also be examined.

Personality. The influence that personality characteristics have on cross source performance profiles will be investigated. Although research over the past decade has substantiated the influence of personality on managerial effectiveness (Judge, Bono, Ilies, & Gerhardt, 2002), this research has typically only investigated individual source's performance

ratings in isolation. No research has examined the influence of ratee personality characteristics on cross source performance profiles. This paper takes a first step toward gaining insight on the characteristics of ratees that facilitate the receipt of either congruent or incongruent performance evaluations.

Prior research has examined the relationships between cross source performance ratings and personality characteristics. Vecchio and Anderson (2009) found a negative relationship between social dominance and peer and subordinate ratings but no relationship with supervisor ratings. They also found a positive relationship between social sensitivity and peer and subordinate ratings but no relationship with supervisor ratings. Blair, Hoffman, and Helland (2008) found a negative relationship between narcissism and supervisor ratings of both interpersonal performance and integrity but no relationship between subordinate ratings and narcissism. Barrick, Parks, and Mount (2005) found that although supervisor ratings were positively related to extraversion, agreeableness, and conscientiousness, peer ratings were unrelated with these constructs. Conway, Lombardo, and Sanders' (2001) meta-analysis found that that peer and supervisor ratings had similar relationships with 'getting ahead' personality traits (e.g., potency and achievement), while peer and subordinate ratings had similar relationships with interpersonal 'getting along' (e.g., agreeableness and affiliation) personality traits. In contrast, subordinate ratings had little similarities with supervisor ratings on any personality traits. Brutus, Fleenor, and McCauley (1999) examined differences in the predictability of ratee personality characteristics on the CPI. They found that dominance was a significant predictor of peer ratings and self ratings, but not a predictor of supervisor or subordinate ratings. Social presence and well-being significantly predicted subordinate ratings, but not self, peer, or supervisor ratings. These results suggest differential relationships between

ratings from different sources and personality characteristics, providing indirect evidence for the possibility that ratee personality will be related to cross source performance patterns.

Although somewhat limited, research has investigated the relationship between ratee personality characteristics and self-other agreement. Goffin and Anderson (2006) found that achievement and self-esteem were significantly related to differences in self and supervisor ratings and self and peer ratings. Specifically, ratees with high achievement levels and high self-esteem had higher self ratings compared to peer and supervisor ratings. Additionally, anxiety was related to differences between self and supervisor ratings. Ratees with high anxiety tended to have lower self-ratings compared to supervisor ratings. However, differences in self and subordinate ratings were not influenced by personality characteristics. Overall, their results show that personality characteristics can differentially influence agreement among different sources. As a result, it is likely that personality characteristics will predict cross source performance profiles.

Six personality characteristics on the CPI, correspond to dimensions associated with the Big Five Personality Inventory (Costa & McCrae, 1992; Gough & Bradley, 1996) will be investigated. Specifically, well-being (neuroticism), social presence (extraversion), empathy (openness), good impression (agreeableness) and achievement vs. conformance (conscientiousness) will be examined. Additionally, the CPI scale of dominance will also be investigated. Since there has been limited research on the investigation of personality characteristics and cross source rater agreement; this study will take an exploratory approach in investigating whether personality predicts cross source performance patterns.

Research Question 4: Does personality predict cross source performance profiles?

Derailment. Derailment is a term that describes managers who fail to reach the anticipated level of managerial achievement by being involuntarily plateaued, demoted or fired (Lombardo & McCauley, 1988; Lombardo, Ruderman, McCauley, 1988; Kovach, 1988). MSPRs may be an important tool for identifying potential derailment. Low performance ratings may be indicative of poor relationships and patterns of disagreement in performance ratings may be indicative of an individual's inability to maintain working relationships across different organizational levels. Several studies have identified characteristics that are likely to be attributed to managers who derail including inability to handle uncertainty, weak relationships with supervisors, and failure to maintain positive interpersonal relationships (Bray, Campbell & Grant, 1974; Hogan & Hogan, 2001; Lombardo & McCauley, 1988; McNally & Parry, 2002; Stoner, Ference, Warren & Christensen, 1980; Skinner & Sasser, 1978; Van Velsor & Leslie, 1995). Gabarro (1987) found that managers who tended to derail did not get along with two rater sources typically included in MSPR systems. Additionally, Hogan and Hogan (2001) found that the inability to balance interpersonal relationships across different organizational levels was related to derailment.

Cross source performance profiles may be helpful in identify ratees at risk for derailment. Further, potential to derail is likely to be different based on cross source performance profile patterns. Ratees who receive high performance ratings from each rater source will likely have the lowest ratings of potential to derail and ratees who receive low performance ratings from each rater source will likely have the highest potential to derail. However, the expected outcomes of derailment for the more inconsistent cross source performance profiles are more difficult to predict, especially given specific cross source performance profiles are unknown. Although no specific hypothesis will be made, two general trends are likely to emerge. Individuals with

profile patterns with low performance ratings from two sources or low performance ratings from supervisors will likely have higher potential to derail. First, if two rater sources view performance as inadequate, then the working relationships between these two sources are likely to be strained and may be more difficult to fix than if one working relationship is strained. Second, supervisors tend to hold the power with regard to promotion and raises. If supervisors view performance as inadequate then promotion is less likely and being demoted or fired is more likely. Again, an exploratory approach will be taking to investigate the relationship between cross source performance profiles and potential to derail.

Research Question 5: Does cross source performance profiles predict potential for derailment?

Overall performance effectiveness. Overall performance effectiveness is an overall assessment of general performance level provided by the supervisor and is akin to supervisor ratings use to determine pay and promotion decisions in a traditional downward rating system. Overall performance effectiveness ratings will be used to determine a rank order of cross-source performance profiles and to determine if there are significant differences between individuals who receive ratings with cross-source disagreement on overall performance. Individuals who receive high ratings from all sources should have the highest overall performance ratings and individuals who receive low ratings from all sources should have the lowest overall performance ratings. However, the distinction between performance levels may not be as clear for individuals who receive ratings that diverge among sources. An exploratory approach will be taken to assess overall performance effectiveness differences among cross-source performance profiles.

Research Question 6: Does cross source performance profiles predict overall performance effectiveness?

Table 2.1

Possible Cross Source Performance Profiles

Source	1	2	3	4	5	6	7	8
Supervisors	High	Low	High	Low	High	Low	Low	High
Peers	High	Low	High	Low	Low	High	High	Low
Subordinates	High	Low	Low	High	Low	High	Low	High

CHAPTER 3

METHODS

Participants

Participants included 3,723 managers who participated in the Center for Creative Leadership's leadership development program. The managers included in the sample are primarily Caucasian (71%), Males (62%) with a mean age of 41 years. Managers were primarily upper middle (51%) and middle management (35.4%) with moderate experience (53.6%). The majority of the managers were in the business sector (74.2%).

Measures

Multisource performance ratings. The Center for Creative Leadership's (2004) BENCHMARKS® is a well-validated multisource performance instrument that is primarily used for leadership development purposes (Carty, 2003; Lombardo & McCauley, 1994; Lombardo, McCauley, McDonald-Mann, & Leslie, 1999; McCauley & Lombardo, 1990; McCauley, Lombardo, & Usher, 1989; Spangler, 2003; Zedeck, 1995). BENCHMARKS® is comprised of 115 items and 16-sub dimensions that are intended to measure three broad dimensions of managerial behavior. For each item, managers were rated on a scale of 1 (*Not at all*) to 5 (*To a Very Large Extent*).

Previous studies have assessed the factor structure of the BENCHMARKS® and 13 of the sub dimensions fit onto three broad performance dimensions which map on to the performance taxonomy proposed by Borman and Brush (1993) including: leadership and supervision, technical activities and mechanics of management, and interpersonal dealings and

communication (e.g., Conway, 1999; Hoffman, et al., 2010). Classifications include: (a) *Leadership and Supervision*: Leading Employees, Confronting Problem Employees, Participative Management, and Change Management; (b) *Technical Activities/Mechanics of management*: Resourcefulness, Being a Quick Study, and Decisiveness; and (c) *Interpersonal Dealings and Communication:* Compassion and Sensitivity, Building and Mending Relationships, Straightforwardness and Composure, Self-Awareness, Putting People at Ease, and Differences Matter. For each manager, self-ratings and ratings from supervisors, peers and subordinates were collected.

Confirmatory Factor Analyses (CFA) was used to examine the factor structure of the MSPRs using ratings from one randomly selected boss, peer and subordinate (Table 3.1). Three models were tested: A 3-factor model with a general performance factor for each source (boss, peer, and subordinate), a 6-factor model with two performance dimensions (technical activities and interpersonal dealings and leadership combined to a single latent factor) for each source, and a 9-factor model with three performance dimensions for each source (technical activities, interpersonal dealings, and leadership). The nine factor model is consistent with prior research and applied uses of the BENCHMARKS® (Hoffman, et al., 2010; Lance, Hoffman et al., 2008; McCauley, & Lombardo, 1990). The 9-factor model fit significantly better than the 3-factor model ($\Delta\chi^2 = 10158.09$, $\Delta df = 33$, $\Delta CFI = .05$) and the 6-factor model ($\Delta\chi^2 = 2400.67$, $\Delta df = 21$, $\Delta CFI = .01$). Although the fit was only adequate, the model did not specify any cross source variance. This was omitted because convergence across sources is picked up and interpreted at the individual level in the pattern approach analyses. Thus, this model is only used to test a baseline to determine the dimension structure. Based on these analyses, three performance dimensions per source were retained for subsequent analyses.

Multiple ratings were collected from each source. Ratees with at least two raters from each source were included in the analyses. On average, each manager received ratings from two supervisors (*range:* 2-15), four peers (*range:* 2-18), and four subordinates (*range:* 2-25). To justify aggregation, interrater agreement was computed using $r_{wg(j)}$ (Lebreton & Senter, 2008, James, Demaree, & Wolf, 1993). $r_{wg(j)}$ values were computed for supervisors, peers, and subordinates for each of the three performance dimensions (technical activities, j = 3; leadership, j = 4; and interpersonal dealings, j = 6). For supervisors, $r_{wg(3,4,6)}$ values ranged from .90 -.93 among the three performance dimensions. For peers, $r_{wg(3,4,6)}$ values ranged from .85-.88 and, for subordinates, $r_{wg(3,4,6)}$ values ranged from .84-.88. $r_{wg(j)}$ values of .70 or higher are typically considered as justification for aggregation (Lance, Butts, & Micheals, 2006; Lebreton & Senter, 2008). As a result, ratings from each source were aggregated to produce three mean performance dimension rating for each source.

California psychological inventory. The California Personality Inventory 260 (CPI 260®; CPP, 2002) will be used to assess personality. The CPI 260® is a shortened version of the CPI and is intended to be used in work and organizational settings. The CPI 260® is composed of 260 items with 29 scales. In accordance with *Research Question 4*, a subset of the CPI 260® scales will be used to investigate the relationship among personality and patterns of rater agreement. Six scales will be used from the CPI 260®: Dominance, Social Presence, Empathy, Good Impression, Well-Being, and Achievement via Conformance. Dominance is a 32 item scale intended to measure pro-social interpersonal power and influence (α = .86). Social Presence is a 29 item scale intended to measure poise and comfort with attention and recognition (α = .65). Empathy is a 25 item scale intended to measure capacity to understand and respond to other's needs (α = .60). Good Impression is a 27 item scale intended to measure tact and positive self

presentation (α = .77). Well Being is a 20 item scale intended to measure over all sense of health and optimism (α = .76). Finally, Achievement via Conformance is a 29 item scale intended to measure motivation within organized settings (α = .76). Composite scores from each personality dimension will be used in the analyses.

Derailment. The Center for Creative Leadership's measure of derailment potential will be used to measure derailment. Derailment is a 40 item measure tapping into five broad dimensions including: (a) *Problems with Interpersonal Relationships* (10 items), (b) *Difficulty Changing or Adapting* (10 items), (c) *Difficulty Building and Leading a Team* (8 items), (d) *Failure to Meet Business Objectives* (7 items), and (e) *Too Narrow Functional Orientation* (5 items). Each rater provided ratings of derailment. However, to avoid the influence of common method variance, derailment scores from an upper level supervisor or, when not available, an immediate supervisor was used. MSPRs from the supervisor providing derailment ratings were not included in the analyses. CFAs were conducted to examine the structure of the derailment scale. The five derailment dimensions were examined as indicators of overall derailment potential. Consistent with previous research (Gentry, Mondore, & Cox, 2007; McCartney & Campbell, 2006) a single factor model of derailment was specified. The model fit the data adequately (χ^2 =684.88, *df* =5, CFI = .96, TLI = .91, RMSEA = .19, SRMR = .03; see Table 3.2). Based on these results, scores were aggregated to form a single derailment score which were used in subsequent analysis.

Overall performance effectiveness. Overall performance effectiveness was measured using a single-item, "How would you rate this person's performance in his/her present job?" Managers were rated on a scale of 1 (Among the worst) to 5 (Among the best). Each rater provided ratings of overall effectiveness, but similar to derailment, to avoid the influence of

common method variance, ratings from an upper level supervisor or, when not available, the immediate supervisor were used.

Data Analysis

A latent clustering approach was used to identify systematic patterns of ratings. Latent clustering techniques are used to identify unobserved heterogeneity within the population and to identify groups with similar patterns on the variables of interest. Subpopulation membership or the variables causing the heterogeneity is unknown, but group membership is inferred from the data (Lubke & Muthén, 2005). A class or group is defined by a pattern of conditional probabilities that indicate the chance that the observed variables take on a certain value. Classes are determined based on differences between mean, variance, and covariance estimate. The differences between groups are maximized and the similarities within groups are minimized to determine the most homogeneous groups. A convenient way of conceptualizing clustering approaches is using a factor analytic framework where people are the object of measurement, rather than items (Cattell, 1966). The results reflect groups of individuals with similar characteristics.

In contrast to traditional clustering techniques, latent class clustering techniques are model based (Pastor, Barron, Miller, & Davis, 2007). Specifically, latent class models are composed of a measurement model and a structural model. The measurement model is composed of observed dependent variables and categorical latent variables. The categorical latent variables represent classes of individuals and latent class indictors, or dependent variables, represent the measures on which classes differ. The structural portion of the model measures the relationship among the categorical latent variables, the relationship among the observed variables and relationships among the categorical latent variable and observed variables that are not used as

indicators of class memberships (Muthén & Muthén, 1998-2007). For the purposes of this study Latent Profile Analysis (LPA) was used. LPA is a latent clustering technique that uses continuous latent class indicators.

