College of Liberal Arts & Social Sciences Theses and Dissertations

College of Liberal Arts and Social Sciences

6-2010

# Self-other agreement and leader effectiveness: An examination of differences across rater sources and leader behaviors

Devon Nicole Riester
*DePaul University*, driester@depaul.edu

SELF-OTHER AGREEMENT AND LEADER EFFECTIVENESS:

AN EXAMINATION OF DIFFERENCES ACROSS RATER SOURCES

AND LEADER BEHAVIORS

A Dissertation

Presented in

Partial Fulfillment of the

Requirements for the Degree of

Doctor of Philosophy

BY

DEVON NICOLE RIESTER

MARCH, 2010

Department of Psychology

College of Liberal Arts and Sciences

DePaul University

Chicago, Illinois

DOCTORAL COMMITTEE

Suzanne Bell, Ph.D.

Chairperson


Douglas F. Cellar, Ph.D.

Robert S. Rubin, Ph.D.

Annette J. Towler, Ph.D.

D. Joel Whalen, Ph.D.

<u>ACKNOWLEDGEMENTS</u>

I would like to acknowledge several individuals who played a significant role in the completion of this work. First, I would like to thank my dissertation chair, Suzanne Bell, for her insightful feedback and advice throughout this process, as well as her support and mentorship throughout graduate school. I would also like to sincerely thank the other members of my committee for their valuable contributions.

This study would not have been possible without the help of Rob Kabacoff at Management Research Group, who provided the data for this dissertation and graciously answered my questions along the way.

To my colleagues at Vantage Leadership Consulting, who gave me the encouragement I needed to reach this milestone: I am tremendously grateful to work with such a supportive, caring, and truly fun group of individuals.

Lastly, my friends and family deserve my deepest gratitude. To my brother, who has an uncanny ability to make me laugh, and whose honesty and sincerity keeps me focused on what really matters. To my dad, who continues to set the ultimate example of hard work, persistence, and integrity toward which I will always strive. And to my mom, for her overwhelming patience, selflessness and warmth. Thank you.

VITA

Devon N. Riester was born in Appleton, Wisconsin on July 24, 1980. She graduated from Appleton West High School in 1998 and received a Bachelor of Arts degree in Psychology from Kenyon College in 2002. She received her Master of Arts degree in Industrial/Organizational Psychology with distinction from DePaul University in 2007. She currently works as an Associate Consultant at Vantage Leadership Consulting in Chicago, IL.

TABLE OF CONTENTS

## LIST OF TABLES

## LIST OF FIGURES

CHAPTER I

INTRODUCTION

Multi-source feedback (MSF) is regarded as "one of the most popular

industrial-organizational psychology, organization development, and human

resource development interventions of the decade" (Church, 2000, p. 99).

MSF refers to the process of gathering anonymous feedback about a person's

behavior from multiple sources on a number of performance dimensions

(Lepsinger & Lucia, 1997). A typical MSF process involves collecting

feedback, most often in the form of a questionnaire, from the individual's

supervisor, direct reports, peers, customers, and/or suppliers. The term 360-

degree feedback is also used to describe this process because feedback is

solicited from sources "all around" an employee with the goal of providing a

comprehensive viewpoint of an employee's capabilities, behaviors, and

leadership style (Nowack, 1993). A qualified facilitator (e.g., HR specialist,

executive coach) typically reviews MSF results with the target individual in

order to interpret the feedback (Antonioni, 1996).

MSF can be used for a variety of purposes, including employee

development, appraisal, selection, and/or facilitation of organizational change;

however, the majority of MSF systems are used as leadership or managerial

development tools (Morgeson, Mumford, & Campion, 2005). One of the

underlying assumptions of MSF is that by focusing on the discrepancies

between self and other (i.e., supervisor, peer, direct report) ratings, the MSF

recipient is able to increase his/her self-awareness and as a result, change

his/her behavior or develop skills or capabilities to become more effective (Morgeson et al., 2005).  Despite the fact that MSF provides richer and broader data from a variety of sources, one of the major drawbacks is that it can also provide an overwhelming amount of information, making it difficult for the MSF recipient to process (DiNisi & Griffin, 2001).  To the MSF recipient, it can be a challenge to reconcile the differences between self-other ratings and understand which self-other discrepancies should be the focus of their ongoing development.  The current study contributes to the growing body of knowledge on MSF systems by examining the *source* and *behavior* for which self-other agreement (conceptualized as self-awareness) is more related to leader effectiveness.  In other words, this research explores the type of alignment (between self and other ratings) that is most critical to perceived effectiveness.  Results will help guide MSF discussions so that feedback recipients feel less overloaded and more able to set behavioral objectives based on the specific discrepancies that are most highly related to their perceived effectiveness.

<div align="center">Brief History of MSF</div>

Although MSF has become increasingly popular within organizations during the last 20 to 30 years, the concept of evaluating performance began much earlier.  Psychologists have a history of helping organizations with the development and implementation of effective performance appraisal systems. During the early 1900s, psychologists began assisting the military with the design of officer performance instruments (Wiese & Buckley, 1998).  A

variety of tools and procedures for evaluating performance have been developed over time, including global essays, judgmental rank order, graphic and trait ratings, and critical incident surveys (Landy & Farr, 1980). During the 1960s and 1970s, psychologists focused on developing a variety of rating formats including behaviorally anchored rating scales (BARS) and behavioral observation scales (Austin & Villanova, 1992). At that time, researchers were interested in developing training for raters and exploring the impact of individual differences in raters. The content of performance appraisal has also evolved over time, shifting from a focus on one global measure of performance to several broad traits, and then to a number of specific behaviors or goals (McGregor, 1957).

The concept of MSF was first explored by researchers in the 1950s and 1960s when they began experimenting with the concept of using other raters in addition to supervisors. One of the first researchers to explore different rater sources was Lawler (1967) who discovered that each rater group (i.e., supervisor, peers, subordinates, and self) provided a useful viewpoint of the employee's performance. Thornton (1968) also identified meaningful differences between self and supervisor ratings in predicting promotability, indicating that each source offered a unique perspective on the employee's ability to advance to higher levels within the organization. Overall, these early studies established that discrepancies in ratings from different sources are not considered error, but instead, provide meaningful and unique feedback which can help leaders better understand their evaluations of performance.

Although MSF received some attention from researchers during the 1950s and 1960s, research was relatively scarce until MSF became increasingly adopted within organizations.  Some of the earliest origins of MSF within organizations were in the form of employee opinion surveys which were administered to employees across all organizational levels (Fleenor & Prince, 1997).  The underlying premise was that input gathered from multiple perspectives was believed to be more comprehensive and objective than information obtained from only one source (e.g., the senior leadership team).  Employee opinion surveys were used to gather information about specific aspects of the organization such as satisfaction with salary, attitudes toward leadership, and feelings toward co-workers.

The popularity of MSF increased considerably when progressive organizations began developing MSF surveys in the late 1970s and 1980s.  At the Center for Creative Leadership, Robert Bailey and Robert Dorn began conducting research on multiple rating sources and proposed the idea of assessment as a means for developing leaders (Bracken, Timmreck, & Church, 2001).  After the first MSF surveys were developed in the late 1970's, other progressive organizations began to follow, and offerings of MSF surveys grew rapidly throughout the 1980s and 1990s.  In 1991, Van Velsor and Leslie provided a list of 16 available MSF instruments, and the estimate grew to 24 surveys in 1998 (Leslie & Fleenor, 1998).  Prominent organizations began using MSF as a development tool, and in 1998, it was estimated that at least 90% of Fortune 1000 organizations used some form of

360-degree feedback, including companies such as Proctor & Gamble, Motorola, Federal Express, and United Airlines (Atwater & Waldman, 1998; Waldman & Atwater, 1998).

The increased focus and usage of MSF in applied settings was likely due to a number of factors. Because organizations have become increasingly more dynamic and fast-paced, there was a need for continuous measurement and improvement of capabilities (Nowack, 1993). In addition, organizational structures flattened over time as traditional hierarchical organizations became less prominent and matrix organizations became more popular. Flatter organizations created the need to gather feedback from sources other than a traditional supervisor or manager. Also, with the increase in team-based structures, individual's roles and responsibilities have become broader in scope, which requires employees to gather feedback from a wider range of employees across all levels of the organization. Lastly, the peak in organizational information technology during the 1990s likely contributed to the increased usage of MSF, as new software made it possible to summarize ratings from multiple rater sources (i.e., peers, supervisor, direct reports) on multiple performance dimensions in customized feedback reports (Nowack, 1993).

<u>Benefits and Purposes of MSF</u>

Employees and organizations began to realize the benefits of implementing MSF, which contributed to its continued popularity and use. Most notably, MSF provides recipients with feedback from sources that is

otherwise not readily available. Each rater source is believed to provide unique information regarding their perceptions and assessments of the leader (Morgeson et al., 2005). As evidence, Mount, Judge, Scullen, Sytsma, and Hezlett (1998) collected self-ratings as well as ratings from supervisors, peers, and subordinates and concluded that each rater source provided partially unique information. Each group is a valuable source of information, but each group *on their own* may not provide the full picture because they may observe the leader in different settings and thus may observe different behaviors (Borman, 1974; Morgeson et al., 2005). Supervisors are believed to provide a valuable perspective because they are uniquely familiar with the job and what is required for success; however, supervisors may not have as many opportunities to observe the leader across performance settings. Peers, on the other hand, are believed to observe a higher proportion of the leaders' behaviors because they typically interact with the leader on a more regular basis (e.g., project teams, executing day-to-day responsibilities). Direct reports also provide a valuable viewpoint because they observe the individual in a leadership role (e.g., delegating tasks, delivering feedback, communicating expectations). In summary, each rater source is likely to observe the leaders' behaviors in different contexts, and for this reason, provides a unique and valuable perspective (Morgeson et al., 2005).

One of the primary benefits of MSF is that it serves a developmental purpose. Feedback from a range of sources on a number of performance dimensions is used to direct attention to an individual's strengths and

weaknesses (Tornow, 1993b). Awareness of discrepancies between how we rate ourselves and how others rate our behaviors is believed to enhance self-awareness (Church, 1997, 2000). Many users of MSF believe that identifying differences between self and observer perceptions is an important step in maximizing individual performance, which then becomes a foundation for management and leadership development (Tornow, 1993b).

Feedback is especially important for individuals in managerial or leadership roles. Through in-depth interviews with eighty-four executives, Longenecker and Gioia (1992) found that as a leader advances to higher job levels within an organization, he/she is less likely to receive quality feedback about his/her job performance. The lack of feedback for higher level leaders is likely to limit their ability to perform their job effectively, develop professionally, and improve their management and leadership skills. Thus, MSF is one way to provide leaders with detailed feedback from a variety of sources across performance domains. MSF provides leaders with crucial information in terms of identifying their strengths, weaknesses, and potential "blind spots" (i.e., performance areas where the leader believes he/she is effective, but others see weaknesses).

An additional benefit of MSF is that these systems can be used to reinforce organizational values (Fleenor & Prince, 1997). In order for an organizational value to become part of the organizational culture, it must be fully developed and reinforced among employees (Parker-Gore, 1996). MSF can be used to emphasize the value of certain behaviors or leadership

capabilities within an organization. For example, if a technology organization values innovation, MSF can be used to deliver feedback to employees on their ability to creatively problem solve and develop new ideas. When MSF is aligned with the organization's values, individual feedback is likely to be considered more useful and valid. In fact, Gebelein (1996) proposed that MSF is most beneficial when it supports not only individual development, but also encourages individuals to change in ways that are consistent with the organizational strategy.

<u>Uses of MSF: Developmental vs. Appraisal</u>

One of the most highly debated topics within the MSF literature focuses on the uses of MSF within organizational systems. The debate is centralized around the question of whether or not MSF should be used for development purposes or as part of the performance appraisal process (Garavan et al., 1997). Many researchers warn against using MSF for anything other than leadership or managerial development (Morgeson et al., 2005). The primary argument is that using MSF for performance appraisal affects how the raters evaluate the target individual. For example, Waldman and Atwater (1998) explain that when MSF is used for evaluative purposes, employees tend to inflate their ratings, which may not contribute to the uncovering of leaders' improvement areas. In fact, research has shown that when MSF is used for evaluative purposes (i.e., versus purely developmental), up to 40% of raters change their ratings in order to influence outcomes, thus making the ratings less reliable and ultimately less helpful for developing

leaders (Waldman & Atwater, 1998). Other researchers have also found that when raters are told that their ratings will be used for decision-making (versus developmental) purposes, both self and peer ratings are inflated (Antonioni, 1996).

On the other hand, a possible advantage of using MSF for appraisal purposes is that multiple sources of feedback may provide a more well-rounded perspective of an employee's performance (McGarvey & Smith, 1993), and as such, many organizations are now attempting to incorporate MSF into their appraisal and development systems (London & Beatty, 1993). In these situations, researchers warn against implementing MSF for both purposes at the onset, and instead, suggest using MSF for developmental purposes for several years before using it as an input to performance appraisals or decisions about pay and promotion (London & Beatty, 1993).

<u>Performance Improvement Following MSF</u>

To demonstrate the empirically-based benefits of MSF, researchers have examined the long-term outcomes of MSF, focusing mainly on performance improvement over time. In one of the earlier studies on the impact of upward feedback, Tuckman and Oliver (1968) showed the usefulness of gathering student evaluations as a way to improve teacher performance. In the organizational setting, Hegarty (1974) found that managers who received feedback from their subordinates improved their behavior and had increased subordinate ratings of managerial performance over time.

Several more recent studies have also shown that MSF leads to performance improvement in the future. Through an examination of 13 longitudinal studies, Smither, London, Reilly, Flautt, Vargas, and Kucine (2002) reported initial evidence of significant performance improvements following MSF. As a follow-up study, Smither, London, and Reilly (2005) analyzed the results of 24 longitudinal studies and also found modest, yet positive improvements in employee behaviors and attitudes following MSF interventions.

Evidence of performance improvement following MSF, however, has not been consistently positive. Performance feedback is a critical component to many organizational interventions, including MSF, and as such, Kluger and DeNisi (1996) sought to understand when feedback will have an effect on subsequent performance. In a large-scale meta-analysis, Kluger and DeNisi showed an overall moderate effect size ($d = .41$) of performance improvement following feedback, indicating that feedback generally leads to performance improvement. However, over one-third of the feedback interventions resulted in decreased performance over time. Based on their findings, Kluger and DeNisi proposed a feedback intervention theory (FIT) which is helpful in understanding the underlying mechanisms of MSF. Their theory proposes that behavior is regulated by comparing feedback to standards or goals, and that goals are hierarchically arranged. At the top of the hierarchy, goals are related to the self (i.e., self-concept), whereas goals at the bottom of the hierarchy are related to specific tasks. An individual's attention is typically

directed at a moderate level within this hierarchy; however, when an individual receives feedback, their attention shifts toward the level at which the feedback is focused.

Using FIT, Kluger and DeNisi (1996) found that the effectiveness of feedback increased as attention focused on task details and decreased as attention shifted away from the task and toward the self. Therefore, FIT suggests that in order to promote performance improvement following MSF, feedback should be focused on the task itself and not the individual receiving the feedback (i.e., to avoid cues related to self-esteem or other meta-level processes). In addition, feedback is more likely to result in performance improvement if it is combined with goal-setting activities; thus, the authors recommend including a formal goal-setting plan when delivering MSF in order to have the most positive impact on performance (DeNisi & Kluger, 2000).

<center>Review of MSF Research</center>

Underlying Mechanisms of MSF

Several theories have been used to describe the underlying mechanisms involved with MSF. One theory is that MSF allows individuals to use feedback on specific behaviors to set developmental goals (e.g., solicit more input from team members when making decisions). Goal setting theory proposes that goals serve a directive function in that they focus attention and effort on goal-relevant activities and away from irrelevant activities (Locke & Latham, 2002). MSF directs individuals toward goal-relevant activities based

on the dimensions rated in MSF system (e.g., teamwork). In addition, setting

goals that are specific and difficult leads to the greatest increase in

performance (Locke & Latham, 2002), and MSF allows individuals to set

more specific goals because they receive feedback from each rater source on

specific behavioral dimensions.

Control theory (Carver & Scheier, 1982) has also been used to explain

the underlying mechanisms of MSF. This theory suggests that individuals are

motivated to reduce discrepancies between their behaviors and a performance

standard or goal. In the context of MSF, these discrepancies typically exist

between self and observer ratings on specific leadership behaviors. Based on

these theories, it is believed that managers who observe the largest

discrepancy between their self- and other-ratings (e.g., supervisor, peers,

subordinates) will demonstrate the largest gains in performance (Smither,

London, Vasilipoulos, Reilly, Millsap, & Salvemini, 1995) because they are

motivated to reduce the difference between their own perceptions and those of

their observers (e.g., peers, supervisor, subordinates).

Context of MSF Systems

The organizational context and perception of the MSF process are

important factors which have received significant attention from researchers

in the field. The organizational context could either contribute to or interfere

with the success of the MSF intervention. For example, Atwater, Waldman,

Atwater, and Cartier (2000) found that employees who were cynical about the

MSF process (e.g., believed that change was not possible) were less likely to

improve performance after receiving feedback. The absence of integration within other existing HR systems may also limit the success of MSF. In a study of over 100 organizations, Brutus and Derayeh (2002) found that when MSF processes were not integrated with other HR initiatives (e.g., performance appraisal, training), employees resisted the MSF process. As previously discussed, it is also important for organizations to clarify the purpose of MSF (i.e., developmental vs. appraisal), and Atwater and Waldman (1998) suggests that in most cases MSF should be used for developmental purposes.