A LPA approach outlined by Pastor, et al. (2007) was used in the present study. LPA models were run using Mplus (version 5.21) EM algorithm maximum likelihood estimation. The latent class indicators were average supervisor, peer, and subordinate ratings for each of the three dimensions. The latent classes represent groups of ratees who have similar patterns of ratings from supervisor, peer, and subordinate raters. Figure 1 depicts the latent profile model. LPAs were initially conducted on half of the ratees and the other half of the sample was used to verify the model solution. LPAs were initially conducted with a class size of two, and class size was incrementally increased in successive steps until the best fitting model was determined. LPAs were conducted for each performance dimensions separately and then for all performance dimensions together. Analyses using logistic regression and ANOVA were used to investigate the relationship between class membership, personality, derailment, and overall performance effectiveness.

Table 3.1

CFA Performance Dimensions

	χ^2	*df*	CFI	TLI	RMSEA	SRMR	$\Delta\chi^2$	Δdf
1. 3-factors: one general performance factor for each rater	29830.48	1077	.85	.84	.08	.05	--	--
2. 6-factors: One factor for each rater, for two performance dimensions (Tech & IP/Lead)	22073.06	1065	.89	.88	.07	.04	--	--
1 vs. 2	--	--	--	--	--	--	7757.42*	12
3. 9-factors: One factor for each rater, for each performance dimension	19672.39	1044	.90	.90	.07	.04	--	--
1 vs. 2	--	--	--	--	--	--	10158.09*	33
2 vs. 3	--	--	--	--	--	--	2400.67*	21

Table 3.2

CFA Derailment

	χ^2	*df*	CFI	TLI	RESMA	SRMR
3. One factor model-using sub scales	684.88	5	.96	.91	.19	.03

CHAPTER 4

RESULTS

Latent Profile Analyses

Split sample. In order to cross-validate the results the full sample was randomly split into two samples. A uniform random sampling technique was used where each observation was assigned a number between zero and one. Observations that were assigned a number less than 0.5 were assigned to sample one and observations that were assigned a number greater than or equal to 0.5 were assigned to sample two. Sample one included 1865 individuals and sample two included 1859 individuals. Latent profile analyses were conducted using Sample 1 and results were confirmed using Sample 2. Finally, results from the two split samples were replicated using the full sample.

Decision criteria. Five criteria were used to determine the best class solution: (a) Sample size adjusted BIC, (b) Lo-Mendell-Rubin likelihood, (c) Entropy Statistic, (d) Classification Accuracy Tables, and (e) Theoretical class fit. Class solutions with lower Adjusted BICs are considered better fitting models. In addition, specific attention was paid to large drops in the adjusted BIC from the class solution with one less the number of classes (*k*-1) to the *k* class solution and the point at which the drop in adjusted BIC flattens out (Nylund, Asparouhov, Muthen, 2008). The Lo-Mendell-Rubin likelihood statistic (LMR) assesses the difference in likelihood estimates for the *k*-1 class solution and the *k* class solution (Lo, Mendell, & Rubin, 2001; Vuong, 1989). A significant LMR suggests that the *k* class solution fits better then the *k*-1 class solution. The entropy statistic is an indication of classification utility, values range from

zero to one with higher values showing clear delineation of classes (Celeux & Soromenho, 1996). There is no predetermined cutoff value that would suggest sufficient classification utility. As such, the entropy statistic is best used to compare between models (Pastor et al., 2007). The classification accuracy table gives the probability of being in the class assigned as well as other classes. High in-class values and low out-of-class values suggest more accurate solutions. Finally, mean values for each class were examined to determine if the class structure was theoretically plausible. Additionally, the class added in the *k* class solution compared to the *k*-1 class solution was examined to determine if the mean structure added information over and beyond other classes and whether the sample size was large enough to suggest the group was determined beyond chance.

All five criteria were conjunctively examined to determine the best class solution. The different fit criteria did not always point to the same class solution. As a result, emphasis was put on the theoretical class fit, specifically, whether additional classes were theoretically plausible and added meaningful information. The fit criteria were independently examined for each sample to determine the best fitting class solution. The best fitting model across both samples was replicated using the full sample. Fit criteria for the full sample were examined to confirm the final class solution.

Technical activities. For technical activities, the six-class solution fit best for the split samples and the full sample (Table 3). When comparing the five class solution to the six class solution, the additional class included 5% of the sample and was composed of high ratings from each source (N =198). The group in the five-class solution that split to add the additional group in the six-class solution had low values of classification accuracy suggesting that the group was composed of a heterogeneous sample. By splitting the groups the classification accuracy

increased. Examining the seven-factor solution, the group that was added included less than 1% of the sample (N = 22). Due to the small sample size of the additional group in the seven-class solution, the six class solution was retained. For the six class solution the LMR statistic suggested significantly better fit then the five class solution (LMR=144.66, $p < .001$), the adjusted BIC dropped in fit from the five class solution but leaved off at the seven class solution. Additionally, the classification accuracy table (Table 4.1) and entropy statistic (.70) indicated high classification utility.

The cross-source performance profiles for the six-class solution are presented in Table 4.2. Four classes emerged with patterns of cross-source agreement and two classes emerged with patterns of cross-source disagreement. The four in-agreement classes included a high performer group (Class 6) with mean boss, peer and subordinate ratings ranging from 4.47 to 4.54 and a mid-high performer group (Class 2) with mean ratings ranging from 4.10-4.19. Both of these groups' mean ratings were above the full sample mean ratings, ranging from 3.99-4.06. A mid-low performer group (Class 3) emerged with in-agreement ratings ranging from 3.74-3.93. Finally, a group with low ratings across all three sources (M = 3.15-3.46; Class 4) also emerged. The two groups with cross-source disagreement included a group with low subordinate ratings (M = 3.18) and average peer and supervisor ratings (M = 3.94, M = 3.87; Class 1) and a group with average subordinate and peer ratings (M = 4.12, M = 3.85) and low supervisor ratings (M = 2.96; Class 5). Figure 2 display the pattern of cross-source agreement and disagreement among the different classes for the technical activities dimension.

Leadership. For leadership the seven-class solution fit best for the split samples and the full sample (Table 4.3). When comparing the six-class solution to the seven class solution the addition of the seventh class included a group with 41 people, approximately 1% of the sample,

and was composed of individuals who received high subordinate ratings and low peer and supervisor ratings. Although a low numbers of individuals were classified into this additional group, the finding of a correspondence between peers and supervisors is consistent with the theoretical and empirical basis of MSPRs. Peers and supervisors are similar in terms of knowledge of job demands and technical aspects of performance, a proposition supported by higher levels of supervisor-peer agreement relative to raters form other vantage points. Accordingly, the seven-class solution was retained over the six-class solution. The eighth group (N =22) included less than 1% of the sample and overlapped substantially with an existing class in terms of class members and theoretical interpretation. As a result, the seven-class solution was retained over the eight-class solution. For the seven-class solution the LMR was significant (LMR=84.55, p = .02), the drop in the adjusted BIC leveled off, and both the entropy statistic (.73) and the classification accuracy table suggested good classification utility (Table 4.4).

The cross-source performance profiles for the leadership dimension can be found in Table 4.5. Of the seven classes, four emerged with cross-source agreement. A high performer group emerged with ratings ranging from 4.32-4.44 (Class 7). A mid-high performer group emerged with ratings ranging from 3.83-3.96 (Class 1). Both of these groups were above the mean ratings for the full sample (M =3.63-3.78). Additionally, a mid-low performer group emerged with ratings ranging from 3.41-3.62 (Class 4) and a low performer group emerged with rating ranging from 2.99-3.00 (Class 2). Three groups emerged with cross-source disagreement. Class 6 emerged with low subordinate ratings (M = 2.75) and average peer and supervisor ratings (M =3.59, M =3.69). Class 3 emerged with low supervisor ratings (M =2.40) and average peer and subordinate ratings (M =3.50, M =3.86). Finally, Class 5 emerged with low supervisor and

low peer ratings (M =3.16, M =2.29), but average subordinate ratings (M =3.62). Figure 3 display the pattern of cross-source agreement and disagreement among the different classes.

Interpersonal dealings. The five-class solution was determined to be the best fitting model for the interpersonal dealing performance dimension for the split samples and the full sample (Table 4.6). When comparing the four-class solution to the five-class solution the addition of the fifth class added a group composed of 4% (N = 141) of the sample. The additional class was composed of low supervisor ratings and high subordinate and peer ratings. Further, the classification accuracy table for the five class solution included high in-class probabilities and low out-of-class probabilities. Comparing the six-class solution to the five-class solution showed a drop in stability of the classification table. Additionally, the sixth class that was added to the solution included less than 1% of the total sample (N = 48) and overlapped with a group already in the five-class solution The five-class solution had a significant LMR statistic (LMR =132.91, p = .03), the drop in adjusted BIC leveled off after the five-class solution, and the entropy (.73) and classification accuracy table showed strong classification utility with in-class probability values ranging from .74-.85 (See Table 4.7).

The cross-source performance profiles for the five-class solution can be found in Table 4.8. Two classes emerged with cross-source agreement. A high performer group emerged (Class 5) with rating ranging from 4.22-4.30 and an average performer group (Class 4) emerged with ratings ranking from 3.76-3.88. Mean ratings for the full sample range from 3.82-3.93. Three groups emerged with cross-source disagreement. Class 3 includes low subordinate ratings (M = 2.95) and average supervisor and peer ratings (M = 3.92, 3.62). Class 1 emerged with low peer ratings (M = 2.82) and average supervisor and subordinate ratings (M = 3.66, 3.49). Finally, Class 2 emerged with overall low ratings (M = 3.01-3.43), but slightly higher peer and

subordinate ratings. Figure 4 depicts the cross-source agreement and disagreement among the five-class solution.

Self ratings. Mean self-ratings for each class for each performance dimension were computed. Results are presented in Table 4.2, 8 and 11. Generally, individuals in the high performer group tended to underrate while individuals in mid-low and low performing groups tended to overrate. No distinct patterns emerged in the class with cross-source disagreement; however, self-ratings were more similar to the highest rating source. For example, when individuals were classified into a class with high subordinate ratings and low supervisor and peer ratings, average self ratings were similar to the subordinate ratings. Overall, mean self-ratings for each class were very similar across all classes (technical activities: M = 3.91, SD = .03; leadership: M = 3.74, SD = .16; interpersonal dealings: M = 3.91, SD = .11). Self-ratings for each class are depicted in Figures 2, 3, and 4, for technical activities, leadership and interpersonal dealings.

Simultaneous investigation of all performance dimensions. LPA was conducted for the nine dimensions by source class indicators (three performance dimensions rated by three sources). Comparing the seven-class solution to the six-class solution showed an additional group with all low performance ratings. The additional seven-class group was comprised of 6% (N = 213) of the sample. Additionally the classification accuracy table was stable with high in-class probabilities and low out-of-class probabilities. As a result, the seven-class solution was retained over the six-class solution. The eight-class solution fit well; however, the addition of the eighth class added an overlapping group with a similar rater pattern. Additionally, there was a slight drop in the classification accuracy when the eighth class was added. As a result, based on the parsimony, the seven-class solution was retained over the eight-class solution. The goodness-

of-fit indices can be found in Table 4.9 and the classification accuracy table can be found in Table 4.10.

Similar to the individual performance dimensions, a high performer class emerged with high ratings by each source on all three performance dimensions (Class 7). A low performer group also emerged with low ratings by each source on all three performance dimensions (Class 6). Two mid-high performer classes emerged (Class 5 and Class 1). Class 5 had slightly higher supervisor ratings, and slightly lower peer and subordinate ratings. In contrast, Class 1 had slightly higher subordinate and peer ratings and slightly lower supervisor ratings. A low subordinate performance group (Class 4) and a low supervisor performance group (Class 2) emerged. Finally, a low peer group emerged (Class 3) with peer ratings three-fourths of a standard deviation lower than supervisors and subordinate ratings. Specific means for each cross source performance profile can be found in Table 4.11. Self ratings corresponding to each class and performance dimensions can be found in Table 4.12. Overall mean leadership ratings for all sources were lower than technical activities and interpersonal dealings ratings, resulting in less straight forward interpretation of the difference among source and performance dimensions ratings. Standardized mean values were computed and used to graph cross-source performance profiles. Standardized performance profiles are depicted in Figure 5.

Antecedents

Personality. Multinomial Logistic Regression was used to examine the relationship between personality and class membership. A substantial number of ratees had missing personality data resulting in a sample size of 856 for subsequent analyses. Sample sizes for each class are reported in Table 4.15, 4.17, 4.18, and 4.19. When examining multiple group comparisons there is a high likelihood of false positive error rates. As a result, alpha values were

adjusted to control for the False Discovery Rate (FDR) using the Benjamini-Hochberg technique (Thissen, Steinberg, & Kuang, 2002). Essentially, this procedure relies on a more stringent alpha value to reject the null hypotheses. The critical *p* value is adjusted based on the number of group comparisons.

Personality was a significant predictor of class membership for technical activities and leadership, but not for interpersonal dealings or the classes defined by all performance dimensions. Specifically, dominance and well-being were significant predictors of technical activities class membership and dominance, well-being, and social presence were significant predictors of leadership class membership. Empathy, Achievement via Conformance, and Good Impressions were not significant predictors of class membership for any of the dimension-based classes. Table 4.13 shows the overall likelihood ratio model fit statistic and significance values for each personality variable. Personality means for technical activities, leadership, interpersonal dealings and all performance dimensions can be found in Table 4.15, 4.17, 4.18, and 4.19 respectively.

For technical activities, parameter estimates, Wald-statistics, odds ratios, and FDR p-value can be found in Table 4.14. The high performer group (Class 6; M =25.39) had significantly higher dominance scores the mid-low performer group (Class 3; M = 23.46) and the low supervisor group (Class 5; M = 22.67). For well-being, the high performer group (Class 6; M = 18.09) had significantly higher scores than the mid-low performers (Class 3; M =16.89) and the low subordinate performers (Class 1; M =17.08).