Perceptions of the MSF process, including acceptance and trust in the appraisal and feedback process, are also critical factors for ensuring successful implementation. Because employees are often rating their supervisors, peers, and direct reports, they may be concerned that the recipient could trace their responses back to them, which could result in retaliation. If raters do not believe they are anonymous, they are less likely to participate in the process, or if they choose to participate, they may inflate their ratings to avoid confrontation. Thus, rater anonymity among peer and subordinate raters has been shown to be related to more honest responses (Brutus & Derayeh, 2002). Antonioni (1994) also found that employees who perceive anonymity in the process are more likely to provide honest feedback compared to employees who believe their responses could be associated with them. Overall, research has shown that rater anonymity and trust in the integrity of the process are crucial components for creating accurate MSF ratings.

Process of Gathering MSF

The actual process of gathering multisource feedback has also received significant attention from both researchers and practitioners. In their survey of 20 organizations, London and Smither (1995) found several trends in the way MSF surveys are typically administered. Organizations commonly ask different rater groups to respond to the same set of MSF survey items, which typically focus on behaviors rather than traits. The number of raters from each source is generally between four and six raters per group. In addition to gathering feedback from supervisors, peers, and subordinates, 60% of the organizations in their study collected ratings from both internal and external customers, 20% of the organizations gathered feedback from internal customers only, and the remaining 20% did not collect customer ratings. In terms of the delivery of MSF results, ratings are usually presented separately from each rater source (i.e., rather than as a composite) assuming that each rater group consists of at least three raters. Additionally, it is common for self-ratings to be contrasted with others' ratings when MSF results are delivered to the target individual. According to Yukl (2006, pg. 398): "It is common practice to highlight large discrepancies between what others say about a manager's behavior and self-ratings by the manager." London and Smither (1995) found that 90% of MSF results specifically contrasted self and other ratings, and 70% provided an indicator of within-source agreement (e.g., range, standard deviation), implying that comparisons would be made across rater groups.

Recently, many organizations have started collecting MSF using electronic survey methods (Atwater et al., 2007). Web-based MSF surveys offer several advantages including increased speed, convenience, security, and confidentiality. Research has confirmed that electronic MSF surveys do not result in different ratings, as Smither, Walker, and Yap (2004) found no differences in feedback scores as a function of the data collection method (e.g., electronic versus paper). Despite the increased efficiency of web-based MSF surveys, the process can still be time-consuming because of the burden involved with completing a large number of surveys at one time (Atwater et al., 2007). For example, some supervisors may have 10 or more subordinates; thus, if the entire team is involved with the MSF process, the supervisor is required to dedicate a significant amount of time to completing separate surveys for each individual. One suggestion is for supervisors to rate only half of their team each year, which may work well if MSF processes have been in place for awhile (Brutus & Derayeh, 2002).

Characteristics of Feedback

It appears as though the MSF process is a different experience for individuals who receive more positive feedback compared to those who receive more negative feedback. Recipients of MSF who receive positive feedback view the ratings as more accurate and useful compared to negative feedback (Brett & Atwater, 2001). Individuals who receive negative feedback from supervisor and peers describe feelings of discouragement and anger immediately following the delivery of MSF feedback. However, these

negative feelings have been shown to diminish over time (i.e., several weeks after receiving MSF; Brett & Atwater, 2001). Smither et al. (2005) reported similar findings in that leaders who initially received negative feedback had immediate negative reactions; however, six months later, these individuals had developed more improvement goals for themselves compared to leaders who received more positive feedback. Researchers conclude that "negative feedback may take awhile to sink in or recipients may need some time to reflect and absorb the feedback (Smither et al., 2005, p. 203)." Thus, although recipients may initially have an adverse reaction to negative feedback, they are capable of setting improvement goals if the feedback intervention provides them with a coach or facilitator to help interpret the feedback and minimize any negative reactions (Atwater et al., 2007).

In addition to the positive or negative nature of the feedback, research has also examined the extent to which self and other ratings differ, and how this impacts outcomes of MSF. Research has shown that one's self-evaluation often differs from feedback received from others. For example, meta-analyses have demonstrated relatively low correlations among rater source (i.e., superior, self, and peer ratings; Mabe & West, 1982). Additionally, Harris and Schaubroeck (1988) found that self-ratings were not as highly related to other ratings (i.e., peers, superiors, or subordinates) compared to correlations among "other" sources (i.e., ratings from peers, subordinates, and superiors with one another). Rather than being viewed as error, these findings suggest that performance may be different, or may be perceived differently, across

various rater groups, and that MSF is needed to capture these variations in perspective (Day, 2001).

Self-rating inflation is the most common type of discrepancy found between self and other ratings (Podsakoff & Organ, 1986). Although there are many possible reasons for the inflation of self-ratings (e.g., self-presentation as a way to enhance one's public image; Baumeister, 1982) one commonly discussed reason is that self-raters may be unaware of how they are viewed by others. Also, self-raters may rate themselves highly in an attempt to produce stronger ratings from others (Harris & Schaubroeck, 1988). Although self-rating inflation is the most common form of discrepancy, not all raters inflate their own ratings. In fact, some raters actually deflate their ratings while others rate themselves similar to others (Atwater & Yammarino, 1992).

The Fundamental Attribution Error (FAE; Mitchell & Kalb, 1982) is a theory that can help explain differences in self-other ratings. FAE suggests that different groups may rate performance differently because they attribute behaviors to different factors (e.g., internal vs. external attributions). Although FAE was primarily offered as an explanation for the discrepancies between supervisors' ratings of their subordinates' performance, this theory can also be applied to behavioral ratings from other sources. The premise of FAE is that "observers" are more likely to attribute negative behaviors to internal attributes (e.g., skills, abilities) of the target individual compared to external or situational factors (e.g., inadequate support, unclear direction, poor supervision). For example, supervisors are more likely to blame their

subordinates for poor performance because blaming situational factors might suggest that the supervisor did not provide adequate direction or oversight. The same theory could be applied to peers and direct reports in that, as "observers," they are more likely to attribute negative behaviors to the individual (as opposed to the situation) which could result in lower ratings from observers compared to self-ratings.

Several studies have examined the impact of self-other rating agreement in terms of performance improvement, which is typically measured in terms of change in MSF ratings on the same measure *over time*. One of the earlier studies on the impact of self-other agreement examined student leaders and their followers at the U.S. Naval Academy (Atwater, Roush, & Fischthal, 1995). Self-other agreement was categorized into three groups: in-agreement (i.e., similar self and subordinate ratings), over-estimators (i.e., self-ratings higher than subordinate ratings), and under-estimators (i.e., self-ratings lower than subordinate ratings). Although feedback generally led to positive behavioral change, differences emerged based on the agreement between self and other ratings. Specifically, for individuals who were in-agreement, neither their behaviors nor their self-ratings changed over time. For under-raters, these individuals significantly raised their self-ratings following feedback, but did not change their behavior based on ratings from their subordinates (likely because the feedback informed them that they were performing better than they expected, thus there was no need to change their behavior). Lastly, for over-raters, these individuals reduced their self-ratings

after feedback and significantly improved their behavior as rated by subordinates. This study demonstrated that over-raters were most responsive to feedback in that they altered both their behaviors and self-ratings following feedback (Atwater et al., 1995).

Consistent with findings by Atwater et al. (1995), Johnson and Ferstl (1999) also examined change in self-other ratings in terms of MSF results over one year. Similarly, Johnson and Ferstl found that over-raters improved their performance most over time (based on subordinate ratings). They also found that individuals either increased or decreased their self-ratings based on their initial feedback (i.e., over-raters decreased self-ratings over time, under-raters increased self-ratings over time) in order to gain consistency in the way they rate themselves compared to their subordinates. The authors offered self-consistency theory (Korman, 1976) to explain that managers are motivated to reduce the discrepancy between how they perceive themselves and how others perceive them in order to minimize feelings of cognitive dissonance. Their findings could also be explained with control theory (Carver & Scheier, 1992), which states that individuals are motivated to reduce discrepancies, and in this case, over-raters did so by adjusting their self-ratings and changing their behavior over time in order to align self and other ratings.

<div align="center">Self-Awareness</div>

In addition to examining the *outcome* of self-other agreement in terms of performance improvement or behavioral change, researchers have also examined self-other agreement as an indicator or measure of managerial self-

awareness (MSA; Church, 1997).  Self-awareness, within an organizational

context, is defined as "the ability to reflect on and accurately assess one's own

behaviors and skills as they are manifested in workplace interactions"

(Church, 1997, pg. 281).  Measuring self-awareness directly is challenging

because individuals are not accurate at rating their own or others' self-

awareness (Fleenor, McCauley, & Brutus, 1996).  Thus, self-awareness has

frequently been measured as the degree to which a discrepancy exists between

self- and other-ratings, where smaller discrepancies indicate greater self-

awareness and larger discrepancies indicate less self-awareness.

Discrepancy-defined self-awareness has become a frequently

measured construct by organizational researchers (e.g., Church, 1997; Church,

2000; Fleenor et al., 1996; Tekleab, Sims, Yun, Tesluk, & Cox, 2008).  The

rationale for using discrepancy-defined self-awareness is that it provides an

estimate of the extent to which leaders and their colleagues agree in their

descriptions of the leaders' behaviors.  Alignment in self-other ratings

indicates that the leader has an accurate self-perception, while misalignment

indicates that the leader either overestimates or underestimates the extent to

which he/she exhibits key behaviors (which represents a lack of insight in

terms of how others perceive his/her behavior).  Although self-other

agreement is not a direct measure of self-awareness, initial evidence supports

the reliability and content validity of discrepancy defined self-awareness in

the MSF context (Kulas & Finkelstein, 2007).

<u>Self-Awareness and Effectiveness</u>

Self-awareness is considered an important capability which is required for managerial or leader effectiveness (Ashford & Tsui, 1991; McCall, Lombardo, & Morrison, 1988). Ashford (1989) explains that leaders need to become skilled at observing and evaluating their own leadership behavior and understanding how others perceive that behavior. Self-aware individuals have an accurate view of their behavioral tendencies and frequency of engaging in specific behaviors. In other words, someone who is self-aware is likely to be aligned with his/her observers in terms of the behaviors that he/she displays more or less frequently. An individual who is less self-aware might believe that he/she engages in behaviors more or less often than his/her observers perceive. A lack of self-awareness may indicate a leader who is not attuned with his/her strengths and weaknesses, not receptive to feedback, or someone who may ignore or not respond appropriately to past failures or mistakes (Atwater, Ostroff, Yammarino, & Fleenor, 1998).

One of the early studies to examine the relationship between self-awareness (defined as agreement between self-ratings and subordinate ratings on a number of managerial behaviors) and effectiveness was conducted by Church (1997). Effectiveness was measured by dichotomizing managers into two groups: high-performing and average-performing (i.e., based on a variety of performance measures). Consistent with their hypothesis, results indicated that high-performing managers had significantly higher congruence between self and direct report ratings compared to average-performing managers. Self-

other agreement was measured in several different ways, including difference

scores and between-manager correlations (i.e., correlation between the

average self-rating and the corresponding average others' rating).

Interestingly, the method used to measure agreement did not have an impact

on their results. Overall, the findings by Church (1997) imply that high-

performing managers are able to more accurately assess their own leadership

behaviors in the workplace. However, the authors point out that because their

data is descriptive and not causal, it is unknown if high-performing managers

became more self-aware (i.e., a skill that they developed) or if the presence of

self-awareness actually contributed to their designation as a high-performing

manager. However, regardless of the direction of the effect, the authors assert

that self-awareness is associated with a managers' performance.

Self-Other Agreement and Effectiveness

One of the most important advancements within self-other agreement

research was a shift in the way that self-other agreement was measured.

Previous research by Church (1997) primarily used single indices (e.g.,

difference scores, correlations between self and other ratings) to represent the

degree of self-other agreement. However, Atwater and Yammarino (1997)

argued that the preferred method was to consider both *degree* (i.e., high or low

ratings) and *type* of agreement (i.e., in-agreement or disagreement), resulting

in a four group categorization including in-agreement/good, in-

agreement/poor, overestimators, and underestimators. It is important to

distinguish between in-agreement/good (i.e., consistently above average

ratings) and in-agreement/poor (i.e., consistently below average ratings) because simply being aligned with other ratings is unlikely to be related to effectiveness if the individual and the observers rate the individual as consistently below average. On the other hand, if self and other ratings are *not* aligned, it is important to consider the direction of the discrepancy (i.e., whether the individual rates themselves higher or lower compared to observers) because these two groups are likely to have a different relationship with effectiveness. For example, overestimators may be overconfident in their abilities (and thus, may be unaware of significant weaknesses which are limiting their effectiveness), while underestimators may set extremely high expectations for themselves (and thus, strive to continually improve and develop their capabilities; Atwater et al., 1998).

The results presented by Atwater et al. (1998) offer support for the importance of simultaneously considering self and other ratings of managerial effectiveness, as well as the *magnitude* of ratings and *direction* of disagreement (i.e., self greater than other ratings vs. self less than other ratings). Their findings indicate that the relationship between self ratings, other ratings, and managerial effectiveness is more complex than previously believed. Atwater and colleagues found that effectiveness was highest for in-agreement/good estimators and underestimators. Effectiveness was lowest for overestimators when self-ratings were moderate and subordinate ratings were low. Overall, the authors noted a general trend that managerial effectiveness tended to *increase* for underestimators and *decrease* for overestimators,

indicating that individuals who underestimate themselves are typically viewed as more effective.

Several explanations are offered for the relationships found by Atwater et al. (1998). First, the authors' results supported that self-other agreement is related to higher managerial effectiveness, assuming that ratings are consistently in the positive direction (i.e., in-agreement/poor ratings had no significant relationship with effectiveness). In terms of underestimators being rated as more effective, one explanation is that these individuals are committed to continually improving themselves, not becoming overconfident in their abilities, and not becoming complacent. These individuals may also set extremely high standards and goals for themselves, which results in harsher self-ratings. For those individuals who were rated moderate or low by others, but higher by themselves (i.e., overestimators), they are likely seen as less effective because they may unknowingly possess significant weaknesses which are negatively impacting their performance. In other words, overestimators may have serious "blind spots" which limit their ability to be effective in managerial or leadership roles.

Self-Awareness of Specific Leader Behaviors

Recent research by Tekleab et al. (2008) confirmed the findings by Atwater et al. (1998) in terms of the relationship between self-other agreement and leadership effectiveness. Similar to previous research, Tekleab et al. measured self-awareness as the degree of agreement between a leader's self-description and his/her followers' perceptions of leader behaviors; however,

their focus was on the specific behaviors related to transformational and empowering leadership. Bass (1985) described the original theory of transformational leadership, in which followers feel trust, loyalty, and respect toward their leader (Yukl, 2006). Followers of transformational leaders are motivated to do more than they originally expected because their leaders articulate a clear vision, increase awareness of important task outcomes and long-term goals, and motivate followers to go beyond acting in their own self-interest for the sake of the larger organization (Howell & Avolio, 1993; Yukl, 2006). In comparison, empowering leaders delegate significant responsibilities which enable followers to satisfy their higher order needs for autonomy and growth by employing self-control and self-direction (Manz & Sims, 1987). Tekleab and colleagues (2008) examined these specific types of leadership behaviors because previous research had aggregated leadership or managerial behaviors (i.e., using a composite score) and failed to examine differences in the relationship between self-other agreement and effectiveness for different types of leadership styles.

Using the polynomial regression method outlined by Edwards (1994), Tekleab et al. (2008) found that the effects of self-awareness for transformational leadership were different than the effects of self-awareness for empowering leadership. Specifically, self-awareness of transformational leadership was related to higher leader effectiveness; however, self-awareness of empowering leadership had no significant relationship with leader effectiveness. Within transformational leadership, findings were consistent

with previous research by Atwater et al. (1998) in that the most effective leaders were those who underestimated themselves (i.e., self-ratings were significantly lower than follower ratings) as well as those who had similarly *high* self and other ratings (i.e., in-agreement/good estimators).

Overall, Tekleab and colleagues' (2008) results indicated that self-awareness of different types of leadership (e.g., transformational and empowering) have different relationships with leadership outcomes. Although self-awareness of transformational leadership is important for perceived leader effectiveness, self-awareness of other leadership behaviors (e.g., empowering) is more related to outcomes such as followers' self-leadership (e.g., self-management or self-control). In other words, self-awareness may not be critical in determining perceived effectiveness for all types of leadership behaviors.

Conclusions and Directions for Self-Awareness Research

To this point, self-other agreement research has concluded that leaders who are aligned with others (and are consistently rated above average) are more effective leaders (Atwater at al., 1998; Church, 1997; Tekleab et al., 2008). Additionally, if self-ratings are not aligned with other ratings, individuals who *underrate* themselves are more effective than individuals who *overrate* themselves. This information is potentially useful when delivering MSF to individuals because it allows them to understand that their alignment with observer groups has an impact on their effectiveness. Ideally, a coach or specialist would facilitate a conversation with the MSF recipient around the

areas in which the individual has the greatest "gaps" in self-other ratings, particularly if the individual *overrates* themselves (because this is likely to diminish their perceived effectiveness). However, the MSF recipient is likely to be confused if results reveal multiple gaps in different directions depending on the rater source examined (i.e., above average supervisor ratings, but below average direct report ratings). Additionally, the MSF recipient might feel overwhelmed if he/she has multiple gaps on multiple behaviors. It may be that self-other agreement is more critical in determining the leader's effectiveness for a certain group when considering a specific behavior (e.g., the ability to collaborate or compromise may be most important to peers). If this is the case, the leader can then examine alignment between sources and focus on uncovering the reasons behind that specific gap because agreement for that group is particularly important in determining his/her effectiveness. In fact, researchers agree that "more research is needed to determine the forms of agreement that are appropriate for different comparison groups (Atwater et al., 1998, p. 595)."

Cognitive load theory (CLT) also supports the need to examine self-other agreement for specific rater groups when interpreting MSF. CLT explains that humans are capable of processing only a limited amount of information, both in terms of our short-term memory and attention span (Rader, 1981). CLT could be applied to the delivery and interpretation of MSF in that feedback recipients are likely capable of processing only two to four pieces of complex information at one time. Each piece of information

could be thought of as one "gap" in self-other ratings, which requires the leader to interpret and analyze the potential reasons behind these discrepancies.  Other researchers agree that MSF has the potential to overwhelm managers with information, making it difficult for them to understand the feedback and use it effectively to improve performance (Nikolaou, Vakola, & Robertson, 2006).  DeNisi and Kluger (2000) suggest that organizations minimize the amount of MSF data presented to employees, or provide a personal coach to help MSF recipients interpret the substantial amount of feedback provided from various sources.  One way for personal coaches to help MSF recipients interpret their results is to narrow their attention to the specific rater groups and behaviors where alignment is most critical to their effectiveness.

Feedback intervention theory (FIT; Kluger & DeNisi, 1996) and goal setting theory (Locke & Latham, 2002) also support the value of examining specific sources and behaviors within self-other agreement.  First, FIT asserts that improvement following feedback is most likely when feedback is specific to the task.  Within the context of MSF, task-specific feedback is more likely if the leader understands the specific behavior and context in which they need to change or improve.  FIT also explains that because our attention is limited, only the feedback that receives direct attention will result in behavioral change or improvement (DeNisi & Kluger, 2000).  Thus, in order for MSF to have optimal results (e.g., behavioral change), it is important for MSF recipients to focus their attention on the feedback gaps that are most highly

related to their perceived effectiveness. Second, goal setting theory explains that establishing difficult and specific goals motivates individuals to improve (Locke & Latham, 2002). Thus, it is likely that focusing MSF recipients' attention on behavioral ratings from specific rater groups will help them set more specific goals (e.g., the need to be more collaborative and team-oriented with peers) which, according to goal setting theory, will contribute to their ability to improve because they are able to focus their attention and monitor their progress toward this specific goal.

## Rater Sources

The current study asserts that the relationship between self-other agreement and effectiveness will depend on the specific rater source and behavior examined. One of the underlying reasons for gathering MSF is to better understand the *perception* of each rater source because each group observes and perceives behaviors in a unique and valuable way. Recent research by Dierdorff and Surface (2007) supports the idea that the *context* of performance ratings is important to consider. Context can be described as "situational opportunities and constraints that affect the occurrence and meaning of organizational behavior (Johns, 2006, p. 386)." In general, differences in performance ratings can be attributed to both systematic and unsystematic variance in ratings (Scullen, Mount, & Goff, 2000). Previous research has shown that actual performance of the target individual is not the only systematic influence on ratings of performance, but that context can also systematically affect performance ratings. For example, Dierdorff and

Surface (2007) demonstrated that in "strong situations" in which peers had a clear understanding of what behaviors were expected, more variance was due to actual performance because raters understood was constituted effective versus ineffective performance. Context is believed to influence the range of behaviors that are viewed as appropriate within a specific situation and impact perceptions of what determines effective behavior within a given situation. Although their research was specific to peers, the findings could be applied to other observer groups in that each rater source (e.g., supervisor, peers, direct reports) has a unique viewpoint of what represents desirable or ideal behavior based on each group's experience and knowledge of what is required for effective performance.

Empirical evidence indicates that different rater sources provide unique information, as agreement across sources (i.e., between-source rating correlations) is generally found to be low (Conway & Huffcutt, 1997; Harris & Schaubroeck, 1988). For example, Conway and Huffcutt (1997) compared correlations between all possible combinations of self, peer, supervisor, and direct report ratings and found that between-source rating correlations were lower than within-source ratings correlations, implying that each source provides relatively distinct feedback on the target individual.

Several theories have been offered on the unique perspective that each rater source provides. Based on the findings by Mount et al. (1998), Morgeson and colleagues (2005) describe that supervisors are accustomed to conducting performance appraisals (and thus, understand what behaviors are

needed to be promoted), peers are likely to observe the individual in a high proportion of situations, and direct reports provide a unique perspective on managerial behaviors. Likewise, Greguras, Ford, and Brutus (2003) argued that certain sources may be better suited to provide feedback on specific behaviors. First, a particular group may have a better understanding of what it takes to be effective at certain behaviors (e.g., what it means to "think strategically"). Second, a particular group may have more opportunities to observe certain behaviors because they work more closely with the individuals in situations which reveal unique sides of their work style. In summary, Greguras et al. explain that *supervisors* have a unique perspective of what it takes to be promoted to the next level, *peers* often work closely with the leader in team settings, and *direct reports* have the clearest line of sight to the leaders' ability to manage and direct the work of others.

Research has also shown that individuals *attend* to feedback from different rater sources depending on the performance dimension being rated (Greguras et al., 2003). Specifically, research by Greguras et al. (2003) demonstrated that MSF recipients attended to feedback from their peers (more than supervisors and direct reports) on the dimension of "general administrative performance." This dimension included a variety of behaviors including how well the individual administers day-to-day activities, coordinates work efforts across work groups, and develops short and long-term plans. In addition, MSF recipients attended to feedback from their direct reports (more than supervisors and peers) on the dimension of "ability to lead

others" which included delegating to employees, providing opportunities to employees, and treating employees fairly. In summary, Greguras and colleagues (2003) found that feedback from peers was attended to when ratings were related to day-to-day project management skills, likely because peers have the greatest opportunity to observe the individual performing these tasks and also have the most experience performing these tasks themselves. Feedback from direct reports was considered more important when rating general leadership behaviors, likely because direct reports have the clearest perspective on the individual's ability to manage and direct their work.

Although the research by Greguras et al. (2003) answers the question of which rater source is *attended to* when leaders receive MSF, the question still remains of which rater source self-ratings need to be in agreement with in order to be most predictive of effective leadership. While Greguras et al. focused on the rater source and behavior the MSF recipient *chooses to* attend to, the current study answers the question of which rater source and behavior the individual *should* attend to (i.e., because alignment with specific groups for certain behaviors is related to the individual's perceived effectiveness). Similar to Greguras et al., the current study assumes that the importance of the rater source is dependent on the specific leadership behavior examined. For example, for certain behaviors, self-*supervisor* agreement is believed to have the strongest relationship with leader effectiveness, whereas for other behaviors, self-*peer* or self-*direct report* agreement is believed to be more strongly related to leader effectiveness.

Rationale

The current study examines the relationship between self-awareness (operationalized as self-other agreement) and leader effectiveness by hypothesizing that the relationship depends on the specific rater source and behavior being rated. Although there is a substantial amount of research on self-other agreement, most prior research has examined only one or two "other" sources (e.g., Tekleab et al., [2008] examined only direct report ratings). Even in previous studies that included a range of rater sources (e.g., Atwater et al., 1998), comparisons were not made across sources to examine whether or not a different relationship between self-other agreement and effectiveness would exist. Thus, researchers recommend that future studies investigate self-other agreement across a broad range of sources and behaviors (London & Smither, 1995; Tekleab et al., 2008). According to Church (1997), "future studies should also explore the extent to which managers differ in assessing themselves in specific content areas (e.g., task vs. people behaviors, work group climate, leadership skills, or customer service), and with respect to different constituents in the workplace (e.g., peers, subordinates, supervisors and customers; p. 289)."

When receiving MSF, individuals may be overwhelmed with the amount of feedback presented to them, and as a result, may not be able to effectively focus their development efforts to improve their performance, which is considered one of the primary benefits of MSF (Morgeson et al., 2005). Thus, MSF recipients are likely to benefit from the results of the

current study because it will help them understand where alignment is needed in order to be viewed as an effective leader. According to Greguras et al. (2003), "research should investigate the boundaries of providing feedback on numerous dimensions by numerous rater sources (p. 358)." Kluger and DeNisi (1996) explain that an individual's attention is limited and that only the feedback receiving attention can direct behavioral change or performance improvement. Therefore, in order for MSF systems to have the most positive impact on performance, both the MSF recipient and the facilitator need to understand where self-other agreement is most critical. By understanding where to *direct* the MSF recipient's attention, practitioners (e.g., executive coaches) will be able to deliver feedback in a way that optimizes the individual's resources with the goal of helping the individual set specific objectives to develop and improve their performance. Even without the assistance of an executive coach, this research has the potential to help leaders more easily interpret and analyze their MSF results. By focusing on specific "gaps" in self-other ratings, the leader will be less likely to become overloaded with the amount of feedback and more likely to focus their development efforts in areas (or behaviors) that contribute most to their effectiveness as a leader.

<u>Statement of Hypotheses</u>

Each hypothesis in the current study contains four components. Components a, b, and c describe the proposed relationship between self-other agreement and leader effectiveness (i.e., one component for each rater source).

As the primary focus of the current study, part d of each hypothesis discusses the proposed *differences* in the relationship between self-other agreement and leader effectiveness based on the source and behaviors examined (which will be described on the following pages).

For components a, b, and c, all hypotheses include an examination of both the *degree* of agreement and the *direction* of ratings (i.e., high versus low ratings). All hypotheses incorporate the following self-other agreement groups: in-agreement/high (i.e., consistently high ratings), in-agreement/low (i.e., consistently low ratings), underestimators (i.e., leaders whose self-ratings are lower than observers), and overestimators (i.e., leaders whose self-ratings are higher than observers). Consistent with previous findings (Atwater et al., 1998; Tekleab et al., 2008), this study hypothesizes that in-agreement/high, followed by underestimators, will be rated as most effective. In-agreement/high leaders are aligned with their observers and display a high amount of desirable leadership behaviors. Underestimators, although misaligned with their observers, are also likely to be seen as effective because they are likely to set high expectations and challenging goals for themselves, which causes them to continually strive for improvement (Atwater et al., 1998). Overestimators are hypothesized to be rated as least effective, followed by in-agreement/low leaders. Overestimators are likely to significantly misdiagnose their leadership strengths and weaknesses, which may cause leaders to be unreceptive to feedback and unlikely to set self-improvement goals. Lastly, although in-agreement/low leaders are aligned

with their observers' ratings, their consistently low ratings indicate that they

are not displaying enough of the critical leadership behaviors, which is likely

to limit their effectiveness in a leadership role. Further, because in-

agreement/low estimators do not observe a discrepancy in self-other ratings,

control theory suggests that these individuals are not as motivated to improve

or change their behaviors because they do not feel the tension needed to

motivate discrepancy reduction (Carver & Scheier, 1982). In summary, the

relationship described above is predicted for Hypotheses 1, 2, and 3, described

as follows.

Supervisor as the Rater Source

Traditionally, supervisors provide formal and informal feedback to

their direct reports, either on a regular basis or during formal performance

reviews (Murphy & Cleveland, 1995). In fact, employees tend to prefer using

their immediate supervisor as the main source of information for performance

evaluations. Because of their involvement with formal employee reviews,

supervisors are assumed to be familiar with the performance dimensions on

which individuals are rated, implying that they have an understanding of the

skills, capabilities, and behaviors that are needed to be effective, both in the

individual's current role and in future roles. In many cases, supervisors have

performed the job of the individual being rated and were then promoted to

their current role as their supervisor. Because of their familiarity with

required tasks and responsibilities, as well as their unique understanding of

what it takes to be effective in higher level roles, supervisors are likely to

understand some of the more complex, conceptual responsibilities of the role, such as problem-solving or long-term planning.

There are two primary reasons that alignment between self-*supervisor* ratings of conceptual behaviors is important to perceived effectiveness. First, it is believed that supervisors, compared to other rater sources, have the clearest line of sight to conceptual behaviors. For example, if the individual is creating a budget for his/her team (i.e., which requires long-term planning), he/she will likely present the proposed budget to his/her supervisor, thus exposing his/her ability to think broadly and plan ahead. Similarly, if the individual is attempting to solve a complex business problem, he/she will likely approach his/her supervisor for direction and guidance. It is during these conversations where the supervisor is exposed to the individual's ability to think strategically about the business (e.g., making decisions not just based on current needs, but on the direction and vision of the organization over the next three to five years).

Second, it is most critical to have self-*supervisor* alignment on ratings of conceptual behaviors because the supervisor typically has a unique understanding of what it takes to succeed at the next level of leadership, which likely requires broader and more strategic, innovative thinking. Most leadership experts agree that as a leader moves into larger leadership roles, one of the most critical shifts is spending less time on the tactical, day-to-day aspects of the job and significantly more time engaging in longer-range strategic planning (Charan, Drotter, & Noel, 2001). The individual's

supervisor (unlike peers and direct reports) likely has the experience needed to understand what "being strategic," for example, looks like at the next level of leadership. Further, the supervisor (unlike peers and direct reports) likely has more exposure to other executives and understands what behaviors will be effective or ineffective in different roles across the organization. Thus, being misaligned with a supervisor on ratings of conceptual behaviors might indicate that the leader misunderstands what is required of him/her as he/she advances to higher level roles. This leads to the first set of hypotheses:

*Hypothesis 1a*: A significant relationship will exist between self-supervisor agreement on ratings of conceptual behaviors and leader effectiveness, such that ratings of effectiveness will be highest for in-agreement/high leaders, second highest for underestimators, third highest for in-agreement/low leaders, and lowest for overestimators.

*Hypothesis 1b*: A significant relationship (as described above) will exist between self-peer agreement on ratings of conceptual behaviors and leader effectiveness.

*Hypothesis 1c*: A significant relationship (as described above) will exist between self-direct report agreement on ratings of conceptual behaviors and leader effectiveness.

*Hypothesis 1d*: The relationship between self-supervisor agreement on ratings of conceptual behaviors and leader effectiveness will be significantly stronger compared to self-peer and self-direct report agreement.

Peers as the Rater Source

Peers are viewed as a critical source of information because they often work closely with the leader and thus, have many opportunities to observe the individual displaying relevant leadership behaviors (Greguras et al., 2003). In fact, Murphy and Cleveland (1995) explain that peers may be the most well-informed source of performance information because of the frequency with which they work with the target individual, which allows them to observe a wide range of their interpersonal behaviors.

In terms of an individual's ability to collaborate effectively with their peers, the concept of participative leadership has been used to describe an individual's ability to involve others in decision-making processes (Yukl, 2006). Utilizing a participative leadership style has several benefits, including increased quality of decision-making because the individual consults people in different functional areas. In addition, if a leader requires cooperation from other individuals or groups, participation is a way to increase others' commitment and understanding of the task's importance. Lastly, in order to "win over" a group of individuals (e.g., if their approval is required to implement an idea), participation is a way to gain an understanding of their unique preferences and concerns so that their needs can be met and a common ground can be established.

Given the behaviors involved with participative leadership (Yukl, 2006), peers (compared to direct reports and supervisors) should have the clearest line of sight to behaviors involving collaboration. Peers have frequent

opportunities to observe the target individual in various meetings and interactions during which the individual may (or may not) actively involve others in decision-making and solicit input from others.  Peers also have ample opportunities to observe the individual's ability to resolve conflict by accommodating others' needs or being willing to find a common ground in order to resolve issues.  Because of the frequency with which peers observe the individual on a regular basis, this group is also able to see the individual's ability to relate to others on a personal level and form supportive relationships.  If an individual is misaligned with his/her peers (in terms of collaborative behaviors), this could indicate that the leader may not take advantage of opportunities to involve others in decisions, align his/her thinking with others, reach out to others for support, or accommodate others interests when needed.  Thus, misalignment with peers is likely to limit his/her effectiveness because he/she may be viewed as insensitive or unresponsive to others' needs  or as someone who may miss important ideas or information (i.e., because he/she does not effectively involve or listen to others).  This leads to the second set of hypotheses:

> *Hypothesis 2a*: A significant relationship will exist between self-supervisor agreement on ratings of *collaborative* behaviors and leader effectiveness, such that ratings of effectiveness will be highest for in-agreement/high leaders, second highest for underestimators, third highest for in-agreement/low leaders, and lowest for overestimators.

*Hypothesis 2b*: A significant relationship (as described above) will

exist between self-peer agreement on ratings of *collaborative*

behaviors and leader effectiveness.

*Hypothesis 2c*: A significant relationship (as described above) will

exist between self-direct report agreement on ratings of *collaborative*

behaviors and leader effectiveness.

*Hypothesis 2d*: The relationship between self-peer agreement on

ratings of *collaborative* behaviors and leader effectiveness will be

significantly stronger compared to self-supervisor and self-direct

report agreement.