For leadership, parameter estimates, Wald-statistics, odds ratios and FDR p-values can be found in Table 4.16. Similar to technical activities, the high performer group (Class 7; M =25.19) had higher dominance scores than the mid-high performers (Class 1; M = 24.25), the mid-low

performers (Class 4; M = 23.68), and the low subordinate performers (Class 6; M = 21.86). The high performers (Class 7; M = 18.21) also had lower social presence scores compared to mid-high performers (Class 1; M =19.43), the mid-low performers (Class 4; M =19.46), and the low performers (Class 2; M = 18.95). The high performer group (Class 7; M = 18.11) had significantly higher well-being scores compared to the mid-low performers (Class 4; M = 16.85), the low performers (Class 2; M = 16.50), and the low peer/low supervisor performers (Class 5; M = 12.60). The mid-high performers (Class 1; M = 17.77) had significantly higher well being scores than the mid-low performers (Class 4; M = 16.85) and the low peer/low supervisor performers (Class 5; M = 12.60).

Outcomes

Derailment. ANOVA was used to examine the relationship between class membership and potential to derail. Derailment potential was examined for each performance dimension separately. Results suggest that class memberships for Technical Activities ($F(5, 3612) = 101.24$, $p < .01$), Leadership ($F(6, 3611) = 96.82$, $p < .01$), Interpersonal Dealings ($F(4, 3613) = 182.84$, $p < .01$), and All Performance Dimensions ($F(6, 3611) = 137.88$, $p < .01$) are significant predictors of potential to derail. That is, the pattern of rating received from boss, peer and subordinates predicts one's potential to derail. Post Hoc comparisons using Boferroni test indicated mean differences among the classes for each performance dimension. Alpha values were adjusted to control for FDR using the Benjamini-Hochberg technique (Thissen, Steinberg, & Kuang, 2002). A full list of results can be found in Table 4.20 and mean potential to derail scores can be found in Table 4.21 for each performance dimension and each class.

For technical activities, the high performer group (Class 6) had the lowest potential to derail (M =1.26) and was significantly lower than the other five technical activities classes.

Similarly, the low performer group (Class 4) had the highest potential to derail (M = 2.31) and was significantly higher than the other five classes. The mid-high performer group (Class 2) had the second lowest potential to derail (M = 1.64) and was significantly different from the other five classes. The two groups with the most dispersed pattern of disagreement, the low subordinate rating group (Class 1) and the low supervisor rating group (Class 5), had similar potential to derail scores (M = 1.92 and M = 2.03, respectively), along with, the mid-low performer group (Class 3; M = 1.95). Although these groups were not significantly different from each other, they were significantly lower than the high performers and mid-high performers, and significantly higher than the low performers.

For leadership, the high performer group (Class 7) had the lowest potential to derail (M = 1.23) and was significantly different from the other six classes. The mid-high performer group (Class 1) had the second lowest potential to derail scores (M = 1.6) and similarly was significantly different from all other classes. The low subordinate group (Class 6), the low supervisor group (Class 3), the mid-low performer group (Class 4), and the low supervisor and low peer group (Class 5) had similar potential to derail scores that were not significantly different from each other (M = 1.92, 1.93, 1.95, 2.18, for Classes 6, 3, 4, and 5 respectively). The low performer group (Class 2) had the highest potential to derail scores (M = 2.26) and was significantly different from the other classes with the exception of the low supervisor and low peer group (Class 5).

Similarly, the high performer group (Class 5) of interpersonal dealings had the lowest potential to derail (M = 1.45) which was significantly different from all other performance classes. The average performer group (Class 4) had the second lowest potential to derail scores (M = 1.80) and were significantly different from the other classes. The low subordinate group

(Class 3), the low peer group (Class 1) and the low performer group had similar potential to derail scores (M = 2.14, 2.23, 2.23, respectively) that were not significantly different from each other.

Finally, the classes defined by all performance dimensions had similar potential to derail compared to the individual dimension classes. The high performer group (Class 7) had the lowest potential to derail (M = 1.31) and the mean was significantly different from all other classes. The mid-high performers groups (Class 5 and Class 1) had similar potential to derail scores (M = 1.59, M = 1.63) that were not significantly different from the high performer group. However, the mean potential to derail scores for these groups was significantly different from the lower performer classes (Class 2, 3, 4, 6). The low peer group (Class 3; M = 1.88) was significantly higher than the high (Class 7) and mid-high (Class 5, Class 1) performance groups and significantly lower than the low subordinate (Class 4; M = 2.02), low supervisor (Class 2; M = 2.06) and all low performer groups (Class 6; M = 2.33). The low subordinate and low supervisor groups had similar potential to derail scores that were not significantly different from each other, but were significantly different from the low performer group (Class 6) and the high, mid-high and low peer group. Finally, the low performer group (Class 6) had the highest potential to derail scores (M = 2.33) that were significantly lower than all other classes.

Overall performance effectiveness. ANOVA was used to examine the relationship between class membership and overall performance effectiveness. Class membership was a significant predictor of overall performance effectiveness for Technical Activates ($F(5, 3665) = 141.99$, $p < .01$), Leadership ($F(6, 3664) = 70.95$, $p < .01$), Interpersonal Dealings ($F(4, 3666) = 84.23$, $p < .01$) and the classes defined by all performance dimensions ($F(6, 3664) = 106.41$, $p < .01$). Post Hoc comparisons using Boferroni test indicated mean differences among the classes

for each performance dimension. Alpha values were adjusted to control for FDR using the Benjamini-Hochberg technique (Thissen, Steinberg, & Kuang, 2002). A full list of ANOVA results can be found in Table 4.22 and the mean overall performance effectiveness scores for each class can be found in Table 4.23.

There were similar group differences on overall performance effectiveness as for potential to derail. For technical activities, the high performer group (Class 6) had the highest performance effectiveness ratings (M = 4.85) while the low performer group (Class 4) had the lowest performance effectiveness ratings (M = 3.23). The mid-high performer group (Class 2) had the second highest ratings (M = 4.26). The low subordinate group (Class 1), mid-low performer group (Class 3), and low supervisor group (Class 5) all had similar performance effectiveness ratings (M = 3.76, 3.86, 3.68, respectively for Class 1, 3, and 5).

For leadership, the high performer group (Class 7) and mid-high performer group (Class 1) had the highest two ratings of performance effectiveness (M = 4.77, M = 4.35). The low performer group (Class 2) had the lowest effectiveness ratings (M = 3.52) that were significantly lower than all other class groups with the exception of the low supervisor/low peer performer group (Class 5; M = 3.60) and the low subordinate performer group (Class 6, M = 3.86). The low supervisor performer group (Class 3) and the mid-low performer group (Class 4) had similar ratings of effectiveness (M = 3.90, M = 3.93).

For interpersonal dealings, the high performer group (Class 5) had the highest ratings of effectiveness (M = 4.46) and the average performers (Class 4) had the second highest ratings of effectiveness (M = 4.09). The low subordinate (Class 3), low peer (Class 1) and low performer (Class 2) groups had similar ratings of effectiveness (M = 3.85, 3.70, 3.69, respectively).

Finally, for classes with all performance dimensions, the high performers (Class 7) had the highest ratings of effectiveness (M = 4.70). The mid-high performer groups (Class 5, Class 1) had similar ratings of effectiveness (M = 4.33, M = 4.31). The low peer group (Class 3) had performance effectiveness ratings (M = 4.06) significantly lower than the high and mid-high performer groups, but significantly higher than the low subordinate (Class 4), low supervisor (Class 2), and all low performer (Class 6) groups. The low subordinate and low supervisor groups had similar ratings of performance effectiveness (M = 3.80, M = 3.75). Finally, the all low performers had the lowest effectiveness ratings (M = 3.45).

Table 4.1

Technical Activities Goodness-of-Fit Indices - Six Class Solution

Sample 1					
Class Size	BIC	Adjusted BIC	Entropy	Lo-Mendell	p-value
2	5091.16	5059.39	0.57	557.18	< 0.001
3	4984.17	4939.69	0.60	132.71	0.119
4	4925.74	4868.56	0.67	85.71	0.64
5	4879.38	4809.49	0.66	74.03	0.008
6	**4852.94**	**4770.34**	**0.69**	**54.75**	**0.001**
7	4834.59	4739.28	0.72	46.92	0.093
8*	4835.86	4727.84	0.73	27.94	0.002
Sample 2					
Class Size	BIC	Adjusted BIC	Entropy	Lo-Mendell	p-value
2	5340.83	5309.06	0.56	541.61	< 0.001
3	5180.47	5136.00	0.66	184.34	0.001
4	5110.56	5053.38	0.72	96.80	0.102
5	5057.68	4987.78	0.76	80.32	0.071
6	**5010.96**	**4928.36**	**0.74**	**74.35**	**< 0.001**
7	4989.76	4894.45	0.76	49.65	0.027
8	4988.91	4880.90	0.78	29.97	0.058

Full Sample

Class Size	BIC	Adjusted BIC	Entropy	Lo-Mendell	p-value
2	10380.29	10348.51	0.56	1098.43	< 0.001
3	10093.02	10048.54	0.63	310.71	< 0.001
4	9966.50	9909.31	0.68	154.71	0.017
5	9854.11	9784.21	0.72	140.99	0.002
6	**9758.82**	**9676.21**	**0.70**	**144.66**	**< 0.001**
7	9698.96	9603.63	0.73	90.02	0.033
8	9662.11	9554.08	0.75	67.68	< 0.001

Note. *Sample 1, class 8 solution did not converge due to a class with N=1.

Table 4.2

Technical Activities Classification Accuracy –Six Class Solution

Classes	*1*	*2*	*3*	*4*	*5*	*6*
N	138	1973	1226	114	74	198
1	**.77**	.06	.15	.03	.00	.00
2	.01	**.80**	.13	.00	.00	.06
3	.04	.17	**.74**	.02	.02	.00
4	.04	.00	.11	**.83**	.03	.00
5	.00	.02	.18	.03	**.76**	.00
6	.00	.22	.00	.00	.00	**.78**

Note. Most likely latent class membership (Row) by latent class (Column)

Table 4.3

Technical Activities Class Means –Six Class Solution

Class	*Rater*	*Mean*	*N*	*Percent*
1	Boss	3.94	138	4%
	Peer	3.87		
	Subordinate	3.18		
	Self	3.97		
2	Boss	4.15	1973	53%
	Peer	4.10		
	Subordinate	4.19		
	Self	3.92		
3	Boss	3.78	1226	33%
	Peer	3.74		
	Subordinate	3.93		
	Self	3.89		
4	Boss	3.32	114	3%
	Peer	3.15		
	Subordinate	3.46		
	Self	3.92		
5	Boss	2.96	74	2%
	Peer	3.85		
	Subordinate	4.12		
	Self	3.87		
6	Boss	4.54	198	5%
	Peer	4.47		
	Subordinate	4.52		
	Self	3.91		

Table 4.4

Leadership Goodness-of-Fit Indices - Seven Class Solution

Sample 1					
Class Size	BIC	Adjusted BIC	Entropy	Lo-Mendell	p-value
2	6901.63	6869.86	0.52	504.16	< 0.001
3	6784.43	6739.95	0.65	142.60	0.239
4	6733.83	6676.64	0.69	78.13	0.025
5	6683.14	6613.25	0.70	78.22	0.190
6	6641.84	6559.24	0.70	69.14	0.071
7	**6629.82**	**6534.51**	**0.73**	**40.79**	**0.382**
8	6634.57	6526.55	0.74	24.56	0.052
Sample 2					
Class Size	BIC	Adjusted BIC	Entropy	Lo-Mendell	p-value
2	7030.23	6998.46	0.53	511.76	< 0.001
3	6880.45	6835.98	0.66	174.10	0.035
4	6798.83	6741.64	0.71	108.14	0.098
5	6768.63	6698.73	0.73	58.37	0.829
6	6745.04	6662.44	0.72	51.97	0.079
7	**6724.69**	**6629.38**	**0.74**	**48.83**	**0.407**
8	6717.01	6608.99	0.74	36.57	0.124

Full Sample

Class Size	BIC	Adjusted BIC	Entropy	Lo-Mendell	p-value
2	13893.38	13861.61	0.50	1004.60	< 0.001
3	13594.03	13549.54	0.65	322.44	0.004
4	13475.09	13417.89	0.69	147.35	0.003
5	13341.91	13272.00	0.71	134.98	0.255
6	13270.68	13188.06	0.70	101.05	0.265
7	**13216.45**	**13121.12**	**0.73**	**84.55**	**0.017**
8*	13182.64	13089.25	0.73	31.38	0.204

Note. *Model did not converge due to low sample size in one class.

Table 4.5

Leadership Classification Accuracy – Seven Class Solution

Classes	*1*	*2*	*3*	*4*	*5*	*6*	*7*
N	1752	101	75	1499	41	79	176
1	**.80**	.00	.00	.15	.00	.00	.05
2	.00	**.74**	.02	.17	.04	.04	.00
3	.02	.04	**.79**	.14	.01	.00	.00
4	.16	.03	.02	**.76**	.01	.03	.00
5	.00	.10	.00	.07	**.83**	.00	.00
6	.03	.04	.00	.19	.00	**.74**	.00
7	.20	.00	.00	.00	.00	.00	**.80**

Note. Most likely latent class membership (Row) by latent class (Column)

Table 4.6

Leadership Class Means – Seven Class Solution

Class	*Rater*	*Mean*	*N*	*Percent*
1	Boss	3.87	1752	47%
	Peer	3.83		
	Subordinate	3.96		
	Self	3.82		
2	Boss	3.00	101	3%
	Peer	2.99		
	Subordinate	2.99		
	Self	3.60		
3	Boss	2.40	75	2%
	Peer	3.50		
	Subordinate	3.86		
	Self	3.63		
4	Boss	3.50	1499	40%
	Peer	3.41		
	Subordinate	3.62		
	Self	3.70		
5	Boss	3.16	41	1%
	Peer	2.29		
	Subordinate	3.60		
	Self	3.69		

6	Boss	3.69	79	2%
	Peer	3.59		
	Subordinate	2.75		
	Self	3.68		
7	Boss	4.39	176	5%
	Peer	4.32		
	Subordinate	4.40		
	Self	4.06		

Table 4.7

Interpersonal Dealings Goodness-of-Fit Indices - Five Class Solution

Sample 1

Class Size	BIC	Adjusted BIC	Entropy	Lo-Mendell	p-value
2	6703.92	6672.15	0.67	860.08	< 0.001
3	6530.81	6486.33	0.64	196.70	< 0.001
4	6509.93	6452.75	0.70	49.36	0.257
5	**6459.22**	**6389.32**	**0.71**	**60.12**	**0.107**
6	6429.08	6346.48	0.70	58.32	0.321

Sample 1

Class Size	BIC	Adjusted BIC	Entropy	Lo-Mendell	p-value
2	6796.16	6764.40	0.62	755.33	< 0.001
3	6597.16	6552.68	0.67	221.75	< 0.001
4	6538.94	6481.75	0.74	85.49	0.038
5	**6454.27**	**6384.38**	**0.76**	**87.89**	**0.001**
6	6439.08	6356.48	0.69	43.84	0.190