Direct Reports as the Rater Source

Direct reports provide unique feedback information because they have

firsthand experience with the individual's ability to manage and direct the

work of others (Greguras et al., 2003). In other words, direct reports

(compared to supervisors and peers) have the most opportunities to observe

the target individual in a managerial or leadership role. Receiving feedback

from direct reports is important because managing effectively is a critical

component of a supervisor's job (Smither et al., 1995). Further, for an

individual in a managerial role, his/her success is partly determined by the

success of his/her direct reports; thus, being aligned with this group is a key

component of a manager's effectiveness.

Consistent with the findings by Greguras et al. (2003), the current

study suggests that individuals need to be aligned with their direct reports in

terms of their ability to manage and direct the work of others. It is critical that the leader is aware of whether or not he/she is displaying the behaviors needed to lead the team toward objectives, such as taking charge when appropriate, providing clear direction, setting expectations, delivering regular feedback, and monitoring progress to ensure that tasks are completed on schedule. If individuals are misaligned with their direct reports (on behaviors related to managing others), it is possible that the leader may not be taking enough authority or directing the actions of others to help the team make progress and deliver results. Misalignment could also be detrimental to the leader's effectiveness because it could be related to vague expectations, frequent misunderstandings, or missed deadlines. This leads to the study's third set of hypotheses:

*Hypothesis 3a*: A significant relationship will exist between self-supervisor agreement on ratings of *managing others* and leader effectiveness, such that ratings of effectiveness will be highest for in-agreement/high leaders, second highest for underestimators, third highest for in-agreement/low leaders, and lowest for overestimators.

*Hypothesis 3b*: A significant relationship (as described above) will exist between self-peer agreement on ratings of *managing others* and leader effectiveness.

*Hypothesis 3c*: A significant relationship (as described above) will exist between self-direct report agreement on ratings of *managing others* and leader effectiveness.

*Hypothesis 3d*: The relationship between self-direct report agreement on ratings of *managing others* and leader effectiveness will be significantly stronger compared to self-supervisor and self-peer agreement.

CHAPTER II

METHOD

Research Participants

Leadership Behaviors

Archival data were used to test the hypotheses.  Participants included 847 leaders within a large, financial organization who completed self-assessments of their leadership behaviors using the *Leadership Effectiveness Analysis* (LEA; Management Research Group, 1992) as part of a broader leadership development initiative.  Participants were identified as a leader if they managed at least one employee.  Job levels ranged from supervisor to the CEO of the organization.  The average age of participants was 44 years old ($SD = 7.6$ years), and participants had been in their current position for an average of 4 years ($SD = 4.2$ years).  Approximately 53% percent of participants were male and all participants were based in the U.S.

For each leader who completed a self-assessment, a combination of observers (i.e., supervisors, peers, and direct reports) also completed the LEA as part of a 360-degree evaluation for the leader.  Each leader selected the individuals in their observer group.  A total of 1,142 supervisor evaluations, 3,025 peer evaluations, and 3,219 direct report evaluations were obtained on the 847 leaders.  This resulted in an average of 1 supervisor, 4 peer, and 4 direct report evaluations for each leader.

<u>Leader Effectiveness</u>

In addition to rating leadership behaviors, each leader's supervisor, peers, and direct reports completed a separate survey to assess leader effectiveness. According to research by Fleenor et al. (1996) and Atwater et al. (1998), supervisor ratings are the most preferred source for ratings of overall performance because supervisors are ultimately in the position to make promotion and salary decisions which represent the leader's success or effectiveness. Accordingly, supervisor ratings of effectiveness were used to test the study's hypotheses; however, exploratory analyses were also conducted to examine the appropriateness of using an aggregate across all rater sources (i.e., supervisors, peers, and direct reports).

<div align="center">Measures</div>

<u>Leadership Behaviors</u>

Each leader completed self-ratings using the LEA (Management Research Group, 1992), while supervisors, peers, and direct reports completed the observer version of the same questionnaire. The purpose of the LEA is to provide information to leaders on how his/her self-perception compares to the perceptions of various stakeholders who work closely with the leader across a variety of situations. The LEA is a descriptive, behaviorally-oriented instrument which provides scores on a wide range of leadership behaviors (Kabacoff, 1998).

The LEA includes 22 behavioral dimensions of leadership. For the purpose of this study, a subset of the LEA dimensions was used to test the

hypotheses. Eight LEA dimensions were the focus of this study because they represent *conceptual* behaviors, *collaborative* behaviors, and behaviors related to *managing others*. C*onceptual* behaviors are those related to making decisions, planning, and solving problems. Two behaviors on the LEA are related to *conceptual* behaviors: strategic and innovative. Strategic is defined as the tendency to "take a long-range, broad approach to problem solving and decision making through objective analysis, thinking ahead, and planning" (Kabacoff, 1998). An example item is: "In general, others see me as planning for the future." Innovative is described as someone who is willing to take a new approach to solving problems and is able to develop creative solutions. An example item is: "This person is an innovative thinker."

C*ollaborative* behaviors are defined as behaviors displayed when working in a team setting. Three behaviors on the LEA are related to *collaborative* behaviors: cooperation, consensual, and empathy. Cooperation is described as the ability to "accommodate the needs and interests of others by being willing to defer performance on one's own objectives in order to assist colleagues with theirs" (Kabacoff, 1998). An example item is: "This person is a helpful teammate." Leaders who exhibit consensual behaviors are described as "valuing the ideas and opinions of others and collecting others' input as part of their decision-making process." An example item on the LEA is: "When in charge, this person tries to get the ideas of his/her colleagues." Lastly, empathy is described as "demonstrating an active concern for people and their needs by forming close and supportive relationships with others" and

an example item is: "People are likely to be impressed by my genuine interest in them."

*Managing others* are behaviors displayed when managing and/or directing the work of others. Three behaviors on the LEA represent behaviors related to *managing others*: management focus, production, and feedback. A leader who is described as demonstrating management focus is someone who "seeks to exert influence by being in positions of authority, taking charge, and leading and directing the efforts of others" (Kabacoff, 1998). An example item is: "In difficult situations, this person displays a willingness to take command." Production is defined as "adopting a strong orientation toward achievement; pushing yourself and others to achieve at high levels." An example item is: "This person is a hard driving achiever." Feedback is described as "letting others know in a straightforward manner what you think of them, how well they have performed, and if they have met your needs and expectations." An example item is: "This person lets people know how they are performing."

The LEA uses a normative, semi-ipsative item format. The normative scale allows comparisons of the individual to a large sample of other leaders (i.e., over 70,000) who have completed the questionnaire. Semi-ipsative is defined as a combination of an ipsative or forced-choice scale (used to measure relative preference among answer options) and an anchored rating scale (used to measure magnitude of preference). The normative, semi-ipsative format has several advantages (Mahoney & Mahoney, 1996). For

example, the partial forced choice format has been shown to minimize

response set distortions such as acquiescence, nay-saying, and social

desirability.  Combining the two methodologies also shows the relative

preference within individuals (i.e., forced choice) as well as the magnitude

(i.e., anchored ratings) of an individual's preference.  The semi-ipsative

format is also associated with lower scale inter-correlations, which indicates

independent dimensions (Kabacoff, 1998; Mahoney & Mahoney, 1996).

The semi-ipsative format is described as follows.  Each question

consists of a stem (e.g., "In supervising people, I am…") and three alternative

options, each of which represents a different leadership behavior (e.g.,

"tactful, demanding, easy to please").  First, the participant is instructed to

choose the option which seems *most* characteristic of him/her and rate it as

either a "5" or a "4," where "5" represents most characteristic.  Then, the

participant is instructed to select the option that is *next* most characteristic of

him/her and rate it as either a "3" or a "2," where "3" represents more

characteristic.  The participant is told to leave the third option blank, and this

option receives a score of "0."  In terms of item scoring, each response

receives a score of 5, 4, 3, 2, or 0 (based on the previous description).  Each

LEA behavior (i.e., Strategic) includes eleven scale items; therefore, each

behavior receives a total score ranging from 0 to 55.  Raw scores for each

behavior are then compared to a normative database of over 70,000

individuals who have completed the survey and a percentile rank is calculated.

The percentile rank represents the relative importance placed on the behavior

compared to a large group of other individuals who have completed the survey. Percentile rank scores were used as the dimensions ratings for the LEA.

In terms of scale development, the LEA was developed in 1986 by a group of organizational consultants and psychologists with the purpose of measuring a broad range of behaviors and practices that tended to lead to success over a wide variety of management situations (Kabacoff, 1998). During the empirical phase of item development, an initial sample of 200 leaders were administered the questionnaire. Items were included in the final version of the LEA if they met the following standards. First, the item was correlated with its target set of items in the range of .30 to .60. Second, the item was essentially uncorrelated (close to zero) with any other item set. Third, the item contributed to the internal consistency (i.e., coefficient alpha) of the scale. And fourth, the item was judged to be appropriate by a panel of experts (i.e., organizational psychologists and senior organization consultants at two consulting firms).

Previous research by Kabacoff (1998) has established strong reliability and validity of the LEA. First, in terms of test-retest reliability, two separate test-retest studies were conducted in 1991 and 1997. The combined results produced test-retest reliability coefficients ranging from .59 to .86 (uncorrected for attenuation), depending on the specific leadership behavior examined. The average test-retest coefficient (i.e., across all leadership behaviors) was .78. Second, several studies of inter-rater reliability were

conducted on the observer version of the LEA using a large database that

represented a wide range of companies, business functions, and geographic

locations (Kabacoff, 1998). Intra-class correlation coefficients were used to

measure inter-rater reliability. For supervisor ratings on the LEA, mean inter-

rater reliabilities ranged from .58 for two raters to .80 for four raters. For peer

ratings, mean inter-rater reliabilities ranged from .67 for four raters to .80 for

eight raters. Lastly, for direct report ratings, mean inter-rater reliabilities

ranged from .66 for four raters to .79 for eight raters.

Several studies have examined the construct validity of the LEA.

First, a multi-trait/multi-method (MTMM) matrix was used to examine the

construct validity of the LEA in a sample of over 120,000 individuals. Rater

groups (i.e., self, supervisor, peer, and direct reports) represented the methods

and each leadership behavior represented the traits. The patterns of the

correlations in the MTMM matrix supported both convergent and discriminant

validity of the measure (Kabacoff, 1998). Second, relationships have been

examined between the LEA self questionnaire and other assessment

instruments (i.e., Sixteen Personality Factors Questionnaire, Wesman

Personnel Classification Test (WPCT), Individual Directions Inventory (IDI))

in a sample of 464 individuals who completed two or more tests as part of

selection or development processes. Although the behaviors in the LEA are

not considered personality variables, this validation study anticipated small to

moderate correlations between certain LEA behaviors and personality

indicators. For example, as expected, the highest correlation was found

between the measure of "innovative" on the LEA and that of "creating" on the IDI ($r = .67$). In addition, the study expected zero or very small correlations between the LEA behaviors and cognitive abilities. In support, there were no significant correlations found between any LEA behaviors and either Verbal or Numerical sub-tests on the WPCT.

Leader Effectiveness

The observer version of the LEA also contains a separate 22-item questionnaire to assess the effectiveness of the leader being rated. Although the measure was designed to capture a wide range of leader effectiveness behaviors, four items were intended to measure "overall effectiveness," which included credibility with management and ability to inspire confidence with superiors (i.e., communicates well, delivers on promises, thinks in similar ways), overall effectiveness as a leader/manager (i.e., total level of performance against expectations, total impact in role), future potential (i.e., has the ability to go beyond present level versus has reached his/her highest potential, is likely to be a major resource to the organization) and ability to make effective decisions. Each item was ranked on a 7-point Likert scale and behavioral anchors were unique to the question being asked. For example, when asked about "credibility with management and ability to inspire confidence with superiors," anchors ranged from "has little credibility" to "inspires complete confidence."

Procedure

Participants completed the LEA questionnaire as part of their ongoing leadership development program within the organization.  Each participant selected a group of observers who then agreed to complete the observer version of the LEA.  Observers were informed that evaluations would remain anonymous and would be used for the purposes of delivering 360-degree feedback to the leader for his/her development.  Both self and observer questionnaires were completed online and took approximately 30 minutes to complete.  All questionnaires were completed between 1996 and 2008.

In addition to the LEA, each individual's supervisor, peers, and direct reports provided ratings of leader effectiveness on a separate survey which was administered at the same time as the observer version of the LEA. Participants did not complete a self-version of the leader effectiveness survey. Separate instructions were provided for the leader effectiveness measure. Participants were instructed to "answer the following questions related to the effectiveness of the person you are rating."  Instructions also provided a frame of reference by asking the participant to "think of an imaginary, average leader/manager who would be considered moderately successful in his/her field.  This person would be placed at point 4 ("average") on the 7-point scale."

In terms of the structure of the data, each "case" (i.e., individual who received 360-degree feedback) is composed of a self-rating and other-ratings from a combination of observers (i.e., supervisor, peers, direct reports).  Thus,

each case includes a self-rating, average supervisor rating, average peer rating, and average direct report rating on each of the following LEA behavioral dimensions: strategic, innovative, cooperation, consensual, empathy, management focus, production, and feedback.  Also, each case includes an average supervisor, average peer, and average direct report rating on each of the leader effectiveness items (i.e., 22-item scale).

CHAPTER III

RESULTS

Results include a series of preliminary analyses conducted prior to testing the hypotheses. First, results are reported from an Exploratory Factor Analysis conducted to determine the extent that the eight LEA behaviors clustered into the three theoretical groupings used to organize the hypotheses. Second, an Exploratory Factor Analysis was conducted to explore the extent to which the four items intended to measure overall leader effectiveness grouped together into one factor. Third, a test of measurement equivalence was conducted to examine the extent to which different rater groups (i.e., supervisors, peers, direct reports) interpreted the leader effectiveness items similarly. Lastly, hypotheses were tested using polynomial regression analyses.

<u>Preliminary Analyses</u>

<u>Factor Analysis of Leadership Behaviors</u>

Hypotheses were theoretically organized in three groups, labeled as follows: *conceptual* behaviors (which consists of two dimensions: strategic and innovative), *collaborative* behaviors (which consists of three dimensions: cooperation, consensual, and empathy), and behaviors related to *managing others* (which consists of three dimensions: management focus, production, and feedback). An Exploratory Factor Analysis (EFA) was conducted to determine the extent to which the LEA dimensions clustered into the three broader groups. Hypothesis testing was conducted on each of the eight

separate LEA dimensions (i.e., to examine unique relationships for each LEA

behavioral dimension); however, the study's hypotheses were organized in a

way that assumed the eight dimensions clustered into these three broader

groups, thus the EFA was conducted to confirm this hypothesized framework.

The SPSS program was used to conduct an EFA using the principal

components analysis as the extraction method and a varimax rotation. The

EFA was conducted on the entire dataset (i.e., all three rater sources

combined) because hypotheses were organized in the same three theoretical

groupings for all rater groups. To determine the appropriate factor solution,

the eigenvalues, scree plot, and percent of variance explained were examined.

First, Kaiser's criterion states that only factors with an eigenvalue of 1.0 or

greater should be retained in the factor analysis (Tinsley & Tinsley, 1987).

Following Kaiser's criterion, the current analysis produced a two-factor

solution ($\lambda_1 = 2.75$, $\lambda_2 = 1.46$). However, a three-factor solution also produced

an eigenvalue very close to the 1.0 cutoff ($\lambda_3 = .94$) and thus, a three-factor

solution was considered as a viable option. Second, the scree plot indicated

that the point at which the curve levels off (or becomes horizontal) was at

three factors. Interpretation of the scree plot implies that all factors to the left

of the scree (i.e., the point at which the curve levels off) are considered real

factors, while all factors to the right of the scree are considered error factors

(Tinsley & Tinsley, 1987). Third, the percentage of variance accounted for by

the factor solution was examined. Tinsley and Tinsley (1987) explain that

there is no firm theoretical guideline for establishing a limit; however, higher

percentages of explained variance are desirable because this indicates a lower percentage of common variance that is unexplained.  In this case, the percent of variance accounted for by the factors was approximately 52.6% for a two-factor solution and 64.3% for a three-factor solution.  Given that this percentage increases to above 60% for the three-factor solution, combined with results from the eigenvalue and scree plot analyses, the three-factor solution was determined to be the most appropriate fit to the data.

Examining the structure matrix for the rotated solution also provided theoretical evidence for the three-factor solution.  Thurston's criteria is the most widely accepted standard for determining a good factor structure, which states that each variable should load highly on only one factor (Tinsley & Tinsley, 1987).  In this case, each item (i.e., LEA behavior) had a relatively high loading on their respective factor, and a relatively low loading on the other factors, fulfilling Thurstone's criteria of finding the simplest structure.  Two behaviors (strategic and innovative) correlated with Factor I, which is consistent with the *conceptual* behaviors grouping.  Although innovative behaviors also loaded on the *managing others* factor, the loading was much stronger for the *conceptual* grouping (e.g., .70 for *conceptual* versus .37 for *managing others)*.  Three behaviors (cooperation, consensual, and empathy) correlated with Factor II, which is consistent with the *collaborative* behaviors grouping.  Lastly, three behaviors (management focus, production, and feedback) correlated with Factor III, which is consistent with the *managing others* grouping.  Table 1 reports the factor loadings for the three factors

described above.  Thus, results of the EFA support the theoretical groupings

of leadership behaviors used for the hypotheses.

Table 1

*Factor Loadings for LEA Behaviors*

| LEA Behavior | Factor I (*Conceptual*) | Factor II (*Collaborative*) | Factor III (*Managing Others*) |
|---|---|---|---|
| Strategic | **.86** | -.02 | -.17 |
| Innovative | **.70** | .13 | .37 |
| Cooperation | .04 | **.76** | -.34 |
| Consensual | .19 | **.75** | -.18 |
| Empathy | -.09 | **.81** | -.11 |
| Management Focus | .21 | -.28 | **.70** |
| Production | .13 | -.25 | **.65** |
| Feedback | -.27 | -.12 | **.73** |

Factor Analysis of Leader Effectiveness Measure

Exploratory and confirmatory factor analyses were conducted to determine the survey items that should be used to measure the underlying construct of leader effectiveness. First, an EFA was conducted with four items related to overall effectiveness. The SPSS program was used to conduct the EFA using the principal components analysis as the extraction method. A varimax rotation was performed on all factors satisfying Kaiser's criterion. Although supervisor ratings were of primary interest to the hypothesis testing, separate EFAs were conducted for each rater group (i.e., supervisor, peer, and direct report) and a test of measurement equivalence was conducted to determine the suitability of combining across the rater groups.

To determine the appropriate factor structure, the eigenvalues, scree plot, and percent of variance were examined (Tinsley & Tinsley, 1987). First, using an eigenvalue-of-greater-than-one criterion, the EFA produced a one-factor solution for all three rater groups (supervisors: $\lambda_1 = 2.82$; peers: $\lambda_1 = 2.94$; direct reports: $\lambda_1 = 3.00$). The second component had an eigenvalue that was not close to 1.0 (for all rater groups); thus, a one-factor solution seemed most appropriate. Second, examination of the scree plots indicated that the point at which the curve levels off was at one factor (for all rater groups). This was consistent with the interpretation of the eigenvalues. Third, the percent of variance accounted for by the factor was approximately 70.5% for the supervisor group, 73.4% for the peer group, and 74.9% for the direct report group. Therefore, following these criteria for the EFA results, a one-

factor solution seemed most appropriate. In addition, coefficient alphas were calculated on the scores of the four-item scale of leader effectiveness, which indicated acceptable internal consistency (supervisors: $\alpha = .85$; peers: $\alpha = .87$; direct reports: $\alpha = .88$).

With the goal to confirm a well-fitting measurement model of leader effectiveness, confirmatory factor analysis (CFA) was conducted separately for each rater group. The LISREL program was used to conduct the CFA, using the maximum likelihood estimation technique. Because relatively large sample sizes were available, each rater group was randomly split in half to form a calibration sample (Group 1) and a holdout sample (Group 2). The calibration sample was used to test the degree of fit for the one-factor model which was tested in the EFA, whereas the holdout sample was used to cross-validate the model among an independent sample (Facteau & Craig, 2001). The purpose of this methodology was to minimize the chance that the previous analyses capitalized on chance versus reflecting the true model underlying the data (Byrne, 1989).

The fit of the measurement model in both the calibration and holdout samples is shown in Table 2. As seen in Table 2, the measurement model cross-validated well in the respective holdout samples for each rater group. First, it was evident that the parameter estimates for each proposed relationship were significant, as the t-values for each of the factor loadings were significant at $p < .05$ for all of the subgroups tested. Second, for most of the measurement models, the Chi-Square statistic was significant; however,

the Chi-Square was not interpretable in this case because the sample size was large (*N* was approximately 300 to 350 for each subgroup). As evidence, the Critical N (CN) was much lower than the actual sample size submitted to the CFA, indicating that a smaller sample size would guarantee a significant Chi-Square, regardless of model fit. Thus, other fit statistics were more interpretable. Other fit indices examined were the NFI, CFI and GFI, which were all well above .90 (as seen in Table 2), indicating that the model provided a good fit for the data from all rater groups. The RMSEA statistic is commonly reported as a measure of discrepancy between the model and the data. Some authors argue that a value below .08 indicates an acceptable model fit (Browne & Cudeck, 1993). For the supervisor group, RMSEA values exceed the acceptable limit according to the .08 cutoff. However, the RMSEA statistic can be inflated when the degrees of freedom are small (Kenny, 2008). Because the analyses included only two degrees of freedom, the RMSEA values are likely inflated. Thus, other fit statistics (i.e., NFI, CFI, and GFI) are more appropriate and confirm that the one-factor (4-item) model provides an acceptable fit to the data. Further, the fact that this model was confirmed among both the calibration and holdout samples provides initial evidence that a common factor structure underlies all raters' responses on these four items.

Table 2

*Results of Confirmatory Factor Analyses Conducted to Establish Baseline*

*Measurement Models Within Each Rater Group for the LE Measure*

| Rater group and sample | $\chi^2$ | *df* | NFI | CFI | GFI | RMSEA |
|---|---|---|---|---|---|---|
| Supervisors | | | | | | |
| Calibration Sample | 39.48* | 2 | .94 | .94 | .94 | .24 |
| Holdout Sample | 56.67* | 2 | .92 | .92 | .93 | .28 |
| Peers | | | | | | |
| Calibration Sample | 2.92 | 2 | 1.00 | 1.00 | 1.00 | .04 |
| Holdout Sample | 7.31* | 2 | .99 | .99 | .99 | .08 |
| Direct Reports | | | | | | |
| Calibration Sample | 6.87* | 2 | .99 | .99 | .99 | .08 |
| Holdout Sample | 1.61 | 2 | 1.00 | 1.00 | 1.00 | .00 |

*Note.* LE = Leader Effectiveness; NFI = normed fit index; CFI = comparative

fit index; GFI = goodness of fit index; RMSEA = root mean square error of

approximation.

* $p < .05$.

Measurement Equivalence of Leader Effectiveness Measure

The next step in analyzing the data was to test for measurement invariance across rater groups on the leader effectiveness measure. Measurement equivalence does not require that the distributional properties of the measure (e.g., means, variances) are identical across groups. Instead, it requires that the empirical relationship between indicators and the latent constructs they are intended to measure are equal (Drasgow & Kanfer, 1985). Without measurement invariance, observed scores from different groups cannot be aggregated and are not directly comparable.

The most widely held standard of measurement equivalence, which is the establishment that factor loadings are invariant across groups (i.e., metric invariance), was used to determine measurement equivalence. Although the Chi-Square statistic is the most widely used test for comparing nested models in multi-group analyses, Chi-Square is highly sensitive to sample size. Thus, the criteria specified by Cheung and Rensvold (1999) were used, which asserts that changes in the comparative fix index (CFI) of .01 or less provide evidence of invariance across groups. Results of the measurement equivalence analysis are shown in Table 3, which indicate that the metric invariance model resulted in a significant loss of fit (in terms of changes in the CFI).

The measurement equivalence analysis failed to demonstrate invariant factor loadings, which indicates that the leader effectiveness scale might be used differently in these three rater groups. Observed scores from the three

groups cannot be aggregated and are not directly comparable.  Thus,

hypotheses were tested using supervisor ratings of effectiveness rather than a

composite score, and any exploratory analyses using the composite leader

effectiveness ratings could not be conducted.

Table 3

*Fit Indices for the Assessment of Measurement Equivalence across Rater*

*Groups for the LE Measure*

| Model description | $\chi^2$ | df | $\Delta\chi^2$ | RMSEA | TLI | CFI |
|---|---|---|---|---|---|---|
| Configural invariance | 54.36* | 6 | -- | .10 (.08, .13) | .97 | .99 |
| Metric invariance | 320.42* | 12 | 266.06* | .19 (.17, .20) | .91 | .94 |

*Note.* LE = Leader Effectiveness; RMSEA = root mean square error of

approximation (90% confidence interval for RMSEA is in parentheses); TLI =

Tucker–Lewis index; CFI = comparative fit index.

* $p < .05$.

<u>Test of Hypotheses</u>

Analytical procedures recommended by Edwards (1994) were used to test the relationship between self-other agreement and leader effectiveness. Edwards highlights deficiencies with typical congruence measures (e.g., difference scores) and provides solutions for assessing congruence using unconstrained regression analyses. While traditional congruence measures use a set of implied constraints, Edwards suggests that these constraints are freed and then examined to explain relationships between variables. Beyond the methodological problems associated with traditional difference scores (e.g., less reliable, confounding effects), Edwards also explains that they oversimplify what is likely a three-dimensional relationship to only two dimensions. Thus, polynomial regression allows researchers to consider the joint effects of the components on an outcome as a three-dimensional surface.

All hypotheses suggested an examination of both the *sign* (i.e., direction of discrepancy between the self and other ratings) and *magnitude* (i.e., size of the discrepancy between self and other ratings) of agreement. Hypotheses (parts a, b, and c) predicted that ratings of effectiveness would be highest for in-agreement/high rated leaders, second highest for underestimators, third highest for in-agreement/low rated leaders, and lowest for overestimators.

To examine the effect of self-other agreement (in terms of both sign and magnitude) on leader effectiveness, the model implied by the squared difference score was tested (Edwards, 1994). The squared difference model

requires a quadratic equation which captures curvilinearity and allows a test of

the hypothesis that the surface changes shape along the line of perfect

congruence. To test this, the following equation was used:

$$Z = b_0 + b_1X + b_2Y + b_3X^2 + b_4XY + b_5Y^2 + e \qquad (1)$$

where $Z$ = leader effectiveness, $X$ = self-ratings on the LEA behavior being

examined, and $Y$ = "other" (i.e., supervisor, peer, or direct report) ratings on

the same LEA behavior. Tests of the following implied constraints would

allow for support for the underlying model testing squared difference scores:

(1) the coefficients of $X^2$ and $Y^2$ are equal, (2) the coefficient on $XY$ is twice as

large as the coefficient on either $X^2$ or $Y^2$ and is opposite in sign, and 3) the

coefficients on $X$ and $Y$ are zero. In addition, the model must account for a

significant amount of variance beyond the simpler model represented by the

algebraic difference score. For the algebraic difference score model, the

following equation was tested (Edwards, 1994):

$$Z = b_0 + b_1X + b_2Y + e \qquad (2)$$

where $Z$ = leader effectiveness, $X$ = self-ratings on the LEA behavior being

examined, and $Y$ = "other" (i.e., supervisor, peer, or direct report) ratings on

the same LEA behavior. To test whether the model implied by the algebraic

difference score is viable, the following conditions must be met: (1) variance

explained by both $X$ and $Y$ must be significant, (2) $X$ and $Y$ must both

contribute a significant effect, and (3) the coefficients for $X$ and $Y$ are opposite

in sign, but not significantly different in magnitude. Testing the algebraic

difference model allows for the more parsimonious explanation that self-

ratings and/or other-ratings of leadership behaviors (regardless of the direction

of agreement) have an impact on leader effectiveness ratings.

In preparation for the polynomial regression analysis, missing cases

were removed using listwise deletion which resulted in a final sample of 732.

Means, standard deviations, and correlations were calculated for all variables

included in the regression analyses (see Table 4). For all hypotheses, the

independent variable was defined as self-other agreement on the specific LEA

behavior and the dependent variable was defined as supervisor ratings of

leader effectiveness. As suggested by Edwards (1994), *X* and *Y* were centered

on the mean of their means to reduce multicollinearity and to maintain

interpretability at the line of congruence. Additionally, upon examining

normality of the distributions for all variables, it was discovered that the

dependent variable (i.e., supervisor ratings of leader effectiveness) displayed

significantly negatively skewed data. In other words, most ratings clustered at

the high end of the 7-point scale. To normalize the data, each score on the

dependent variable was cubed, which resulted in a more normal distribution of

leader effectiveness ratings.

Results are organized as follows. Tables 5 and 6 represent results for

Hypotheses 1a, 1b, 1c and 1d, which refer to ratings of *conceptual* behaviors

(i.e., strategic and innovative). Tables 7, 8, and 9 represent results for

Hypotheses 2a, 2b, 2c and 2d, which refer to ratings of *collaborative*

behaviors (i.e., cooperative, consensual, and empathy). Tables 10, 11, and 12

represent results for Hypotheses 3a, 3b, 3c and 3d, which refer to ratings of

behaviors related to *managing others* (i.e., management focus, production, and feedback).  Hypotheses were conducted on each separate LEA dimension in order to examine the unique relationship between self-other agreement and leader effectiveness for specific leader behaviors.  Broader groupings of the hypotheses (i.e., conceptual, collaborative, and managing others) were simply used as a framework for the hypotheses based on proposed theoretical differences.

Table 4

*Means, Standard Deviations, and Correlations for All Study Variables*

| Variables | 1. | 2. | 3. | 4. | 5. | 6. | 7. | 8. | 9. | 10. | 11. | 12. |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1. Strategic (Self) | — | | | | | | | | | | | |
| 2. Strategic (Supervisor) | .33* | — | | | | | | | | | | |
| 3. Strategic (Peer) | .37* | .43* | — | | | | | | | | | |
| 4. Strategic (DR) | .33* | .33* | .46* | — | | | | | | | | |
| 5. Innovative (Self) | .24* | .05 | .11* | .06 | — | | | | | | | |
| 6. Innovative (Supervisor) | .19* | .34* | .20* | .12* | .33* | — | | | | | | |
| 7. Innovative (Peer) | .13* | .11* | .26* | .08 | .35* | .45* | — | | | | | |
| 8. Innovative (DR) | .11* | .08 | .12* | .20* | .39* | .42* | .52* | — | | | | |
| 9. Cooperation (Self) | -.15* | -.01 | .03 | -.03 | -.28* | -.13* | -.17* | -.21* | — | | | |
| 10. Cooperation (Supervisor) | -.06 | -.02 | .01 | -.01 | -.09 | -.23* | -.21* | -.22* | .31* | — | | |
| 11. Cooperation (Peer) | -.01 | .01 | .07 | .09 | -.06 | -.19* | -.27* | -.21* | .37* | .47* | — | |
| 12. Cooperation (DR) | -.07 | -.03 | .02 | .04 | -.09 | -.11* | -.16* | -.19* | .37* | .39* | .50* | — |
| *Mean* | 63.15 | 65.29 | 62.10 | 62.01 | 55.80 | 59.03 | 53.45 | 54.65 | 58.06 | 55.19 | 57.63 | 56.75 |
| *Standard Deviation* | 27.90 | 26.60 | 19.85 | 19.37 | 27.64 | 29.51 | 23.13 | 22.87 | 29.13 | 26.27 | 20.29 | 20.35 |

*Note*. $N = 732$. *$p < .05$. DR = Direct Report.

Table 4 (Continued)

*Means, Standard Deviations, and Correlations for All Study Variables (Continued)*

| Variables | 1. | 2. | 3. | 4. | 5. | 6. | 7. | 8. | 9. | 10. | 11. | 12. |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 13. Consensual (Self) | -.04 | .05 | .07 | .03 | -.05 | .01 | -.06 | -.04 | .31* | .15* | .19* | .21* |
| 14. Consensual (Supervisor) | .03 | .15* | .02 | .03 | -.01 | -.07 | -.12* | -.14* | .18* | .54* | .34* | .29* |
| 15. Consensual (Peer) | .07 | .05 | .14* | .11* | .01 | -.07 | -.10* | -.12* | .23* | .34* | .62* | .32* |
| 16. Consensual (DR) | .00 | .04 | .03 | .19* | -.04 | -.08 | -.10* | -.06 | .22* | .28* | .33* | .56* |
| 17. Empathy (Self) | -.22* | -.15* | -.16* | -.11* | -.21* | -.15* | -.12* | -.07 | .39* | .26* | .32* | .30* |
| 18. Empathy (Supervisor) | -.09 | -.16* | -.16* | -.13* | -.05 | -.16* | -.08 | -.08 | .21* | .48* | .33* | .28* |
| 19. Empathy (Peer) | -.08* | -.09* | -.10* | -.04 | -.06 | -.15* | -.10* | -.06 | .24* | .29* | .62* | .38* |
| 20. Empathy (DR) | -.13* | -.14* | -.14* | -.07 | -.07 | -.11* | -.09 | -.05 | .24* | .28* | .37* | .58* |
| 21. MF (Self) | .03 | -.03 | -.02 | -.02 | .06 | .04 | .08 | .09 | -.36* | -.20* | -.29* | -.27* |
| 22. MF (Supervisor) | -.01 | -.02 | -.03 | -.02 | .01 | .13* | .12* | .14* | -.26* | -.55* | -.44* | -.37* |
| 23. MF (Peer) | -.04 | -.03 | -.04 | -.07 | .03 | .08 | .22* | .17* | -.31* | -.42* | -.69* | -.48* |
| 24. MF (DR) | .00 | -.01 | -.04 | -.02 | .05 | .05 | .12* | .18* | -.36* | -.36* | -.48* | -.70* |

*Note*. $N = 732$. *$p < .05$. DR = Direct Report. MF = Management Focus.