Full Sample

Class Size	BIC	Adjusted BIC	Entropy	Lo-Mendell	p-value
2	13452.24	13420.47	0.64	1609.37	< 0.001
3	13052.03	13007.54	0.65	420.32	< 0.001
4	12945.81	12888.62	0.70	135.00	0.042
5	**12807.83**	**12738.03**	**0.73**	**132.91**	**0.029**
6	12736.81	12654.19	0.73	100.95	0.510

Table 4.8

Interpersonal Dealings Classification Accuracy – Five Class Solution

Classes	*1*	*2*	*3*	*4*	*5*
N	130	299	141	1993	1160
1	**.76**	.14	.03	.07	.00
2	.08	**.77**	.03	.11	.00
3	.03	.05	**.74**	.18	.00
4	.01	.03	.03	**.83**	.10
5	.00	.00	.00	.15	**.85**

Note. Most likely latent class membership (Row) by latent class (Column)

Table 4.9

Interpersonal Dealings Class Means – Five Class Solution

Class	*Rater*	*Mean*	*N*	*Percent*
1	Boss	3.66	130	3%
	Peer	2.82		
	Subordinate	3.49		
	Self	3.90		
2	Boss	3.01	299	8%
	Peer	3.31		
	Subordinate	3.43		
	Self	3.77		
3	Boss	3.92	141	4%
	Peer	3.62		
	Subordinate	2.95		
	Self	3.89		
4	Boss	3.88	1993	53%
	Peer	3.76		
	Subordinate	3.87		
	Self	3.92		
5	Boss	4.30	1160	31%
	Peer	4.22		
	Subordinate	4.29		
	Self	4.09		

Table 4.10

All Performance Dimensions Goodness-of-Fit Indices – Seven Class Solution

Sample 1

Class Size	BIC	Adjusted BIC	Entropy	Lo-Mendell	p-value
2	16900.16	16811.21	0.81	7398.09	< 0.001
3	15791.58	15670.85	0.82	2342.71	0.350
4	15023.19	14870.70	0.80	1783.53	0.162
5	14468.25	14283.98	0.80	1220.74	0.217
6	14063.26	13847.22	0.79	788.79	0.323
7	**13712.02**	**13464.22**	**0.80**	**887.70**	**0.195**
8	13457.46	13177.89	0.80	706.61	0.336

Class Size	BIC	Adjusted BIC	Entropy	Lo-Mendell	p-value
2	17274.53	17185.58	0.81	3697.29	< 0.001
3	16158.63	16037.90	0.83	1175.55	0.163
4	15255.31	15102.82	0.81	965.75	0.002
5	14705.82	14521.55	0.81	616.57	0.110
6	14302.52	14086.48	0.82	472.30	0.155
7	**13957.50**	**13709.69**	**0.82**	**414.77**	**0.464**
8	13610.76	13331.19	0.83	416.47	0.301

Full Sample

Class Size	BIC	Adjusted BIC	Entropy	Lo-Mendell	p-value
2	34046.69	33957.72	0.81	7398.09	< 0.001
3	31757.71	31636.97	0.82	2342.71	0.030
4	30034.71	29882.19	0.80	1783.53	< 0.001
5	28881.35	28697.05	0.80	1220.74	0.005
6	28165.19	27949.12	0.79	788.79	0.071
7	**27348.91**	**27101.06**	**0.80**	**887.70**	**0.121**
8	26715.93	26436.31	0.80	706.61	0.536

Table 4.11

All Performance Dimensions Classification Accuracy – Seven Class Solution

Classes	*1*	*2*	*3*	*4*	*5*	*6*	*7*
N	876	451	708	309	778	213	388
1	**.84**	.02	.03	.00	.07	.00	.04
2	.03	**.83**	.09	.01	.02	.01	.00
3	.04	.06	**.81**	.02	.06	.02	.00
4	.00	.01	.04	**.90**	.03	.02	.00
5	.07	.01	.06	.02	**.83**	.00	.01
6	.00	.03	.04	.04	.00	**.89**	.00
7	.08	.00	.00	.00	.02	.00	**.90**

Note. Most likely latent class membership (Row) by latent class (Column)

Table 4.12

All Performance Dimensions Class Means – Seven Class Solution

Class	*Rater*	*Mean*	*N*	*Percent*
1	Boss TA	4.04	876	24%
	Boss LD	3.77		
	Boss IP	4.02		
	Peer TA	4.05		
	Peer LD	3.78		
	Peer IP	4.00		
	Sub TA	4.36		
	Sub LD	4.17		
	Sub IP	4.34		
	Self TA	3.90		
	Self LD	3.73		
	Self IP	3.90		
2	Boss TA	3.51	451	12%
	Boss LD	3.04		
	Boss IP	3.31		
	Peer TA	3.86		
	Peer LD	3.52		
	Peer IP	3.73		
	Sub TA	4.07		
	Sub LD	3.80		

	Sub IP	3.96		
	Self TA	3.89		
	Self LD	3.89		
	Self IP	3.95		
3	Boss TA	3.98	708	19%
	Boss LD	3.64		
	Boss IP	3.85		
	Peer TA	3.72		
	Peer LD	3.30		
	Peer IP	3.46		
	Sub TA	4.03		
	Sub LD	3.71		
	Sub IP	3.85		
	Self TA	3.93		
	Self LD	3.82		
	Self IP	4.02		
4	Boss TA	3.93	309	8%
	Boss LD	3.55		
	Boss IP	3.77		
	Peer TA	3.83		
	Peer LD	3.47		
	Peer IP	3.62		
	Sub TA	3.39		

	Sub LD	2.97		
	Sub IP	3.14		
	Self TA	3.93		
	Self LD	3.82		
	Self IP	4.03		
5	Boss TA	4.13	778	21%
	Boss LD	3.83		
	Boss IP	4.04		
	Peer TA	3.91		
	Peer LD	3.61		
	Peer IP	3.80		
	Sub TA	4.13		
	Sub LD	3.87		
	Sub IP	4.11		
	Self TA	3.91		
	Self LD	3.75		
	Self IP	3.92		
6	Boss TA	3.54	213	6%
	Boss LD	3.05		
	Boss IP	3.26		
	Peer TA	3.40		
	Peer LD	2.90		
	Peer IP	3.05		

	Sub TA	3.63		
	Sub LD	3.26		
	Sub IP	3.31		
	Self TA	3.87		
	Self LD	3.78		
	Self IP	3.99		
7	Boss TA	4.49	388	10%
	Boss LD	4.33		
	Boss IP	4.51		
	Peer TA	4.39		
	Peer LD	4.21		
	Peer IP	4.36		
	Sub TA	4.48		
	Sub LD	4.32		
	Sub IP	4.46		
	Self TA	3.91		
	Self LD	3.79		
	Self IP	3.99		

Note. Self Ratings were not included in the LPA analyses. Self Ratings reflect the mean self rating for the respective class.

Table 4.13

Logistic Regression Overall Model Fit

	Likelihood Ratio	$\Delta\chi^2$	p-value	FDR p-value
Technical Activities				
Full Model	1962.23	---	---	---
Dominance	1975.93	13.70	0.02	0.05*
Social Presence	1973.82	11.60	0.04	0.06
Empathy	1964.54	2.31	0.80	0.80
Good Impression	1974.42	12.19	0.03	0.06
Well-Being	1979.38	17.15	<.01	0.03*
Achievement via Conformance	1965.88	3.65	0.60	0.72
Leadership				
Full Model	1898.24	---	---	---
Dominance	1914.07	15.83	0.01*	0.03*
Social Presence	1920.62	22.38	<.01*	<.01*
Empathy	1907.60	9.36	0.15	0.23
Good Impression	1905.51	7.27	0.30	0.36
Well-Being	1923.98	25.74	<.01	<.01*
Achievement via Conformance	1903.30	5.06	0.54	0.54

Interpersonal Dealings				
Full Model	1986.45	---	---	---
Dominance	1986.74	0.30	0.99	0.99
Social Presence	1992.31	5.87	0.21	0.46
Empathy	1989.94	3.49	0.48	0.72
Good Impression	1993.36	6.91	0.14	0.46
Well-Being	1992.03	5.58	0.23	0.46
Achievement via Conformance	1988.58	2.13	0.71	0.85
All Performance Dimensions				
Full Model	3178.87	---	---	---
Dominance	3101.89	7.02	0.32	0.38
Social Presence	3107.41	12.54	0.05	0.15
Empathy	3099.47	4.59	0.60	0.58
Good Impression	3103.17	8.30	0.22	0.32
Well-Being	3111.02	16.15	0.01	0.08
Achievement via Conformance	3106.40	11.53	0.07	0.15

Note. * Significant at alpha < .05

Table 4.14

Logistic Regression Results – Technical Activities

Referent Group	Comparison Group	Personality Variable	β	Wald	p-value	Odds Ratio	FDR p-value
Class 6	Class 1	Well-Being	-0.38	10.12	0.01	0.69	0.02
	Class 3	Dominance	-0.12	9.07	<.01	0.89	0.02
		Well-Being	-0.40	8.39	<.01	0.67	0.02
	Class 5	Dominance	-0.26	9.21	<.01	0.77	0.02

Table 4.15

Personality Means – Technical Activities

	Class 1	Class 2	Class 3	Class 4	Class 5	Class 6
N	26	439	276	18	12	85
Dominance	22.88	24.11	23.46	24.17	22.67	25.39
Social Presence	19.00	19.33	19.26	19.56	20.50	18.96
Empathy	14.88	15.60	15.38	15.50	15.33	15.59
Good Impression	16.19	16.05	15.07	16.78	14.00	16.67
Well-Being	16.42	17.52	16.89	16.83	17.08	18.09
Achievement via Conformance	22.42	22.95	22.50	23.39	22.92	23.71

Table 4.16

Logistic Regression Results - Leadership

Referent Group	Comparison Group	Personality Variable	β	Wald	p-value	Odds Ratio	FDR p-value
Class 7	Class 1	Dominance	-0.12	8.81	<.01	0.89	0.02
		Social Presence	0.21	13.29	<.01	1.23	<.01
	Class 2	Social Presence	0.33	9.36	<.01	1.40	0.02
		Well-Being	-0.36	7.47	0.01	0.69	0.03
	Class 4	Dominance	-0.12	8.96	<.01	0.89	0.02
		Social Presence	0.25	18.63	<.01	1.28	<.01
		Well-Being	-0.22	6.937	0.01	0.80	0.04
	Class 5	Well-Being	-0.59	8.84	<.01	0.56	0.02
	Class 6	Dominance	-0.23	7.38	0.01	0.80	0.03
Class 1	Class 4	Well-Being	-0.14	11.76	<.01	0.86	<.01
	Class 5	Well-Being	-0.51	7.76	<.01	0.60	0.04

Table 4.17

Personality Means – Leadership

	Class 1	Class 2	Class 3	Class 4	Class 5	Class 6	Class 7
N	376	20	11	358	5	14	72
Dominance	24.25	22.70	22.82	23.68	19.80	21.86	25.19
Social Presence	19.43	18.95	18.36	19.46	17.00	18.36	18.21
Empathy	15.75	14.85	15.73	15.26	15.00	16.00	15.50
Good Impression	16.34	16.10	15.09	14.95	11.60	16.00	17.29
Well-Being	17.77	16.50	16.91	16.85	12.60	16.43	18.11
Achievement via Conformance	23.08	23.45	22.91	22.44	21.60	22.64	23.87

Table 4.18

Personality Means – Interpersonal Dealings

	Class 1	Class 2	Class 3	Class 4	Class 5
N	34	83	32	431	276
Dominance	23.44	23.73	24.78	23.93	24.08
Social Presence	18.88	19.07	20.28	19.45	19.02
Empathy	15.06	15.07	15.78	15.45	15.75
Good Impression	13.88	15.45	15.38	15.40	16.76
Well-Being	16.24	16.98	17.53	17.17	17.77
Achievement via Conformance	22.06	22.98	23.00	22.63	23.30

Table 4.19

Personality Means – All Performance Dimensions

	Class 1	Class 2	Class 3	Class 4	Class 5	Class 6	Class 7
N	185	99	170	66	165	50	121
Dominance	23.96	24.32	23.30	23.67	24.17	22.96	24.96
Social Presence	19.43	19.71	19.25	19.29	19.38	18.88	18.78
Empathy	15.77	15.11	15.18	15.42	15.78	15.02	15.74
Good Impression	16.24	14.45	15.08	15.95	16.21	14.86	16.88
Well-Being	17.56	17.02	16.70	17.32	17.46	16.68	18.16
Achievement via Conformance	22.82	22.36	22.14	23.11	23.11	22.94	23.92

Table 4.20

ANOVA Results – Derailment

	F	df_b	df_w	*p-value*
Technical Activities	101.24	5	3612	< 0.001
Leadership	96.82	6	3611	< 0.001
Interpersonal Dealings	182.84	4	3613	< 0.001
All Performance Dimensions	137.88	6	3611	< 0.001

Table 4.21

Mean Potential to Derail Scores by Class

Classes	1	2	3	4	5	6	7
Technical Activities	1.92[246]	1.64[a]	1.95[246]	2.31[a]	2.03[246]	1.26[a]	***
Leadership	1.60[a]	2.26[13467]	1.93[127]	1.95[127]	2.18[17]	1.92[127]	1.23[a]
Interpersonal Dealings	2.23[45]	2.23[45]	2.14[45]	1.80[a]	1.45[a]	***	***
All Performance Dimensions	1.63[23467]	2.06[13567]	1.88[a]	2.02[13567]	1.59[23467]	2.33[a]	1.31[a]

Note. Superscript indicates classes that were significantly different based on False Discovery Rate p-value adjustment. [a] Classes significantly different from all other classes.

Table 4.22

ANOVA Results – Overall Performance Effectiveness

	F	df_b	df_w	*p-value*
Technical Activities	141.99	5	3665	< 0.001
Leadership	70.95	6	3664	< 0.001
Interpersonal Dealings	84.23	4	3666	< 0.001
All Performance Dimensions	106.41	6	3664	< 0.001

Table 4.23

Mean Overall Performance Effectiveness Scores by Class

Classes	1	2	3	4	5	6	7
Technical Activities	3.76^{235}	4.36^{a}	3.86^{246}	3.23^{a}	3.68^{246}	4.85^{a}	***
Leadership	4.35^{a}	3.52^{13467}	3.90^{127}	3.93^{127}	3.60^{17}	3.86^{127}	4.77^{a}
Interpersonal Dealings	3.70^{45}	3.69^{45}	3.85^{45}	4.09^{a}	4.46^{a}	***	***
All Performance Dimensions	4.31^{23467}	3.75^{13567}	4.06^{a}	3.80^{13567}	4.33^{23467}	3.45^{a}	4.70^{a}

Note. Superscript indicates classes that were significantly different based on False Discovery Rate p-value adjustment. [a] Classes significantly different from all other classes.

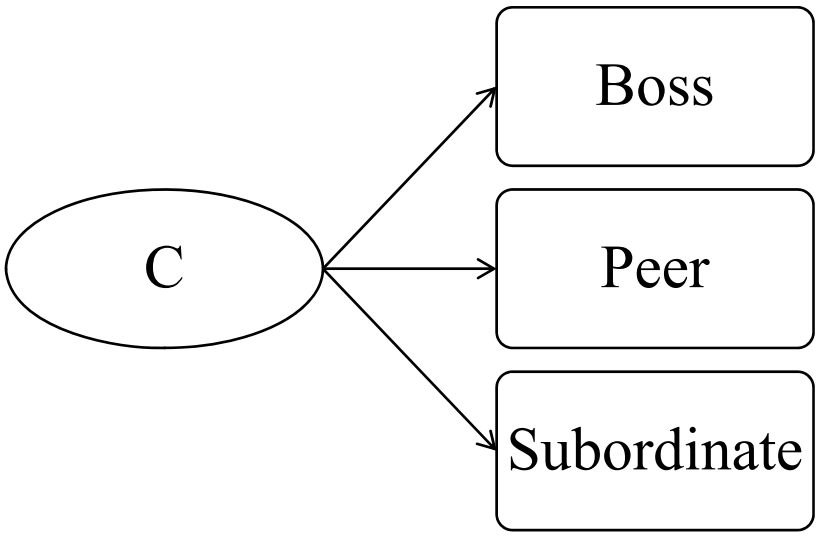

Figure 4.1: Latent Profile Analysis Model of Rater Patterns

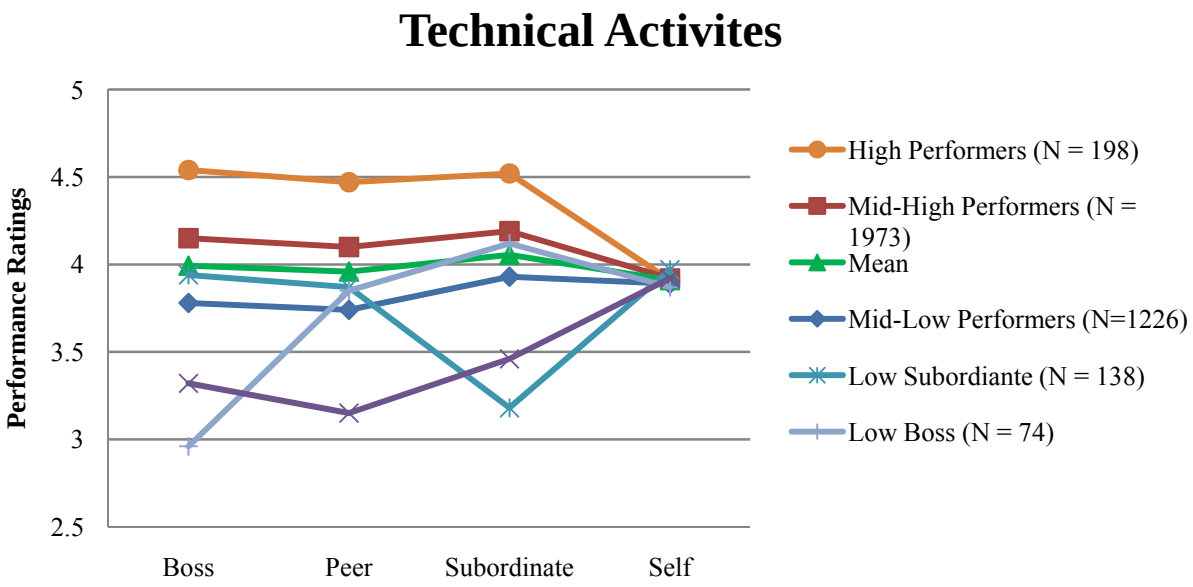

Figure 4.2: *Technical Activities Cross Source Means*

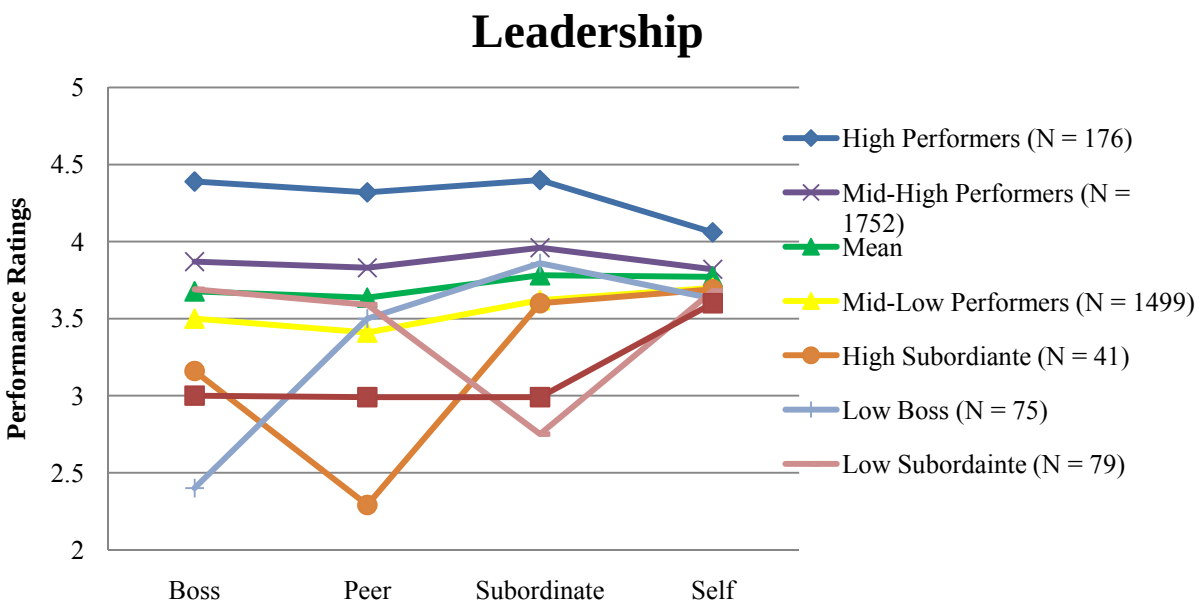

Figure 4.3: *Leadership Cross Source Means*

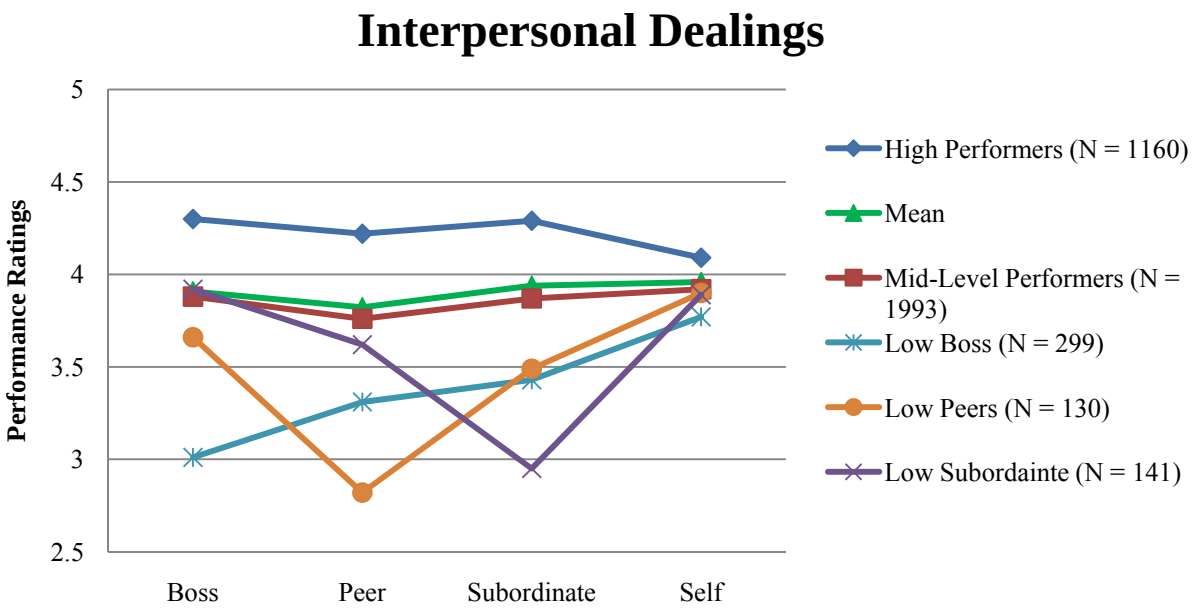

Figure 4.4: *Interpersonal Dealings Cross Source Means*

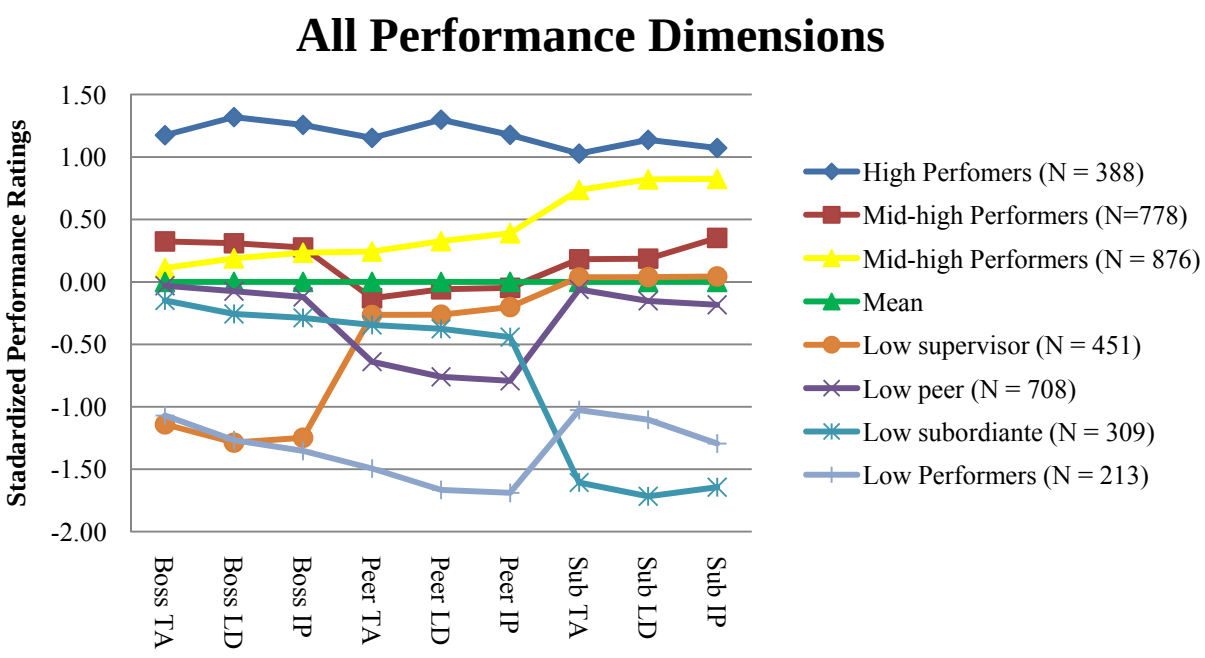

Figure 4.5: *All Performance Dimensions Cross Source Means*

CHAPTER 5

DISCUSSION

The present study contributes to the MSPRs literature by applying a person-centered approach to the investigation of cross-source agreement in an attempt to answer two primary questions. First, can managers be grouped based on the patterns of multisource ratings they receive? Second, is group membership predicted by personality, and is group membership a meaningful predictor of outcomes variables? The latent profile analyses suggest that managers can be meaningfully grouped. However, contrary to expectations, the majority of cross-source performance profiles that emerged were due to performance level and not differences in source ratings. Nevertheless, for a small number of individuals systematic patterns emerged that reveal homogeneous subsamples with ratings that diverge across sources. This study highlights the importance of using a person-centered approach for investigating MSPRs because the results substantially differ from prior studies examining rater agreement (Borman, 1974, Mabe & West, 1982, Conway & Huffcutt, 1997; Harris & Schaubroeck, 1988; Viswesvaran, et al., 2002). Consequently, this study has key implications for researchers and practitioners faced with interpreting and using the results of multisource feedback.

Performance Profiles

Groups characterized by cross source agreement. The analyses indicate the vast majority of ratees were clustered into groups characterized by consistent ratings across sources (e.g., high, medium, or low ratings for all three sources). This pattern of results held when investigating performance dimensions individually and when investigating all performance dimensions

simultaneously. Indeed, among the three performance dimensions, 93% - 95% of individuals received cross-source ratings within one-half of a standard deviation of each other. However, the differences in mean ratings of these groups differed on performance level.

For technical activities and leadership, approximately 5% of the sample received high performance ratings from all three sources and 3% of the sample received consistently low performance ratings. Additionally, a mid-high performer group and a mid-low performer group emerged comprising the majority of the sample (85% and 87%, respectively). These results suggest that the majority of individuals are within one standard deviation of the average performer with about half of those individuals falling above the mean and half falling below the mean, reflecting a normal population distribution, albeit on a truncated scale.

In contrast, for interpersonal dealings, approximately 31% of the sample received high performance ratings. Additionally, an average performer group emerged comprising the majority of the sample (54%) as well as a low performer group, representing 5% of the sample. The high proportion of managers that received high interpersonal ratings across sources may be indicative of the sample. The sample is composed of individuals already in management positions; in order to be promoted into a management position a threshold of interpersonal skills is likely required, leading to a larger proportion of individuals receiving high ratings. Mumford, Campion, and Morgeson (2007) found that interpersonal skills, compared to other leadership skills, were more important in junior level position. In conjunction with the current results, individuals with low levels of interpersonal skills are not likely to even be promoted in to low level management jobs.