Table 4 (Continued)

*Means, Standard Deviations, and Correlations for All Study Variables (Continued)*

| Variables | 1. | 2. | 3. | 4. | 5. | 6. | 7. | 8. | 9. | 10. | 11. | 12. |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 25. Production (Self) | .12* | -.03 | .00 | -.02 | .18* | .08 | .10* | .14* | -.37* | -.19* | -.22* | -.23* |
| 26. Production (Supervisor) | .02 | .03 | .04 | -.02 | .00 | .11 | .09 | .11* | -.16* | -.37* | -.29* | -.25* |
| 27. Production (Peer) | -.06 | -.03 | -.05 | -.11 | .01 | .05 | .17* | .12* | -.23* | -.31* | -.54* | -.36* |
| 28. Production (DR) | .04 | .01 | -.01 | -.13* | .03 | .03 | .13* | .08 | -.27* | -.26* | -.33* | -.57* |
| 29. Feedback (Self) | -.08 | -.10* | -.13 | -.11 | .09 | .02 | .15* | .13* | -.37* | -.32* | -.39* | -.34* |
| 30. Feedback (Supervisor) | -.05 | -.22* | -.22* | -.15* | .03 | -.02 | .07 | .05 | -.24* | -.43* | -.41* | -.33* |
| 31. Feedback (Peer) | -.13* | -.20* | -.33* | -.22* | .00 | -.03 | .07 | .05 | -.24* | -.37* | -.58* | -.39* |
| 32. Feedback (DR) | -.12* | -.19* | -.23* | -.24* | .00 | -.04 | .03 | .04 | -.21* | -.30* | -.39* | -.48* |
| 33. LE (Supervisor) | .12* | .46* | .30* | .25* | -.06 | .22* | .07 | .11* | -.07 | -.17* | -.09 | -.12* |
| 34. LE (Peer) | .09 | .28* | .51* | .29* | -.02 | .11* | .25* | .16* | -.07 | -.10* | -.08 | -.07 |
| 35. LE (DR) | .08 | .19* | .28* | .51* | .01 | .07 | .11* | .29* | -.11* | -.12* | -.09 | -.06 |

*Note*. $N = 732$. *$p < .05$. DR = Direct Report. LE = Leader Effectiveness.

Table 4 (Continued)

*Means, Standard Deviations, and Correlations for All Study Variables (Continued)*

| Variables | 13. | 14. | 15. | 16. | 17. | 18. | 19. | 20. | 21. | 22. | 23. | 24. |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 13. Consensual (Self) | — | | | | | | | | | | | |
| 14. Consensual (Supervisor) | .21* | — | | | | | | | | | | |
| 15. Consensual (Peer) | .22* | .34* | — | | | | | | | | | |
| 16. Consensual (DR) | .35* | .30* | .38* | — | | | | | | | | |
| 17. Empathy (Self) | .25* | .18* | .21* | .25* | — | | | | | | | |
| 18. Empathy (Supervisor) | .15* | .33* | .27* | .25* | .45* | — | | | | | | |
| 19. Empathy (Peer) | .19* | .31* | .51* | .32* | .47* | .51* | — | | | | | |
| 20. Empathy (DR) | .20* | .22* | .30* | .42* | .50* | .45* | .58* | — | | | | |
| 21. MF (Self) | -.15* | -.14* | -.18* | -.14* | -.19* | -.14* | -.17* | -.11* | — | | | |
| 22. MF (Supervisor) | -.08 | -.40* | -.34* | -.24* | -.21* | -.46* | -.38* | -.29* | .26* | — | | |
| 23. MF (Peer) | -.16* | -.32* | -.53* | -.28* | -.24* | -.28* | -.52* | -.35* | .36* | .52* | — | |
| 24. MF (DR) | -.22* | -.29* | -.33* | -.47* | -.26* | -.31* | -.40* | -.50* | .34* | .46* | .58* | — |
| *Mean* | 60.33 | 57.04 | 56.44 | 57.72 | 54.21 | 52.79 | 52.83 | 56.39 | 59.26 | 60.26 | 56.15 | 54.52 |
| *Standard Deviation* | 28.19 | 25.34 | 18.53 | 19.40 | 28.92 | 25.45 | 20.24 | 20.28 | 29.13 | 25.94 | 21.09 | 21.29 |

*Note.* $N = 732$. *$p < .05$.  DR = Direct Report.  MF = Management Focus.

Table 4 (Continued)

*Means, Standard Deviations, and Correlations for All Study Variables (Continued)*

| Variables | 13. | 14. | 15. | 16. | 17. | 18. | 19. | 20. | 21. | 22. | 23. | 24. |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 25. Production (Self) | -.21* | -.12* | -.12* | -.10 | -.25* | -.09 | -.13* | -.12* | .30* | .15* | .21* | .22* |
| 26. Production (Supervisor) | -.07 | -.33* | -.19* | -.16* | -.14* | -.27* | -.22* | -.18* | .14* | .40* | .27* | .27* |
| 27. Production (Peer) | -.11* | -.19* | -.40* | -.22* | -.21* | -.25* | -.39* | -.25* | .26* | .31* | .53* | .38* |
| 28. Production (DR) | -.13* | -.19* | -.21* | -.39* | -.18* | -.16* | -.25* | -.36* | .24* | .27* | .36* | .50* |
| 29. Feedback (Self) | -.22* | -.21* | -.24* | -.25* | -.34* | -.24* | -.32* | -.30* | .19* | .30* | .36* | .35* |
| 30. Feedback (Supervisor) | -.17* | -.27* | -.31* | -.23* | -.16* | -.24* | -.30* | -.24* | .18* | .38* | .34* | .31* |
| 31. Feedback (Peer) | -.23* | -.27* | -.46* | -.33* | -.19* | -.22* | -.43* | -.28* | .21* | .35* | .53* | .42* |
| 32. Feedback (DR) | -.16* | -.21* | -.26* | -.30* | -.14* | -.15* | -.24* | -.29* | .13* | .27* | .37* | .43* |
| 33. LE (Supervisor) | .02 | -.03 | .00 | -.04 | -.10 | -.18* | -.10* | -.14* | .12* | .28* | .16* | .17* |
| 34. LE (Peer) | .00 | .00 | -.02 | .03 | -.04 | -.12* | -.02 | -.08 | .12* | .13* | .22* | .15* |
| 35. LE (DR) | -.02 | -.05 | -.02 | .11* | -.03 | -.06 | -.02 | .05 | .09 | .13* | .15* | .23* |

*Note.* *N* = 732. *p* < .05.  DR = Direct Report.  LE = Leader Effectiveness.

Table 4 (Continued)

*Means, Standard Deviations, and Correlations for All Study Variables (Continued)*

| Variables | 25. | 26. | 27. | 28. | 29. | 30. | 31. | 32. | 33. | 34. | 35. |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 25. Production (Self) | — | | | | | | | | | | |
| 26. Production (Supervisor) | .30* | — | | | | | | | | | |
| 27. Production (Peer) | .43* | .43* | — | | | | | | | | |
| 28. Production (DR) | .43* | .42* | .52* | — | | | | | | | |
| 29. Feedback (Self) | .20* | .15* | .27* | .25* | — | | | | | | |
| 30. Feedback (Supervisor) | .14* | .24* | .24* | .23* | .38* | — | | | | | |
| 31. Feedback (Peer) | .14* | .18* | .37* | .25* | .48* | .51* | — | | | | |
| 32. Feedback (DR) | .13* | .19* | .27* | .32* | .42* | .44* | .55* | — | | | |
| 33. LE (Supervisor) | .08 | .30* | .14* | .20* | .00 | .03 | -.04 | -.03 | — | | |
| 34. LE (Peer) | .12* | .18* | .21* | .13* | .00 | -.07 | -.04 | -.06 | .55* | — | |
| 35. LE (DR) | .14* | .16* | .14* | .10* | .04 | -.01 | -.05 | .03 | .45* | .48* | — |
| *Mean* | 45.90 | 55.99 | 52.44 | 53.44 | 51.72 | 55.44 | 52.02 | 52.30 | 5.55 | 5.35 | 5.46 |
| *Standard Deviation* | 29.01 | 26.87 | 20.58 | 20.70 | 29.37 | 28.30 | 22.19 | 21.18 | .75 | .57 | .62 |

*Note*. $N = 732$. *$p < .05$.  DR = Direct Report.  LE = Leader Effectiveness.

Table 5

*Regressions of Supervisor Ratings of LE on Self-Other Ratings of Strategic*

| Model and LEA Dimension | $b$ | $t$ | $R^2$ | $\Delta R^2$ |
|---|---|---|---|---|
| **Self / Supervisor Ratings of *Strategic*** | | | | |
| Algebraic Difference | | | | |
|     Self rating | -.094 | -1.17 | .201* | |
|     Supervisor rating | 1.11* | 13.14* | | |
| Squared Difference | | | | |
|     Self rating | -.115 | -1.29 | .203* | .002 |
|     Supervisor rating | 1.08* | 12.03* | | |
|     Self x Self | -.001 | -.447 | | |
|     Supervisor x Supervisor | -.003 | -.914 | | |
|     Self x Supervisor | .000 | .014 | | |
| **Self / Peer Ratings of *Strategic*** | | | | |
| Algebraic Difference | | | | |
|     Self rating | .008 | .089 | .085* | |
|     Peer rating | .938* | 7.60* | | |
| Squared Difference | | | | |
|     Self rating | -.025 | -.262 | .086* | .001 |
|     Peer rating | .962* | 7.27* | | |
|     Self x Self | -.001 | -.404 | | |
|     Peer x Peer | .004 | .772 | | |
|     Self x Peer | -.003 | -.675 | | |
| **Self / Direct Report Ratings of *Strategic*** | | | | |
| Algebraic Difference | .075 | .085 | .060* | |
|     Self rating | .774* | 6.13* | | |
|     Direct report rating | | | | |
| Squared Difference | .026 | .261 | .065* | .005 |
|     Self rating | .719* | 5.24* | | |
|     Direct report rating | -.006 | -1.64 | | |
|     Self x Self | -.006 | -1.06 | | |
|     Direct report x Direct report | .006 | 1.14 | | |
|     Self x Direct report | | | | |

*Note.* LE = Leader Effectiveness. $b$ = unstandardized regression coefficients. *$p < 0.05$.

Table 6

*Regressions of Supervisor Ratings of LE on Self-Other Ratings of Innovative*

| Model and LEA Dimension | *b* | *t* | *R²* | *Δ R²* |
|---|---|---|---|---|
| <u>Self / Supervisor Ratings of</u> <u>*Innovative*</u> | | | | |
| Algebraic Difference | -.369* | -4.20* | .071* | |
| Self rating | .594* | 7.21* | | |
| Supervisor rating | | | | |
| Squared Difference | -.390* | -4.27* | .072* | .001 |
| Self rating | .559* | 7.13* | | |
| Supervisor rating | -.002 | -.641 | | |
| Self x Self | .001 | .313 | | |
| Supervisor x Supervisor | .002 | .650 | | |
| Self x Supervisor | | | | |
| | | | | |
| <u>Self / Peer Ratings of *Innovative*</u> | | | | |
| Algebraic Difference | | | | |
| Self rating | -.244* | -2.67* | .014* | |
| Peer rating | .279* | 2.55* | | |
| Squared Difference | | | | |
| Self rating | -.218* | -2.35* | .030* | .016* |
| Peer rating | .186 | 1.64 | | |
| Self x Self | -.003 | -.803 | | |
| Peer x Peer | -.014* | -3.19* | | |
| Self x Peer | .010* | 2.28* | | |
| <u>Self / Direct Report Ratings of</u> <u>*Innovative*</u> | | | | |
| Algebraic Difference | | | | |
| Self rating | -.303* | -3.28* | .025* | |
| Direct report rating | .438* | 3.92* | | |
| Squared Difference | | | | |
| Self rating | -.295* | -3.10* | .026* | .001 |
| Direct report rating | .422* | 3.67* | | |
| Self x Self | -.001 | -.248 | | |
| Direct report x Direct report | -.004 | -.833 | | |
| Self x Direct report | .001 | .272 | | |

*Note.* LE = Leader Effectiveness. *b* = unstandardized
regression coefficients. *p* < 0.05.

Agreement between Self and Supervisor Ratings of Conceptual Behaviors

Hypothesis 1a stated that a significant relationship would exist between self-supervisor agreement (on ratings of *conceptual* behaviors) and leader effectiveness. The resulting unstandardized regression coefficients are reported in Table 5 (strategic behaviors) and Table 6 (innovative behaviors).

For self-supervisor agreement on *strategic* behaviors (Table 5), results indicated that the squared difference model did not account for more variance, $\Delta R^2 = .002$, $F(3, 726) = .433$, $p = .73$, beyond the simpler algebraic difference model. The simpler algebraic difference model accounted for a significant amount of variance, $R^2 = .201$, $F(2, 729) = 91.84$, $p < .01$. However, an inspection of the unstandardized coefficients revealed that only supervisor ratings of strategic behaviors ($b_2 = 1.11$, $p < .01$) were a significant predictor of leader effectiveness. Self-ratings of strategic behaviors ($b_1 = -.094$, $p = .24$) did not significantly predict leader effectiveness. In other words, the regression coefficients suggest that the higher the supervisor ratings of strategic behaviors (regardless of the level of self-ratings), the higher the ratings of leader effectiveness.

For self-supervisor agreement on *innovative* behaviors (Table 6), results indicated that the squared difference model did not account for more variance, $\Delta R^2 = .001$, $F(3, 726) = .283$, $p = .84$, beyond the simpler algebraic difference model. The simpler algebraic difference model accounted for a significant amount of variance, $R^2 = .071$, $F(2, 729) = 27.90$, $p < .01$. An inspection of the unstandardized coefficients revealed that both self ($b_1 = -$

.369, $p < .01$) and supervisor ratings of innovative behaviors ($b_2 = .594$, $p <$

.01) were significant predictors of leader effectiveness and their coefficients

were in the opposite direction.  This supports the fit of the algebraic difference

model.  In other words, the regression coefficients indicate a change in leader

effectiveness ratings as self-ratings and supervisor ratings of innovative

behaviors move in opposite directions.  Specifically, as self-ratings *decrease*

and supervisor ratings *increase*, ratings of leader effectiveness are higher.

Agreement between Self and Peer Ratings of Conceptual Behaviors

Hypothesis 1b stated that a significant relationship would exist

between self-peer agreement (on ratings of *conceptual* behaviors) and leader

effectiveness.  The resulting unstandardized regression coefficients are

reported in Table 5 (strategic behaviors) and Table 6 (innovative behaviors).

For self-peer agreement on *strategic* behaviors (Table 5), results

indicated that the squared difference model did not account for more variance,

$\Delta R^2 = .001$, $F(3, 726) = .403$, $p = .75$, beyond the simpler algebraic difference

model.  The simpler algebraic difference model accounted for a significant

amount of variance, $R^2 = .085$, $F(2, 729) = 33.71$, $p < .01$.  However, an

inspection of the unstandardized coefficients revealed that only peer ratings of

strategic behaviors ($b_2 = .938$, $p < .01$) were a significant predictor of leader

effectiveness.  Self-ratings ($b_1 = .008$, $p = .93$) of strategic behaviors did not

significantly predict leader effectiveness.  In other words, the regression

coefficients suggest that the higher the peer ratings of strategic behaviors

(regardless of the level of self-ratings), the higher the ratings of leader effectiveness.

For self-peer agreement on *innovative* behaviors (Table 6), results indicated that the squared difference model accounted for significantly more variance, $\Delta R^2 = .016$, $F(3, 726) = 3.96$, $p < .01$, beyond the simpler algebraic difference model, implying that self-peer agreement on innovative behaviors was significantly related to leader effectiveness. Following Edwards' (1994) methodology, a surface plot was created to examine the three-dimensional relationship, shown in Figure 1. Centered variables were used in this figure and in the analysis, as recommended by Edwards. The surface in Figure 1 shows a pattern of significant curvature, which is consistent with the significant interaction term ($b_4 = .010$, $p < .05$). To further analyze the complex relationship between agreement and leader effectiveness, the lines of perfect agreement ($y = x$) and disagreement ($y = -x$) were examined. Along the $y = x$ line, self and peer ratings on innovative behaviors are equivalent. The $y = x$ line has been isolated in Figure 2, where it suggests that as self and peer ratings (of innovative behaviors) approach zero, leader effectiveness is the highest. In other words, when self and peer ratings are in-agreement and *moderate*, leader effectiveness is maximized. Leader effectiveness decreases as both self and peer ratings of innovative behaviors become more extreme in either direction. When comparing in-agreement/high ratings to in-agreement/low ratings, there does not appear to be much of a difference in

terms of effectiveness; however, in-agreement/high ratings are related to slightly higher leader effectiveness compared to in-agreement/low ratings.

The y = -x line (shown in Figure 3) represents the points at which peer ratings are equal to self-ratings of the opposite sign. This line shows a concave surface which indicates that when leaders and their peers disagree, ratings of leader effectiveness decline. Further, the degree of decline for *overestimators* (i.e., higher self than peer ratings) is greater than for *underestimators* (i.e., lower self than peer ratings). Therefore, in general, self and peer ratings (of innovative behaviors) that were more aligned were related to higher leader effectiveness ratings compared to self and peer ratings that were not aligned. Further, when self and peer ratings were aligned, *moderate* levels of innovative behaviors resulted in the highest levels of leader effectiveness, which did not support Hypothesis 1b (i.e., which stated that in-agreement/high ratings would be related to the highest leader effectiveness). Less agreement was generally related to lower effectiveness ratings; and further, disagreement in terms of *overestimation* was found to be more detrimental to leader effectiveness compared to *underestimation*, which was consistent with Hypothesis 1b.