Overall, these results suggested that for the majority of individuals, cross-source raters agree on performance level for a given ratee. These results are contrary to prior research that suggests raters tend to disagree (Borman, 1974, Mabe & West, 1982, Conway& Huffcutt, 1997;

Harris & Schaubroeck, 1988; Viswesvaran, et al., 2002). The current study shows that cross source raters agree on performance level and are able to reliably classify and rank order ratees based on performance. That is, the high performer group received mean ratings that did not overlap with any of the mean ratings received by the mid-high performer group. Similarly, the mid-low performer group received mean ratings that did not overlap with any of the mean ratings received by the low performer group, distinctly grouping ratees by performance level. The rank order of the groups is further illustrated in Figures 2, 3, and 4 by the lack of overlap in the line graphs for the high, medium and low performing groups. These results are surprising given that prior research has clearly acknowledged that raters from different sources do not agree (Conway & Huffcutt, 1997; Harris & Schaubroeck, 1988; Mount et al., 1998; Viswesvaran, et al., 2002, Viswesvaran et al., 2005). The current results not only suggesting that raters from different sources agree but that raters from different sources reliable distinguish among people with different performance levels.

Examining the correlations highlights that the current sample is not vastly different from prior MSPR samples. That is, for technical activities, the correlations for the full sample were .45 (supervisor and peers), .38 (peer and subordinate), and .30 (supervisor and subordinate). The correlations examining only classes with agreement were .48 (supervisor and peers), .43 (peer and subordinate), and .38 (supervisor and subordinate). The correlations including only classes that agree are not substantially different from the full sample or prior meta-analytical values of source agreement which suggest low to moderate agreement among sources (Conway & Huffcutt, 1997; Harris & Schaubroeck, 1988; Viswesvaran, et al., 2002). Why do the current results, using a person centered approach, suggest such vastly different interpretation?

There are a couple possible explanations for the discrepancies. First, for each individual, ratings were aggregated within sources. Several researchers have argued against this practice due to the large amount of variance associated with idiosyncratic rater effects (Mount et al., 1998; Viswesvaran et al., 2002; Viswesvaran et al., 2005). Although, $r_{wg(j)}$ results suggested sufficient agreement to aggregate within sources, aggregating confounds idiosyncratic rater effects and source effects, ignoring a substantial proportion of variance associated with MSPRs. Further, aggregating within sources can lead to regression to the mean for each aggregated rater source (Smith & Smith, 2005). If each aggregated source value moves toward the middle of the scale then the similarities across sources may be magnified. For the current study, slight differences among sources that appear in groups with cross-source agreement may appear larger if raters within a source were not aggregated. Future research should use a person-centered approach including idiosyncratic raters to determine if results differ from the current study.

Second, upon closer examination of the performance means for different sources within a given class, there are slight variations among supervisors, peers and subordinates. For example, for technical activities, the mid-low performer group received subordinate ratings one-half of a standard deviation higher than the peer and supervisor ratings. For the low performer group, subordinate and peer ratings were three-fourths of a standard deviation different. Overall both of these groups were distinctly different from other performance levels, but there appears to be a scale range within a performance level that produces the differences in source ratings. Statistically these differences may be significant, and without further comparisons of the other groups, these differences may be interpreted as disagreement. If source ratings within each group were compared in isolation with no further comparisons of the other groups, the interpretation

would be drastically different. However, when examining all classes together, the pragmatic differences based on performance level suggest substantial agreement between sources.

Groups characterized by disagreement across sources. Although the majority of individuals were characterized by similar patterns of ratings across sources, several groups emerged with divergent ratings among sources. Among performance dimensions, 5% - 7% of individuals received ratings with one source providing very divergent ratings from the other two sources. For each of the performance dimensions, a class emerged including low supervisor ratings and average peer and subordinate ratings and a class emerged including low subordinate ratings and average peer and supervisor ratings. For both classes, the low ratings were almost two standard deviations lower than the other two sources, indicating a substantial difference in perceptions of performance.

There were several interesting findings that emerged among the groups with disagreement across sources. Peer ratings for both of these classes were average but divergent in terms of which source peers more strongly agreed. For the low subordinate group, peers agreed more strongly with the supervisors, while peers more strongly agreed with the subordinates in the low supervisor group. For leadership, an additional class emerged with low supervisor and low peer ratings, but high subordinate ratings. For the technical activities and leadership dimensions, there were no classes that emerged with strong divergent peer ratings, suggesting that disagreement primarily manifests between supervisors and subordinates. These results more closely support prior finds that supervisors and subordinates have the lowest levels of agreement Conway & Huffcutt, 1997; Harris & Schaubroeck, 1988). Practically, managers who receive large discrepancies between supervisors and subordinates may experience role conflict with doing the most effective job for their supervisors and their subordinates. For example,

supervisors may have expectations such as cutting costs and managing a budget that is in direct conflict with the role a manager has with doing the most effective job for his or her subordinates, such as providing necessary supplies and resources.

For the interpersonal dealings dimension, a class emerged with substantially low peer ratings and below average subordinate and supervisor ratings. For the strong divergent peer ratings to emerge in interpersonal dealings dimension, supports prior theories that peers have more opportunity to observe interpersonal skills. Hoffman and Woher (2009) found that peers ratings were more influenced by manager interpersonal skills than were other source's ratings.

The lack of divergent peer ratings in the other two performance dimensions may be due to organizational structure. With no strong divergent peer ratings for technical activities and leadership, the results suggest that some ratees display divergent behaviors toward their supervisor and subordinates but not their peers. For some organizations, individuals who are peers to the ratee may have very infrequent contact with the ratee's subordinates or may have little contact with the ratee's supervisor. In line with theories that suggest that ratees act differently around different sources (Lance, Baxter, & Mahan, 2006; Lance, Hoffman, et al., 2008), these results appear to support that some ratees may act differently around their supervisors and subordinates, and peers are influenced by ratees' behaviors in the presence of the sources that the peers more often interact.

Regardless of the source that provided the low performance ratings, individuals who receive ratings with large disagreement between sources are likely those individuals who are unable to maintain positive relationships across all three sources. These individuals may be unable or unsure how to navigate the job requirements need to fulfill the expectations of coworkers from different sources, which is reflected in the low performance ratings.

Additionally, these individuals are among the lowest performers. Each group received below average ratings from all three sources, with one rating being well below average. Not surprisingly, individuals who balance relationships among different sources appear to be the best performers and have the highest potential to succeed.

Simultaneous investigation of all performance dimensions. The simultaneous investigation of all performance dimensions together revealed similar classes to those that emerged examining each performance dimension separately. A high performer group, low performer group and two mid-high performance groups, with slight variations in ratings, emerged. The two mid-high performance groups received above average ratings but for one group peer and subordinate ratings were slightly higher than the supervisor ratings and for the other group supervisor ratings were slightly higher than the peer ratings. The differences in source ratings across both groups were within one-half of a standard deviation. Additionally, three groups emerged with cross-source disagreement, including: a low supervisor group, a low subordinate group and a low peer group.

The proportion of individuals who received ratings with cross-source agreement was substantially less than the individuals who received ratings with agreement for the separate performance dimensions. Only 60% of individuals' received similar ratings across all sources. In contrast, 40% received ratings with cross-source disagreement, substantially higher than the 5% that received divergent ratings examining each performance dimension separately. It is surprising that the proportion of individuals receiving ratings with agreement versus disagreement is drastically different, but the performance profiles of the groups that emerged using all performance dimensions are very similar to the groups that emerged using each performance dimension separately. Overall, the results using all performance dimensions fall in line with prior

research suggesting that raters disagree (Borman, 1997; Conway & Huffcutt, 1997; Harris & Schaubroeck, 1988; Mount et al., 1998; Viswesvaran, et al., 2002, Viswesvaran et al., 2005), leading to nearly opposite interpretation of the usefulness of person-centered approach for examining cross-source disagreement.

One explanation for these differences is that performance dimension ratings given by each source influences cross-source performance profiles. However, the three performance dimension ratings provided by the same source are not notably different, suggesting that raters are using a general impression to rate all three performance dimension similarly. These results suggest that differences between performance profiles are due to the source and not difference in performance dimensions.

Overall the primary explanation for these differences could be due to the model assessment using the nine performance ratings as class indicators. Overall, there was better model fit for the latent profile analyses examining all performance dimensions simultaneously than for the analyses examining each performance dimension separately. The finally class solution had higher entropy statistic and high in-class probabilities in the classification accuracy tables, suggesting better classification utility than any of the solutions for the performance dimensions examined separately. Additionally, the sample size was more evenly distributed among the cross-source performance profiles. The proportion of individuals in each group ranged from 24% to 6% where as for the separate performance dimensions 80% of the individuals were classified into two groups. The LPA appears to be working better to identify homogenous groups because there is more information to use to group similar individuals together.

It is unclear why there are such vast differences when examining all performance dimensions simultaneously compared to the separate examination of each performance dimension or which interpretation is more accurate. More research needs to be done to assess the model impact of having more or less class indicators. These results highlight the need for further research to replicate these findings.

Self ratings. A great deal of MSPR literature has been devoted to self-other agreement (Ashford, 1989; Atwater & Brett, 2005; Bass &Yammario, 1991; Brutus, Fleenor, & McCauley, 1999; Church, 1997; Fletcher, 1997; Johnson, & Ferstl, 1999; London & Smither, 1995; McCall & Lombardo, 1983; Ostroff, Atwater, & Feinberg, 2004; Roush & Atwater, 1992; Van Velsor, Taylor, Leslie, 1993; Yammarino & Atwater, 1993). This literature has primarily looked at the agreement between self ratings and each source's ratings separately. The current study examined mean self-ratings for each performance profile, effectively examining agreement between self and others for all raters simultaneously. Overall the results appear to support prior research (i.e. Ostroff et al., 2004) that high performers typically underrated their performance, while low performers overrated their performance. Similarly, high and mid-high performers tend to rate themselves more closely to other raters.

However, it appears that everyone is rating themselves a four, regardless of performance level. Examining self-ratings for each cross-source performance profile highlights these similarities across performance levels that are not captured using differences scores and interaction methodology. Individuals who appear to be more self-aware are those individuals who received ratings of fours because they also rate themselves as four. These results suggest that self raters are not distinct across performance level, giving little justification for interpreting self-other findings as any one self-rater being any more self aware than another. Past research

suggests that high self-other agreement is a predictor of performance level. However because self-ratings are essentially constant, the current findings suggest that the correct interpretation is likely that the correlation between self awareness and self-other agreement is solely a function of the other's performance ratings. As a result, prior research examining the relationship between self-other agreement and performance may be erroneous. The current study does not investigate specific self-other indices, but these findings calls into question the finding between self-other agreement and performance level and dictates that more research needs to be done to examine self-other agreement using a pattern approach.

Nomological Network of Cross Source Performance Profiles

The second purpose of this study was to investigate antecedents and consequences associated with cross-source performance profiles. The relationships between class membership and personality, derailment and overall effectiveness also were based on performance level and not cross-source rater patterns.

Personality. Personality was assessed as a predictor of class membership. The total sample with personality information was 857, approximately 22% of the sample. The small sample size of some of the classes severally limits the interpretation of the personality findings. Six personality dimensions were examined, however only three dimensions emerged as significant predictors of class membership: Dominance, social presence, and well-being. Dominance and well-being were significant predictors of class membership for technical activities and dominance, well-being, and social presence were predictors of class membership for leadership. However, very few group differences emerged for either performance dimension.

For dominance, a few interesting comparisons are important to note. First, technical activities the high performers were more dominant than the mid-low performers, but were no

different than the low performers. These results appear to suggest that for technical activities, the difference between high and low performers on dominance is minimal. However, there may be a curvilinear relationship that is not able to be examined using the current methodology. The results clearly suggest that dominance is important for being a better task performer, but there is likely a point at which being too dominate, with some other combination of traits or situational variables, back fires and leads to low performance ratings.

For leadership, high performers had higher dominance scores than the mid-high, mid-low, and low subordinate groups. In contrast, for leadership the relationship between dominance and performance is linear. Anderson and Kilduff (2009) suggest that individuals who are dominant have more influence over a group because they are perceived as being more competent regardless of their actual ability. As a result, the strong relationship between dominance and high leadership is not surprising. Individuals who are viewed as dominate likely have more influence over a group, are viewed as being more competent and, in turn, receive higher leadership ratings.

For technical activities, high performers had higher dominance scores than the low supervisors group and for leadership the high performers had higher dominance scores than the low subordinate group. Results suggest that low dominance, or an inability to have prosocial interpersonal power and influence, affects subordinate views of leadership, while the same lack of power and influence affects supervisors' views of technical activities. These results highlight that the lack of certain personality characteristics can differentially influence ratings of different performance dimensions among sources. Subordinates view dominance as more important to leadership, which is typically viewed as the most influential performance dimension for subordinates' general impressions (Bynum et al., under review). In contrast, supervisors view dominance as more important for technical activities.

For well-being, a few interesting group comparisons emerged. First, for technical activities the high performers were more emotionally stable than the mid-low performers, but were no different than the low performers. There may be a similar curvilinear relationship between performance and well-being as was suggested for dominance. For leadership, the high performers had higher well-being than the three lowest performance groups (mid-low, low supervisor/low peer, and low performers). Further examination of the means shows a linear relationship between performance group and well-being. Such that as the performance level increases, well-being scores also increase. Being emotionally stable and optimistic appears critical for being viewed as a strong leader.

Social presence was a significant predictor for class membership only for leadership. In conjunction with being high dominant and high well being, high performers had lower social presence than mid-high, mid-low and low performance groups. Although the three groups with cross-source disagreement had similar overall performance levels as the mid-low performers, there were no distinctions on social presence between the high performers and the groups with cross-source disagreement. Examination of the social presence means in conjunction with the dominance and well-being means shows some interesting trends. For the groups with cross-source agreement, as performance increases, dominance increases, well-being increases and social presence decreases. However, when looking at the means for the groups with cross-source disagreement they appear to have low social presence, low dominance and low well-being. The general linear trend in personality associated with the increase in performance level seen in the cross-source agreement groups does not hold for these groups. Based on the mean personality trends, there appears to be an interactive relationship between dominance and social presence in the prediction of cross-source performance profiles. Several studies have used a person centered

approach to examine the interactive relationship between personality characteristics showing that different patterns and levels of personality characteristics are related to performance level (Mumford, et al., 2000; Hautaluoma, Dickinson, & Inada, 2001). Personality profiles may be better suited to predicting class membership with cross-source disagreement.