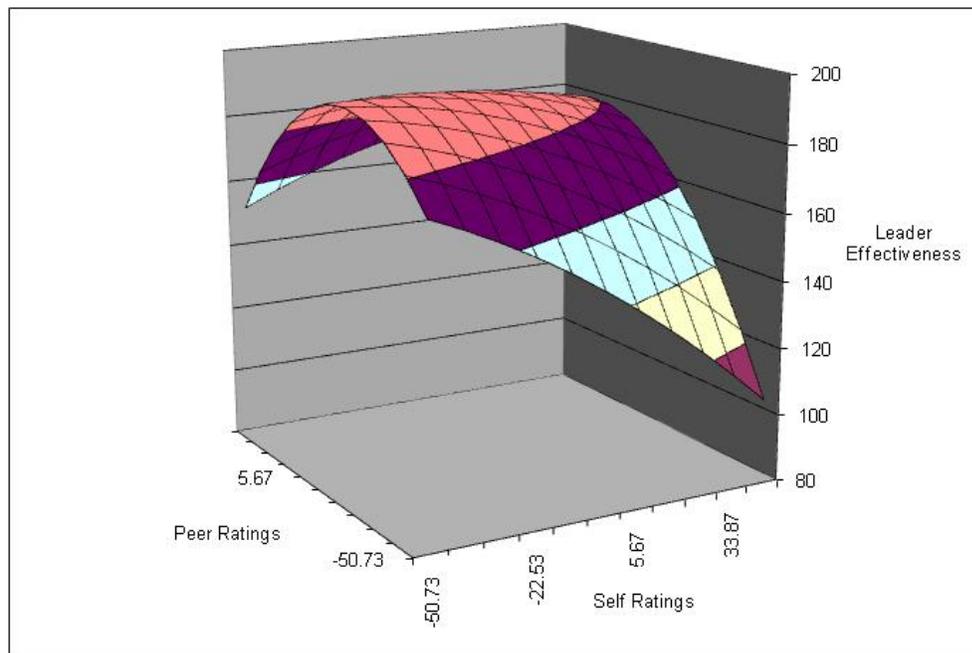Figure 1

*Leader Effectiveness Ratings Relative to Self and Peer Ratings of Innovative*

Figure 2

*Leader Effectiveness Ratings Where Self and Peer Ratings of Innovative are*
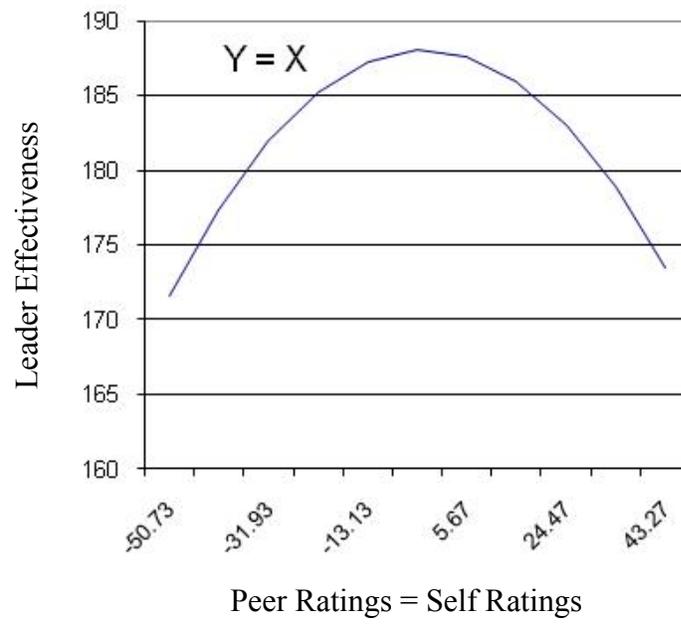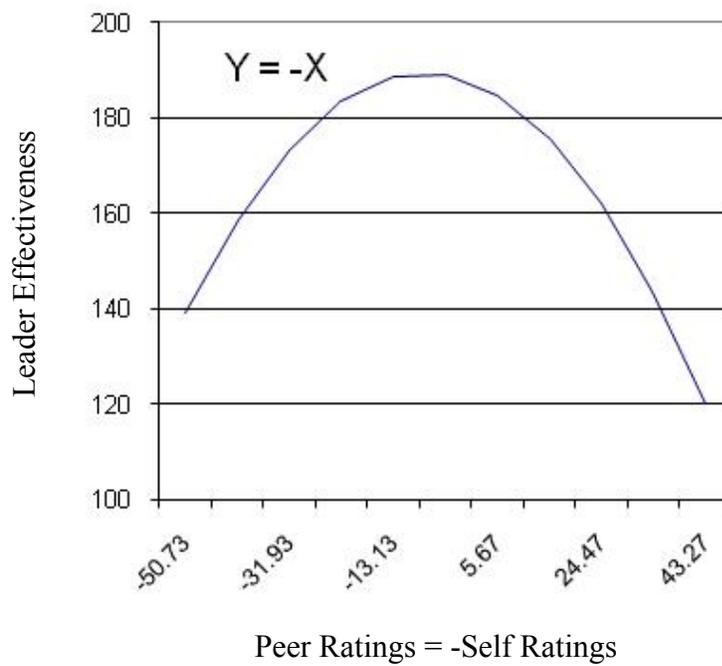
*Equivalent*



Peer Ratings = Self Ratings

Figure 3

*Leader Effectiveness Ratings Where Self and Peer Ratings of Innovative are*

*at Extreme Disagreement*



Peer Ratings = -Self Ratings

Agreement between Self and Direct Report Ratings of Conceptual Behaviors

Hypothesis 1c stated that a significant relationship would exist between self-direct report agreement (on ratings of *conceptual* behaviors) and leader effectiveness. The resulting unstandardized regression coefficients are reported in Table 5 (strategic) and Table 6 (innovative).

For self-direct report agreement on *strategic* behaviors (Table 5), results indicated that the squared difference model did not account for more variance, $\Delta R^2 = .005$, $F(3, 726) = 1.23$, $p = .30$, beyond the simpler algebraic difference model. The simpler algebraic difference model accounted for a significant amount of variance, $R^2 = .060$, $F(2, 729) = 23.47$, $p < .01$. However, an inspection of the unstandardized coefficients revealed that only direct report ratings of strategic behaviors ($b_2 = .774$, $p < .01$) were a significant predictor of leader effectiveness. Self-ratings of strategic behaviors ($b_1 = .075$, $p = .40$) did not significantly predict leader effectiveness. In other words, the regression coefficients suggest that the higher the direct report ratings of strategic behaviors (regardless of the level of self-ratings), the higher the ratings of leader effectiveness.

For self-direct report agreement on *innovative* behaviors (Table 6), results indicated that the squared difference model did not account for more variance, $\Delta R^2 = .001$, $F(3, 726) = .261$, $p = .85$, beyond the simpler algebraic difference model. The simpler algebraic difference model accounted for a significant amount of variance, $R^2 = .025$, $F(2, 729) = 9.50$, $p < .01$. An inspection of the unstandardized coefficients revealed that both self ($b_1 = -$

.303, $p < .01$) and direct report ratings of innovative behaviors ($b_2 = .438$, $p <$ .01) were significant predictors of leader effectiveness and their coefficients were in the opposite direction. This supports the fit of the algebraic difference model. In other words, the regression coefficients indicate a change in leader effectiveness ratings as self-ratings and direct report ratings of innovative behaviors move in opposite directions. Specifically, as self-ratings *decrease* and direct report ratings *increase*, ratings of leader effectiveness are higher.

Differences across Rater Sources on Conceptual Behaviors

Hypothesis 1d sought to examine the differences (on ratings of *conceptual* behavior) across rater sources, stating that the relationship between self-supervisor agreement and leader effectiveness would be significantly stronger compared to self-peer and self-direct report agreement. However, because the squared difference model (used to test agreement between rater sources) was not supported for all of the self-other group comparisons, this hypothesis could not be tested.

Table 7

*Regressions of Supervisor Ratings of LE on Self-Other Ratings of Cooperative*

| Model and LEA Dimension | b | t | $R^2$ | $\Delta R^2$ |
|---|---|---|---|---|
| <u>Self / Supervisor Ratings of</u> <br> *Cooperative* | | | | |
| Algebraic Difference | -.065 | -.767 | .034* | |
|   Self rating | -.421* | -4.50* | | |
|   Supervisor rating | | | | |
| Squared Difference | -.099 | -1.19 | .100* | .066* |
|   Self rating | -.511* | -5.58* | | |
|   Supervisor rating | -.001 | -.286 | | |
|   Self x Self | -.025* | -7.00* | | |
|   Supervisor x Supervisor | .000 | -.053 | | |
|   Self x Supervisor | | | | |
| <u>Self / Peer Ratings of *Cooperative*</u> | | | | |
| Algebraic Difference | | | | |
|   Self rating | -.108 | -1.23 | .014* | |
|   Peer rating | -.286* | -2.28* | | |
| Squared Difference | | | | |
|   Self rating | -.110 | -1.23 | .018* | .004 |
|   Peer rating | -.227* | -2.20* | | |
|   Self x Self | -.002 | -.480 | | |
|   Peer x Peer | -.009 | -1.63 | | |
|   Self x Peer | .001 | .247 | | |
| <u>Self / Direct Report Ratings of</u> <br> *Cooperative* | | | | |
| Algebraic Difference | | | | |
|   Self rating | -.080 | -.919 | .020* | |
|   Direct report rating | -.386* | -3.10* | | |
| Squared Difference | | | | |
|   Self rating | -.090 | -1.01 | .021* | .001 |
|   Direct report rating | -.390* | -3.07* | | |
|   Self x Self | -.003 | -.929 | | |
|   Direct report x Direct report | .002 | .320 | | |
|   Self x Direct report | .002 | .352 | | |

*Note.* LE = Leader Effectiveness. *b* = unstandardized regression coefficients.

*p < 0.05.

Table 8

*Regressions of Supervisor Ratings of LE on Self-Other Ratings of Consensual*

| Model and LEA Dimension | $b$ | $t$ | $R^2$ | $\Delta R^2$ |
|---|---|---|---|---|
| Self / Supervisor Ratings of *Consensual* | | | | |
| Algebraic Difference | | | | |
| Self rating | .038 | .438 | .002 | |
| Supervisor rating | -.122 | -1.27 | | |
| Squared Difference | | | | |
| Self rating | .056 | .644 | .018* | .016* |
| Supervisor rating | -.208* | -2.11* | | |
| Self x Self | .002 | .640 | | |
| Supervisor x Supervisor | -.013* | -3.41* | | |
| Self x Supervisor | .003 | .927 | | |
| Self / Peer Ratings of *Consensual* | | | | |
| Algebraic Difference | | | | |
| Self rating | .024 | .277 | .000 | |
| Peer rating | -.059 | -.450 | | |
| Squared Difference | | | | |
| Self rating | .009 | .101 | .012 | .012* |
| Peer rating | -.113 | -.851 | | |
| Self x Self | .002 | .599 | | |
| Peer x Peer | -.016* | -2.46* | | |
| Self x Peer | -.002 | -.492 | | |
| Self / Direct Report Ratings of *Consensual* | | | | |
| Algebraic Difference | | | | |
| Self rating | .065 | .723 | .004 | |
| Direct report rating | -.208 | -1.60 | | |
| Squared Difference | | | | |
| Self rating | .074 | .808 | .006 | .002 |
| Direct report rating | -.238 | -1.79 | | |
| Self x Self | .001 | .287 | | |
| Direct report x Direct report | -.006 | -1.10 | | |
| Self x Direct report | .004 | .779 | | |

*Note.* LE = Leader Effectiveness. $b$ = unstandardized regression coefficients.

*p < 0.05.

Table 9

*Regressions of Supervisor Ratings of LE on Self-Other Ratings of Empathy*

| Model and LEA Dimension | $b$ | $t$ | $R^2$ | $\Delta R^2$ |
|---|---|---|---|---|
| **Self / Supervisor Ratings of *Empathy*** | | | | |
| Algebraic Difference | | | | |
| Self rating | -.048 | -.526 | .036* | |
| Supervisor rating | -.453* | -4.41* | | |
| Squared Difference | | | | |
| Self rating | .011 | .116 | .062* | .026* |
| Supervisor rating | -.499* | -4.79* | | |
| Self x Self | .003 | .920 | | |
| Supervisor x Supervisor | -.017* | -4.16* | | |
| Self x Supervisor | .002 | .432 | | |
| | | | | |
| **Self / Peer Ratings of *Empathy*** | | | | |
| Algebraic Difference | | | | |
| Self rating | -.138 | -1.49 | .016* | |
| Peer rating | -.268* | -2.03* | | |
| Squared Difference | | | | |
| Self rating | -.132 | -1.40 | .019* | .003 |
| Peer rating | -.271* | -2.00* | | |
| Self x Self | .004 | 1.04 | | |
| Peer x Peer | -.004 | -.646 | | |
| Self x Peer | -.004 | -.668 | | |
| | | | | |
| **Self / Direct Report Ratings of *Empathy*** | | | | |
| Algebraic Difference | | | | |
| Self rating | -.082 | -.877 | .023* | |
| Direct report rating | -.411* | -3.07* | | |
| Squared Difference | | | | |
| Self rating | -.072 | -.752 | .025* | .002 |
| Direct report rating | -.399* | -2.94* | | |
| Self x Self | .003 | .795 | | |
| Direct report x Direct report | -.005 | -.776 | | |
| Self x Direct report | -.001 | -.126 | | |

*Note.* LE = Leader Effectiveness. $b$ = unstandardized regression coefficients.

*p* < 0.05.

Agreement between Self and Supervisor Ratings of Collaborative Behaviors

Hypothesis 2a stated that a significant relationship would exist between self-supervisor agreement (on ratings of *collaborative* behaviors) and leader effectiveness. The resulting unstandardized regression coefficients are reported in Table 7 (cooperative), Table 8 (consensual), and Table 9 (empathy).

For self-supervisor agreement on *cooperative* behaviors (Table 7), results indicated that the squared difference model accounted for significantly more variance, $\Delta R^2 = .066$, $F(3, 726) = 17.96$, $p < .01$, beyond the simpler algebraic difference model. Following Edwards' (1994) methodology, a surface plot was created to examine the three-dimensional relationship, shown in Figure 4. The surface in Figure 4 shows a curvilinear relationship between supervisor ratings of cooperative behaviors and leader effectiveness, which is consistent with the significant squared term for supervisor ratings ($b_5 = -.025$, $p < .01$), but a non-significant interaction term ($b_4 = .000$, $p = .96$). The plot displays a concave surface such that leader effectiveness is highest when supervisor ratings of cooperative behaviors are moderate, regardless of the level of self-ratings. Leader effectiveness decreases as supervisor ratings of cooperative behaviors either increase or decrease, and the rate of decline is greatest when supervisor ratings of cooperative behaviors are *very high*. In other words, leader effectiveness decelerates at a faster pace for high ratings of cooperative behaviors from supervisors. This finding does not support Hypothesis 2a because neither self-ratings nor the agreement between self and

supervisor ratings (of cooperative behaviors) were significantly related to

leader effectiveness.

Figure 4

*Leader Effectiveness Ratings Relative to Self and Supervisor Ratings of*

*Cooperative*
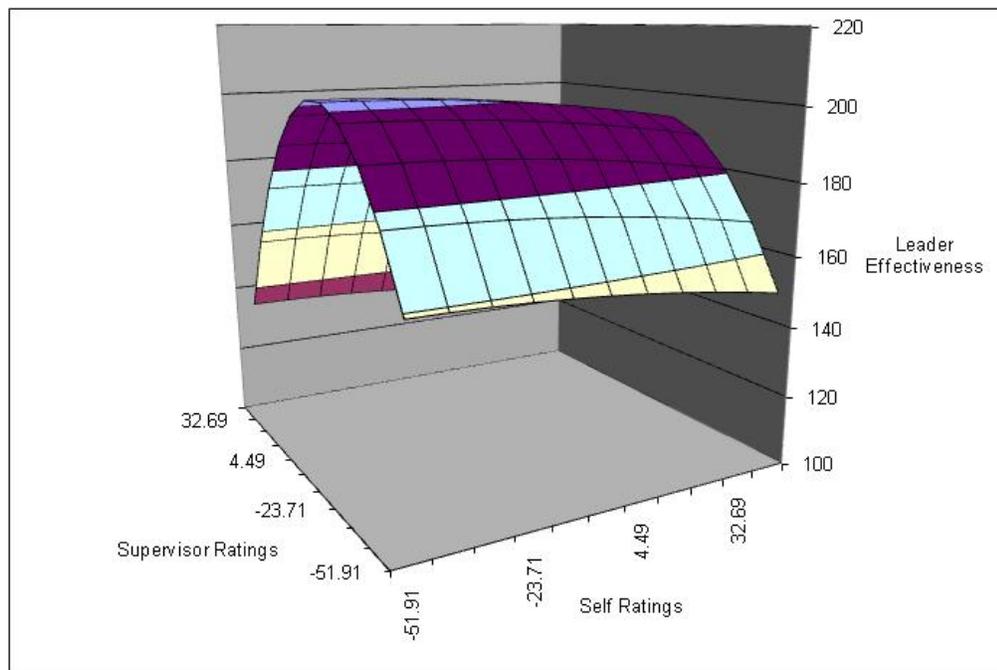
For self-supervisor agreement on *consensual* behaviors (Table 8),

results indicated that the squared difference model accounted for significantly

more variance, $\Delta R^2 = .016$, $F (3, 726) = 4.00$, $p < .01$, beyond the simpler

algebraic difference model.  Following Edwards' (1994) methodology, a

surface plot was created to examine the three-dimensional relationship, shown

in Figure 5.  Similar to Figure 4, the surface in Figure 5 shows a curvilinear

relationship between supervisor ratings of consensual behaviors and leader

effectiveness, which is consistent with the significant squared term for

supervisor ratings ($b_5 = -.013$, $p < .01$), but a non-significant interaction term