Outcomes of performance profiles. Two outcomes of class membership were examined: derailment and overall performance effectiveness. Results supported a linear relationship between class membership for both derailment and overall effectiveness. That is, the high performer group had the lowest potential to derail and the highest overall effectiveness scores. The mid-high performer group fell in line with the next lowest potential to derail and next highest effectiveness scores. The mid-low, low supervisor, and low subordinate performance groups had similar potential to derail scores and overall effectiveness scores, lower than the high and mid-high performance groups. Finally, the low performance group had the highest potential to derail and lowest overall effectiveness. This general linear trend was similar for each performance dimension.

The mid-low, low subordinate and low supervisor performance groups each received potential to derail and overall effectiveness scores that were below the mean. Generally, individuals who receive ratings with cross-source disagreement are perceived as ineffective and more likely to derail, and it appears to make little difference whether the source of lower ratings stems from supervisors or subordinates. In other words, these results suggest that individuals who receive below average ratings from each source are no better performers than individuals who receive low subordinate but average peer and supervisor ratings or those who receive low supervisor but average peer and subordinate ratings. These findings are interesting in that the rater who provided the derailment potential and overall performance effectiveness ratings was

from an independent upper level supervisor or a direct supervisor. It is surprising that the derailment potential and overall effectiveness ratings did not mirror the average MSPRs of the supervisor source. Specifically, for the low supervisor performance group, the supervisor's performance ratings were lower than any of the source ratings in the low performance group, but the supervisor who provided the overall effectiveness ratings rank ordered the low supervisor performance group higher than the low performance group. These results suggest that supervisors are able to identify an overall performance level of a ratee that is indicative of all three sources views of performance. Even though, the rank order differences from the supervisors views of performance.

The groups with only one source providing below average ratings were evaluated significantly more favorably than groups that were rated as less effective by all sources and groups that were rated as less effective by two sources. For instance, the leadership class with low supervisor, low peer and average subordinate ratings were no different in potential to derail or overall effectiveness than the low performance groups. These results support prior findings that individuals who receive two or more low ratings have strained relationship that may lead to higher potential to derail (Gabarro, 1987).

For analyses examining all performance dimensions simultaneously, the low peer group had significantly lower potential to derail than the low subordinate or low supervisor groups. These results suggest that maintaining positive relationships between peers maybe more important than having a strained relationship between either subordinates or supervisors.

Post-hoc analyses. Post-Hoc analyses were done to investigate whether individuals were grouped differently based on organizational level (ranging from hourly to executive level), organizational type (public, private or business sector), and level of experience (very

experienced, moderately experienced, or no experience). For the full population, there was a larger proportion of individuals working as a middle or upper middle manager ($\chi^2 = 4987.87$, $df = 5$), moderate experience ($\chi^2 = 751.12$, $df = 2$), and in the business sector ($\chi^2 = 2974.22$, $df = 2$). For most of the cross source performance profiles there were no proportion differences for organizational level, organizational type, or level of experience than those in the full sample. However, for organizational type there was a higher proportion of individuals in the public sector classified as high performers compared to the full sample. The full sample included approximately 20% of individuals from the public sector, the high performer groups for technical activities, leadership, and all performance dimensions examined simultaneously was composed of 72%, 61%, and 42% of individuals in the public sector, respectively. Public sector includes elementary, secondary, and higher education; government agency; health; and military. More research needs to be done to examine why these differences occurred.

Implications

Feedback. MSPRs are used as a feedback tool and research commonly shows that raters from different levels disagree, yet existing research has not considered the impact of receiving ratings with cross-source disagreement on the individual. Using a person-centered approach in this study allowed for examination of the impact of MSPRs on individuals receiving the ratings. First, the investigation of potential to derail as an outcome of patterns of ratings show that potential to derail was higher for low performing groups, and groups with divergent ratings among sources. It is important for practitioners using MSPRs to be aware that individuals who receive MSPRs with divergent pattern of ratings may be at risk for derailment. Furthermore, the analyses examining all performance dimension simultaneously highlights that individuals who receive low supervisor or subordinate ratings may have an even greater risk of derailment.

Borman (1997) suggest that to make behavioral changes, precise and focused knowledge about performance is needed. However, when appropriate performance behaviors are unclear, such as when ratees receive inconsistent feedback among sources, ratees may be more likely to dismiss feedback from divergent sources or experience conflict and confusion about the appropriate behaviors to change. Overall, the results suggest for feedback purposes most individuals will receive consistent ratings across sources allowing for straightforward reporting and interpretation. However, a subset of individuals received divergent feedback and for those individuals identifying the appropriate action to change their performance behaviors may be difficult. For practitioners, it is critical to identify individuals who receive divergent feedback and provide them with training to respond and clarify what behaviors should change to meet the performance expectations of all of the sources. Interestingly, individuals in the low supervisor and low subordinate performance groups received low supervisor or low subordinate ratings, yet received average ratings from the other two sources. If these individuals were able to make appropriate behavioral changes to increase performance ratings for that one source, their potential to derail may decrease and their overall effectiveness ratings may increase. Although the number of individuals receiving divergent ratings is small, a great deal of resources goes into hiring and training employees. Focusing additional resources to improve performance among these individuals could be far less costly than replacing these individuals because of derailment.

Multisource performance ratings. MSPRs are built on the notion that raters from different sources will provide unique perspectives on performance and, in order, to evaluate the full spectrum of performance behaviors additional sources are needed to provide performance ratings (Borman, 1974). However, the discrepancy of the current studies result with prior research begs the question: Why are organizations devoting resources to MSPR systems, if for

most individuals, raters agree on performance level? The current results seem to suggest that raters from different sources are not providing substantially more information about performance level than what would be gained from using traditional downward ratings systems. However, researchers and practitioners must decide on the practical importance of the small number of individuals who receive ratings with cross-source disagreement. Although the number of individuals is small, the proportion is similar to those who receive consistently high performance ratings and consistently low performance ratings. Organizations and practitioners would not ignore the small subset of high performers and the low performers, and researchers should not ignore the potential practical implications of the small number of individuals who receive divergent ratings.

Limitations and Future Research

There are several limitations with the current study that should be discussed. In conjunction with the limitations, suggests for future research are also outlined.

Latent profile analyses and fit statistics. The person-centered pattern approaches, in general, and latent profile analysis, in particular, are relatively new approaches in the management literature. As a result, there is little guidance on comparing models with different class sizes and determining the number of classes to retain. There are several statistics that can be used to determine the best class solution, but the fit statistics do not always converge on the same solution. Henson, Reise, & Kim (2007) note that none of the relative model fit statistics, used to determine the best fitting class solution, were exceedingly accurate. Further, they found that Adjusted BIC was the most accurate at determining the correct class solution, but was only correct 72% of the time.

In addition, there is a balance between relying on fit statistics and identifying theoretically meaningful classes. In the current study, fit statistics such as adjusted BIC and the entropy statistic suggested solutions with large class sizes; however as the classes sizes increased the sample size within each class become increasingly small. As a result, theoretical rationale was very important in determining the number of classes to retain. Relying on theoretical rationale lead to a fair amount of subjectivity associated with determining the best class solution.

Researchers using latent class analyses suggest that using a hold-out sample to confirm the class solution is the most defensible practice (M. Wang, personal communication, April 9, 2010; R. J. Vandenberg, personal communication, April 14, 2010). The current study uses a hold-out sample in addition to examining several fit statistics; consequently, the results presented in this paper are as precise as possible given the current best-practices provided in the literature. The use of hold-out samples is dependent on large sample sizes and there is still very little research with guidance on using hold-out samples in conjunction with fit statistics. As person-centered approaches become more popular in the management literature, it is critical that best-practices for determining class solutions continue to advance. More research needs to be done to determine better methods for determining the best fitting class solution.

Sample. The small sample size used to examine the relationship between personality and cross-source performance profiles is likely the reason that there were no observed differences between profiles with cross-source disagreement. Overall there were over 3700 raters that were used to determine the cross-source performance profiles, however, only 857 of these individuals provided personality information. Additionally, individuals who were classified in high performance group had a higher percent of individuals with personality data (43% - 24%) compared to the other groups (12% - 28%). Cross-source performance groups with low sample

size using the full sample (e.g. N = 41, 74, 75) were extremely low when individuals without personality data were removed (e.g. N = 5, 12, 11). As a result, there was likely not enough power to detect a significant effect between these groups. The current results should be interpreted cautiously given the lack of power and biased sampling. Further, the results should be replicated using a larger, more representative sample.

The MSPR sample used in this study included individuals who were participating in a leadership program. Organization believed these individuals had enough potential to financially invest in further development. There are likely managers employed by organizations in the current sampled that did not have the opportunity to participate in the leadership program. Consequently, the composition of the current sample may differ from a sample from a large organization where all managers participate in the MSPRs system. The difference may be the reason for less instances of cross-source disagreement. First, the groups who received ratings with cross-source disagreement were small and overall these groups had below average ratings. Practically it is unlikely that an organization would send all managers to a leadership development program and the managers who did not participated are likely those managers that have no potential to stay on the management track and have overall low performance. These results should be replicated in a sample with a difference composition of managers. Different MSPR instrument should be examined as well.

Nomological network. A key limitation of this study is that the potential nomologial network of predictor variables was not examined. Future research should investigate additional ratee characteristics as predictors of pattern of cross source ratings. Specifically, research should examine ratee characteristics such as gender, race, experience, skill/ability based characteristics (e.g., cognitive ability, administrative skills) and other personality characteristics (e.g., self-

monitoring, social sensitivity). In addition, organizational influences should be examined as a possible predictor of disagreement. For instance, organizational structures that create more interaction among sources may have higher cross-source agreement than organizational structures that limit interactions among supervisors, peers, and subordinates. Organizational culture may also influence cross-source ratings. If there is a strong organizational culture with emphasis on the goals and mission of the organization, there may be stronger agreement among sources than if there is little emphasis of the overarching goals of the organization. In this type of culture there may be more divergent views of what an individual's job performance should be, leading to more divergent performance ratings.

Derailment potential. Finally, a cross-sectional sample was used which included potential to derail ratings. The potential to derail ratings were one supervisor's perceptions of how likely that individual would be to fall off the management track. Actual rates of involuntarily career plateau, demotion and firings were not investigated. A retrospective approach has been a common practice for studying derailment (e.g. Gentry, et al., 2007; Lombardo, et al., 1988) but lacks the ability to make casual inferences with regard to the relationship between performance and derailment. In the current study, derailment potential had an inverse relationship with overall performance effectiveness. It is unclear if potential to derail is simply an inverse of overall performance ratings and whether that inverse linear relationship would hold for actual derailment. A longitudinal approach should be used to examine actual derailment and determine if individuals who receive ratings with high cross-source disagreement are more likely to actually derail. Given the wide range of cross-source performance profiles, investigating actual derailment may distinguish the lower level cross-source performance profiles that were not differentiated on potential to derail. This would give insight into which cross-source performance

profiles are necessary and sufficient to continue to be promoted and stay on the management track and which cross-source performance profiles serve as a red-flag for derailment.

General Conclusions

Overall, the current study adds unique information to the MSPR literature. First, the current results call in to question the extent to which raters really disagree. For the majority of individuals, there are slight differences in cross-source ratings but the practical implication of these differences with regard to performance level and rank order is nil. Although not as expected, the lack of disagreement is a good news scenario for organizations and practitioners. The consistent ratings across raters allows for straightforward reporting and interpretation of MSPR feedback. However, the small number of individuals who receive ratings with strong divergent cross-source ratings cannot be ignored. More research needs to be done to determine what types of individual receive ratings with cross-source disagreement and what effect they ultimately may have on the organization. Future research should continue to utilize a person-centered approach to investigate patterns of performance ratings.

REFERENCES

Anderson, C. &, Kilduff, G. J. (2009). Why do dominant personalities attain influence in face-to-face groups? The competence-signaling effects of trait dominance. *Journal of Personality and Social Psychology, 96*, 491-503.

Ashford, S. J. (1989). Self assessments in organizations: A literature review and integrative model. In L. L. Cummings & B. M. Staw (Eds.). *Research in organizational behavior* (Vol. 11, pp. 133-174). Creenwhich, CT: JAI Press.

Atwater, L. E. & Brett, J. F. (2005). Antecedents and consequences of reactions to developmental 360° feedback. *Journal of Vocational Behavior, 66,* 532-548.

Barrick, M. R. Parks, L. & Mount, M. K. (2005). Self-Monitoring as a moderator of the relationship between personality traits and performance. *Personnel Psychology, 58,* 747-767.

Bass, B. M. & Yammario, F. J. (1991). Congruence of self and other's leadership ratings of naval officers for understanding successful performance. *Applied Psychology: An International Review, 40*, 437-454.

Bergman, L. R., Magnusson, D., & El Khouri, B. M. (2003). *Studying individual development in an interindividual context: A person-oriented approach.* Mahwah, NJ: Erlbaum.

Blair, C. A., Hoffman, B. J., & Helland, K. R. (2008). Narcissism in organizations: A multisource appraisal reflects different perspectives. *Human Performance, 21,* 254-276.

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

Borman, W. C. (1991). Job behaviors, performance and effectiveness. In M. D. Dunnette & L. M. Hough (Eds). *Handbook of industrial and organizational psychology* (2nd ed, vol. 2, pp 833-872). Palo Alto, CA: Consulting Psychologists Press.

Borman, W. C. (1997). 360° Ratings: An analysis of assumptions and a research agenda for evaluation 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.

Bray, D. W., Campbell, R. S. & Grant, D. L. (1974). *Formative years in business: A long term AT&T study of managerial lives.* New York: Wiley.

Brutus, S., Fleenor, J. W., & McCauley, C. D. (1999). Demographic and personality predictors of congruence in multi-source ratings. *Journal of Management Development, 18*, 417-435.

Bynum, B. H., Hoffman, B. J. Meade, A. W. and Gentry, W. A. (under review). Reconsidering the Equivalence of Multisource Performance Ratings: Evidence for the Importance and Meaning of Rater Method Factors.

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

Carty, H. M. (2003). Review of BENCHMARKS® [revised]. In B. S. Plake, J. Impara, & R. A. Spies (Eds.), *The fifteenth mental measurements yearbook* (pp. 123-124). Lincoln, NE: Buros Institute of Mental Measurements.

Cattell, R. B. (1966). Multivariate behavioral research and the integrative challenge. *Multivariate Behavioral Research, 1,* 4-23.

Center for Creative Leadership. (2004). *BENCHMARKS® facilitator's manual*. Greensboro, NC: Center for Creative Leadership.

Celeux, G., & Soromenho, G. (1996). An entropy criterion for assessing the number of clusters in a mixture model. *Journal of Classification, 13,* 195-212.

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

Church, A. H. (1997). Managerial self-awareness in high performing individuals in organizations. *Journal of Applied Psychology, 82*, 281-292.

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

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

Costa, P. T., Jr., & McCrae, R. R.(1992). *The revised NEO Personality Inventory (NEO-PI-R) and NEO Five Factor Inventory (NEO-FFI) professional manual.* Odessa, FL: Psychological Assessment Resources.

CPP, Inc. (2002, February). Brief technical report on the CPI 260® instrument (Tech. Rep. No. 2002-3). Mountain View, CA: Author.

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,* 399-425.

Edwards, J. R., & Cable, D. M. (2009). The value of Value congruence. *Journal of Applied Psychology, 94,* 654-677.

Edwards, J. R., & Parry, M. E. (1993). On the use of polynomial regression equations as an alternative to difference scores in organizational research. *Academy of Management Journal, 36*, 1577 – 1613.

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

Ferris, G. L., Witt, L. A. & Hochwater , W. A. (2001). Interaction of social skill and general mental ability on job performance and salary. *Journal of Applied Psychology, 86,* 1075-1082.

Fletcher, C. (1997). Self-awareness – A neglected attribute in selection and assessment. *International Journal Selection and Assessment*, 5, 183-187.

Foti, R. J. & Hauenstein, N. M. (2007). Pattern and variable approaches in leadership emergence and effectiveness. *Journal of Applied Psychology, 92*, 347-355.

Gabarro, J. J. (1987). *The dynamics of taking charge.* Boston: Harvard Business School Press.

Gentry, W. A., Mondore, S. P., & Cox, B. D. (2007). An exploratory study of managerial derailment characteristics and personality preferences. *Journal of Management Development*, *26*, 857–873.

Gillespie, T. L. (2005). Internationalizing 360-degree feedback: Are subordinate ratings comparable? *Journal of Business and Psychology, 19*, 361-382.

Goffin, R. D. & Anderson, D. W. (2006). The self-rater's personality and self-other disagreement in multisource performance ratings: Is disagreement healthy*? Journal of Managerial Psychology, 22*, 271-299.

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

Greguras, G. J. (2005). Managerial experience and the measurement equivalence of performance ratings. *Journal of Business and Psychology, 19*, 383-397.

Hair, J. R., Anderson, R. E., Tatham, R. L., & Black, W. C. (1998). Multivariate data analysis. Upper Saddle River, NJ: Prentice-Hall.

Hannum, K. M. (2007). Measurement equivalence of 360°-Assessment data: Are different raters rating the same construct? *International Journal of Selection and Assessment, 15,* 293-301.

Harris, M. & Schaubroeck, J. (1988) A meta-analysis of self-supervisor, self-peer, and peer supervisor ratings, *Personnel Psychology, 41,* 43-61.

Hautaluoma, J. E., Dickinson, T. L. & Inada, A. R. (2001). Trait patterns, background characteristics, managerial styles, and job practices of industrial middle managers. The *Journal of Social Psychology, 132*, 201-215.

Henson, J. M., Reise, S. P, & Kim, K. H. (2007). Detecting mixtures from structural model differences using latent variable mixture modeling: A comparison of relative model fit statistics. *Structural Equation Modeling, 14,* 202-226.

Hoffman, B. J., Lance, C. E., Bynum, B. H., & Gentry, W. A. (2010). Rater source effects are alive and well after all. *Personnel Psychology.*

Hoffman, B. J., & Woehr, D. J. (2009). Disentangling the meaning of multisource feedback: Expanding the nomological network surrounding source and dimension factors. *Personnel Psychology*.

Hogan, R., & Hogan, J. (2001). Assessing leadership: A view from the dark side. *International Journal of Selection and Assessment, 9,* 40–51.

Hom, P. W., Griffeth, R. W., Palich, L. E., & Bracker, J. S. (1999). Assessing applicant 'fit' within the selection process: Which interviewer perceptions matter? *Personnel Psychology, 52,* 97-112.

Ilgen, D. R., Fisher, C. D., & Taylor, M. S. (1979). Consequences of individual feedback on behavior in organizations. *Journal of Applied Psychology, 64,* 359-371.

Jaccard, J., Turrisi, R., & Wan, C. K. (1990). Interaction effects in multiple regression. Newbury Park, CA: Sage.

James, L. R., Demaree, R. G., & Wolf, G. (1993). Rwg: An assessment of within-group agreement. *Journal of Applied Psychology, 78,* 306-309.

Johnson, J. W. & Ferstl, K. L. (1999). The effects of interrater and self-other agreement on performance improvement following upward feedback. *Personnel Psychology, 52,* 271-303.

Judge, T. A., Bono, J. E., Ilies, R., & Gerhardt, M. W. (2002). Personality and leadership: A qualitative and quantitative review. *Journal of Applied Psychology, 87*, 765-780.

Katzell, R. A. (1994). Contemporary meta-trends in industrial and organizational psychology. In H. C. Triandias, M. D. Dunnette, & L. M. Hough (Eds), *Handbook of industrial and organizational psychology* (3rd ed., vol. 4, pp 1-89). Palo Alto, CA: Consulting Psychologist Press.

Korac-Kakabadse, A., Korac-Kakabadse, N. and Myers, A. (1998), Demographics and leadership philosophy: exploring gender differences. *Journal of Management Development, 17,* 351-388.

Kovach, B. E. (1988). The Derailment of Fast-Track Managers. Organizational Dynamics, 15, 41-48.

Lance, C. E., Baxter, D., & Mahan, R. P. (2006). Multi-source performance measurement: A reconceptualization. In W. Bennett, C. E. Lance, & D. J. Woehr (Eds.), *Performance measurement: Current perspectives and future challenges* (pp. 49-76). Mahwah, NJ: Erlbaum.

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

Lance, C. E., Hoffman, B. J., Gentry, W. A., Baranik, L. E. (2008). Rater source factors represent important subcomponents of the criterion construct space, not rater bias. *Human Resource Management Review, 18,* 223-232.

Lance, C. E., Butts, M. M. & Micheals, L. C. (2006). The sources of four commonly reported cutoff criteria: What did they really say? *Organizational Research methods, 9*, 202-220.

Landy, F. J. & Farr, J. L. (1980). Performance Ratings. *Psychological Bulletin, 87,* 72-107.

Landy, F. J. & Farr, J. L. (1983). The measurement of work performance: methods, theory, and applications. New York : Academic Press.

Lebreton, J. M. & Senter, J. L. (2008). Answer to 20 questions about interrater reliability and interrater agreement. *Organizational Research Methods, 11,* 815-852.

Lo, Y., Mendell, N. R., & Rubin, D. B. (2001). Testing the number of components in a normal mixture. Biometricka, 88, 767-778.

Lombardo, M. M., & McCauley, C. D. (1988). *The dynamics of management derailment* (Tech. Rep. No. 34). Greensboro, NC: Center for Creative Leadership.

Lombardo, M. M., & McCauley, C, D. (1994). *BENCHMARKS®: A manual and trainer's guide*. Greensboro, NC: Center for Creative Leadership

Lombardo, M. M., McCauley, C. D., McDonald-Mann, D., & Leslie, J. B. (1999). *BENCHMARKS® developmental reference points*. Greensboro, NC: Center for Creative Leadership.

Lombardo, M. M., Ruderman, M. N., McCauley, C. D. (1988). Explanations of success and derailment in upper-level management positions. *Journal of Business and Psychology, 2*, 199-216.

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

Lubke, G. H. & Muthén, B. (2005). Investigating population heterogeneity with factor mixture models. *Psychological Methods, 10,* 21-39.

Mabe, P. & West, J. (1982). Validity of self-evaluation of ability: a review and meta-analysis. *Journal of Applied Psychology, 67,* 280-296.

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

McCall, M. & Lombardo, M. (1983). *Off the Track: Why and How successful Executives get Derailed* (Technical Report No. 21). Greensboro, NC: Center for Creative Leadership:

McCauley, C., & Lombardo, M. (1990). BENCHMARKS®: An instrument for diagnosing managerial strengths and weaknesses. In K. E. Clark & M. B. Clark (Eds.), *Measures of leadership* (pp. 535-545). West Orange, NJ: Leadership Library of America.

McCauley, C., Lombardo, M., & Usher, C. (1989). Diagnosing management development needs: An instrument based on how managers develop. *Journal of Management, 15*, 389-403.

McNally, B. and Parry, K. (2002), "A phenomenological examination of chief executive derailment/failure", paper presented at ANZAM-IFSAM 6th World Congress. Gold Coast, Australia, July 10-13.

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.

Mumford, M. D., Zaccaro, S. J., Johnson, J. F., Diana, M. Gilbert, J. A., & Threlfall, V. K. (2000). Patterns of leader characteristics: Implications for performance and development. *Leadership Quarterly, 11,* 115-133.

Mumford, T.V., Campion, M.A., & Morgeson, F.P. (2007). Leadership skills strataplex: Leadership skill requirements across organizational levels. *Leadership Quarterly, 18, 154-166*.

Muthén, L.K. and Muthén, B.O. (1998-2007). Mplus User's Guide. Fifth Edition. Los Angeles, CA: Muthén & Muthén

Nylund, K. L., Asparouhov, T., & Muthén, B. O. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A monte carlo simulation study. *Structural equation Modeling, 14,* 535-569.

Oh, I. S. & Berry, C. M. (2009). The five-factor model of personality and managerial performance: Validity gains through the use of 360 degree performance ratings. *Journal of Applied Psychology, 94*, 1498-1513.

Oshagbemi, T. (1988*). Leadership and Management in Universities*. Berlin: Walter de Gruyter.

Oshagbemi, T. (1997). Job satisfaction profiles of university teachers. *Journal of Managerial Psychology, 12,* 27-39.

Oshagbemi, T. & Ocholi, S. A. (2005). Leadership styles and behaviors profiles of managers. *Journal of Management Development, 25,* 748-762.

Ostroff, C. Atwater, L. E., Feinberg, B. J. (2004). Understanding self-other agreement: A look at rater and ratee characteristics, context and outcomes. *Personnel Psychology, 57,* 333-375.

Owens, W. A. & Schoenfeldt, L. F. (1979). Toward a classification of persons. *Journal of Applied Psychology, 64*, 569-607.

Pastor, D. A., Barron, K. E. Miller, B. J. & Davis, S. L. (2007). A latent profile analysis of college students' achievement goal orientation. *Contemporary Educational Psychology, 32*, 8-47.

Roush, P. & Atwater, L. (1992). Using the MBTI to understand transformational leadership and self perception accuracy. Military Psychology, 4, 17-34.

Schneider, B. (1987). The people make the place. *Personnel Psychology, 40*, 437-455.

Schneider, B. Goldstein, H. W. & Smith, D.B. (1995). The ASA framework: An update. *Personnel Psychology, 48,* 747-773.

Schneider, B. & Schneider, J. L. (1994). Biodata: An organizational focus. In G. S. Stokes, M. D. Mumford, & Owens, W. A. (Eds.), Biodata handbook: Theory, research and use of

biographical information in selection and performance prediction (pp. 423-450). Palo Alto, CA: Consulting Psychologist Press.

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

Skinner. W., & Sasser, E. (1978). *Manufacturing in the corporate strategy.* New York: Wiley & Sons.

Smith, J. A., & Foti, R. J. (1998). A pattern approach to the study of leadership emergence. *Leadership Quarterly, 9,* 147–160.

Smith, G., & Smith, J. (2005). Regression to the mean in average test scores. *Educational Assessment, 10,* 377-399.

Spangler, M. (2003). Review of Benchmarks [revised]. In B. S. Plake, J. Impara, & R. A. Spies (Eds.), *The fifteenth mental measurements yearbook* (pp. 124-126). Lincoln, NE: Buros Institute of Mental Measurements.

Stewart, R. (1984). The nature of management? A problem for management education*. The Journal of Management Studies, 21,* 323-330.

Stewart, R. (1988). *Managers and their jobs: A study of the similarities and differences in the way mangers spend their time* (2nd ed.). London: Macmillan.

Stone-Romero, E. F., & Anderson, L. E. (1994). Relative power of moderated multiple regression and the comparison of subgroup correlation coefficients for detecting moderating effects. Journal of Applied Psychology, 79, 354-359.

Stoner, J. A. F., Ference, T. P., Warren, E. K. & Christensen, H. K. (1980). *Managerial career plateaus: An exploratory study.* New York: Center for Research in Career Development, Columbia University.

Thissen, D. Steinberg, L. & Kuang, D. (2002). Quick and easy implementation of Benjamini Hochberg Procedure for controlling false positive rate in multiple comparisons. *Journal of Educational & Behavioral Statistics, 27,* 77-83.

Van Velsor, E., & Leslie, J. B. 1995. Why executives derail: Perspectives across time and culture. Academy of Management Executive, 9: 62-72.

Van Velsor, E., Taylor, S., & Leslie, J. B. (1993). An examination of the relationship between self-perception accuracy, self-awareness, gender and leader effectiveness. *Human Resource Management, 32,* 249-264.

Vecchio, R. P. & Anderson, R. J. (2009). Agreement in self-other ratings of leader effectiveness: The role of demographics and personality. *International Journal of Selection and Assessment, 17,* 165-179.

Vinson, M. (1996). The pros and cons of 360-degree feedback: Making it work. *Training and Development, 50*, 11-12.

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*, 245-354.

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

Vuong, Q. (1989). Likelihood ratio tests for model selection and non-nested hypotheses. Econometrica, 57, 307–333.

Weiner, B. (1974). *Achievement motivation and attribution theory.* Morristown, N.J.: General Learning Press.

Weiner, B. (1986). An *attributional theory of motivation and emotion*. New York: Springer-Verlag.

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

Yammarino, F. & Atwater, L. (1993). Understanding self-perception accuracy: implications for human resource management. *Human Resource management, 32*, 321-247.

Yang, L. Q., Levine, **E.** L., Smith, M. A., Ispas, D., Rossi, M. E. (2008). Person-environment fit or person plus environment: A meta-analysis of studies using polynomial regression analysis. *Human Resource Management Review, 18,* 311-321.

Yukl, G.A. (2006). *Leadership in organizations* (6th ed.). Upper Saddle River, NJ: Pearson Prentice Hall.

Zaccaro, S. J. (2007). Trait-Based perspectives of Leadership. *American Psychologist, 62*, 6-16.

Zedeck, S. (1995). Review of BENCHMARKS®. In J. Conoley & J. Impara (Eds.), *The twelfth mental measurements yearbook* (Vol. 1, pp. 128-129). Lincoln, NE: Buros Institute of Mental Measurements.