($b_4 = .003$, $p = .35$).  The plot displays a concave surface such that leader

effectiveness is highest when supervisor ratings (of consensual behaviors) are

*moderate*.  Leader effectiveness decreases as supervisor ratings of consensual

behaviors either increase or decrease, and the rate of decline is greatest when

supervisor ratings of consensual behaviors are high.  Additionally, Figure 5

shows a slight saddle-shaped curve, indicating that leader effectiveness is

*slightly* higher when self-ratings are either high or low (and therefore, leader

effectiveness is slightly lower when self-ratings are moderate).  Overall, these

findings do not support Hypothesis 2a because neither self-ratings nor the

agreement between self and supervisor ratings (of consensual behaviors) were

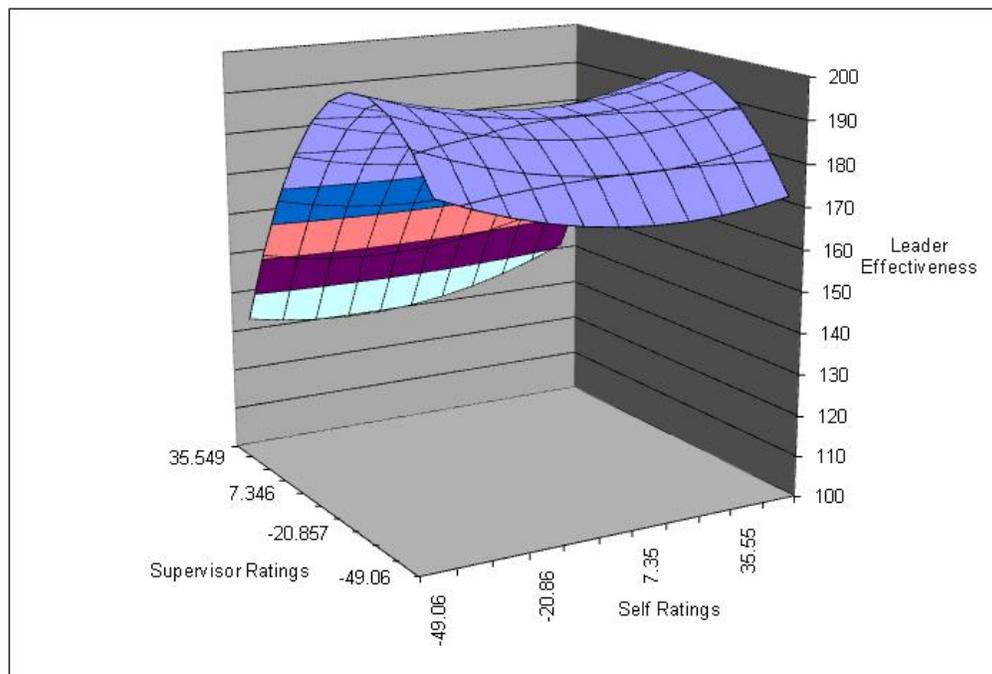significantly related to leader effectiveness.

Figure 5

*Leader Effectiveness Ratings Relative to Self and Supervisor Ratings of*

*Consensual*
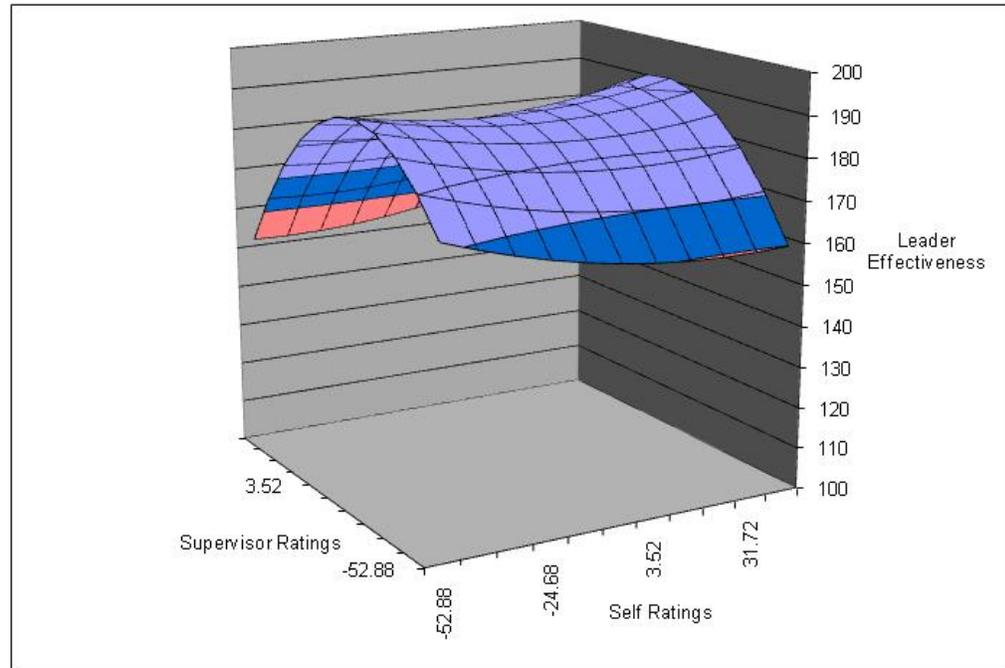
For self-supervisor agreement on *empathy* behaviors (Table 9), results indicated that the squared difference model accounted for significantly more variance, $\Delta R^2 = .026$, $F (3, 726) = 6.75$, $p < .01$, beyond the simpler algebraic difference model. Following Edwards' (1994) methodology, a surface plot was created as shown in Figure 6. Similar to the surfaces in Figures 4 and 5, Figure 6 shows a curvilinear relationship between supervisor ratings of empathy behaviors and leader effectiveness, which is consistent with the significant squared term for supervisor ratings ($b_5 = -.017$, $p < .01$), but a non-significant interaction term ($b_4 = .002$, $p = .67$). The plot displays a concave surface such that leader effectiveness is highest when supervisor ratings (of empathy behaviors) are *moderate*. Leader effectiveness decreases as supervisor ratings of empathy behaviors either increase or decrease, and the rate of deceleration is *much faster* when supervisor ratings of empathy behaviors are high. In other words, leaders are viewed as less effective when their supervisors rate them as highly empathetic. Lastly, Figure 6 also shows a slight saddle-shaped curve, which indicates that leader effectiveness is *slightly* higher when self-ratings are either high or low (and likewise, leader effectiveness is slightly lower when self-ratings are moderate). Overall, these findings do not support Hypothesis 2a because neither self-ratings nor the agreement between self and supervisor ratings (of empathy behaviors) were significantly related to leader effectiveness.

Figure 6

*Leader Effectiveness Ratings Relative to Self and Supervisor Ratings of*

*Empathy*

Agreement between Self and Peer Ratings of Collaborative Behaviors

Hypothesis 2b stated that a significant relationship would exist between self-peer agreement (on ratings of *collaborative* behaviors) and leader effectiveness. The resulting unstandardized regression coefficients are reported in Table 7 (cooperative), Table 8 (consensual), and Table 9 (empathy).
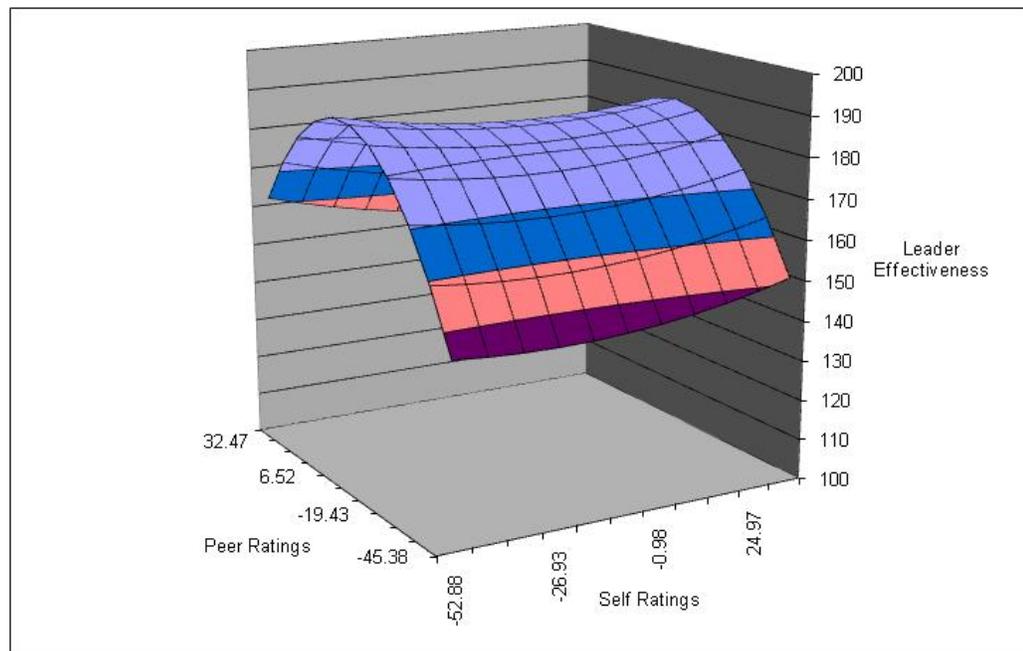
For self-peer agreement on *cooperative* behaviors (Table 7), results indicated that the squared difference model did not account for more variance, $\Delta R^2 = .004$, $F(3, 726) = 1.10$, $p = .35$, beyond the simpler algebraic difference model. The simpler algebraic difference model accounted for a significant amount of variance, $R^2 = .014$, $F(2, 729) = 5.08$, $p < .01$. However, an inspection of the unstandardized coefficients revealed that only peer ratings of cooperative behaviors ($b_2 = -.286$, $p < .05$) were significantly related to leader effectiveness. Self-ratings of cooperative behaviors ($b_1 = -.108$, $p = .22$) were not significantly related to leader effectiveness. In other words, the regression coefficients suggest that the higher the peer ratings of cooperative behaviors (regardless of the level of self-ratings), the *lower* the ratings of leader effectiveness.

For self-peer agreement on *consensual* behaviors (Table 8), results indicated that the squared difference model accounted for significantly more variance, $\Delta R^2 = .012$, $F(3, 726) = 2.79$, $p < .05$, beyond the simpler algebraic difference model. Following Edwards' (1994) methodology, a surface plot was created. Figure 7 shows a curvilinear relationship between peer ratings of

consensual behaviors and leader effectiveness, which is consistent with the significant squared term for peer ratings ($b_5 = -.016$, $p < .05$), but a non-significant interaction term ($b_4 = -.002$, $p = .62$). Figure 7 displays a concave surface such that leader effectiveness is highest when peer ratings (of consensual behaviors) are *moderate*. Leader effectiveness decreases as peer ratings either increase or decrease, and the rate of deceleration is more severe when peer ratings of consensual behaviors are *low*. In other words, leaders are rated as less effective when their peers do not view them as consensual. Lastly, Figure 7 also shows a slight saddle-shaped curve, which indicates that leader effectiveness is *slightly* higher when self-ratings are either high or low (and likewise, leader effectiveness is slightly lower when self-ratings are moderate). Overall, these findings do not support Hypothesis 2b because neither self-ratings nor the agreement between self and peer ratings (of consensual behaviors) were significantly related to leader effectiveness.

Figure 7

*Leader Effectiveness Ratings Relative to Self and Peer Ratings of Consensual*

For self-peer agreement on *empathy* behaviors (Table 9), results indicated that the squared difference model did not account for more variance, $\Delta R^2 = .003$, $F(3, 726) = 0.81$, $p = .49$, beyond the simpler algebraic difference model. The simpler algebraic difference model accounted for a significant amount of variance, $R^2 = .016$, $F(2, 729) = 5.89$, $p < .01$. However, an inspection of the unstandardized coefficients revealed that only peer ratings of empathy behaviors ($b_2 = -.268$, $p < .05$) were significantly related to leader effectiveness. Self-ratings of empathy behaviors ($b_1 = -.138$, $p = .14$) were not significantly related to leader effectiveness. In other words, the regression coefficients suggest that the higher the peer ratings of empathy behaviors (regardless of the level of self-ratings), the *lower* the ratings of leader effectiveness.

Agreement between Self and Direct Report Ratings of Collaborative Behaviors

Hypothesis 2c stated that a significant relationship would exist between self-direct report agreement (on ratings of *collaborative* behaviors) and leader effectiveness. The resulting unstandardized regression coefficients are reported in Table 7 (cooperative), Table 8 (consensual), and Table 9 (empathy).

For self-direct report agreement on *cooperative* behaviors (Table 7), results indicated that the squared difference model did not account for more variance, $\Delta R^2 = .001$, $F(3, 726) = 0.35$, $p = .79$, beyond the simpler algebraic difference model. The simpler algebraic difference model accounted for a

significant amount of variance, $R^2 = .020$, $F(2, 729) = 7.28$, $p < .01$.

However, an inspection of the unstandardized coefficients revealed that only

direct report ratings of cooperative behaviors ($b_2 = -.386$, $p < .01$) were a

significant predictor of leader effectiveness. Self-ratings of cooperative

behaviors ($b_1 = -.080$, $p = .36$) did not significantly predict leader

effectiveness. In other words, the regression coefficients suggest that the

higher the direct report ratings of cooperative behaviors (regardless of the

level of self-ratings), the *lower* the ratings of leader effectiveness.

For self-direct report agreement on *consensual* behaviors (Table 8),

results indicated that the squared difference model did not account for more

variance, $\Delta R^2 = .002$, $F(3, 726) = 0.55$, $p = .65$, beyond the simpler algebraic

difference model. The simpler algebraic difference model was also non-

significant, $R^2 = .004$, $F(2, 729) = 1.29$, $p = .28$. Thus, the data did not

support the squared difference model or the simpler algebraic model.

For self-direct report agreement on *empathy* behaviors (Table 9),

results indicated that the squared difference model did not account for more

variance, $\Delta R^2 = .002$, $F(3, 726) = 0.48$, $p = .70$, beyond the simpler algebraic

difference model. The simpler algebraic difference model accounted for a

significant amount of variance, $R^2 = .023$, $F(2, 729) = 8.59$, $p < .01$.

However, an inspection of the unstandardized coefficients revealed that only

direct report ratings of empathy behaviors ($b_2 = -.411$, $p < .01$) were a

significant predictor of leader effectiveness. Self-ratings of empathy

behaviors ($b_1 = -.082$, $p = .38$) did not significantly predict leader

effectiveness. In other words, the regression coefficients suggest that the higher the direct report ratings of empathy behaviors (regardless of the level of self-ratings), the *lower* the ratings of leader effectiveness.

Differences across Rater Sources on Collaborative Behaviors

Hypothesis 2d sought to examine the differences (on ratings of *collaborative* behavior) across rater sources, stating that the relationship between self-peer agreement and leader effectiveness would be significantly stronger compared to self-supervisor and self-direct report agreement. However, because the squared difference model (used to test agreement between rater sources) was not supported for all of the self-other group comparisons, this hypothesis could not be tested.

Table 10

*Regressions of Supervisor Ratings of LE on Self-Other Ratings of MF*

| Model and LEA Dimension | $b$ | $t$ | $R^2$ | $\Delta R^2$ |
|---|---|---|---|---|
| **Self / Supervisor Ratings of *MF*** | | | | |
| Algebraic Difference | | | | |
| Self rating | .113 | 1.40 | .088* | |
| Supervisor rating | .693* | 7.63* | | |
| Squared Difference | | | | |
| Self rating | .095 | 1.16 | .101* | .013* |
| Supervisor rating | .681* | 7.50* | | |
| Self x Self | .001 | .182 | | |
| Supervisor x Supervisor | -.010* | -2.77* | | |
| Self x Supervisor | .007* | 2.22* | | |
| **Self / Peer Ratings of *MF*** | | | | |
| Algebraic Difference | | | | |
| Self rating | .155 | 1.81 | .036* | |
| Peer rating | .464* | 3.91* | | |
| Squared Difference | | | | |
| Self rating | .175* | 2.00* | .040* | .004 |
| Peer rating | .465* | 3.78* | | |
| Self x Self | .000 | .073 | | |
| Peer x Peer | -.004 | -.727 | | |
| Self x Peer | .007 | 1.64 | | |
| **Self / Direct Report Ratings of *MF*** | | | | |
| Algebraic Difference | | | | |
| Self rating | .154 | 1.81 | .039* | |
| Direct report rating | .491* | 4.22* | | |
| Squared Difference | | | | |
| Self rating | .180* | 2.05* | .044* | .005 |
| Direct report rating | .516* | 4.20* | | |
| Self x Self | -.001 | -.171 | | |
| Direct report x Direct report | .004 | .727 | | |
| Self x Direct report | .006 | 1.40 | | |

*Note.* LE = Leader Effectiveness. $b$ = unstandardized regression coefficients.

*$p < 0.05$. MF = Management Focus.

Table 11

*Regressions of Supervisor Ratings of LE on Self-Other Ratings of Production*

| Model and LEA Dimension | $b$ | $t$ | $R^2$ | $\Delta R^2$ |
|---|---|---|---|---|
| Self / Supervisor Ratings of *Production* | | | | |
| Algebraic Difference | | | | |
| Self rating | -.004 | -.055 | .097* | |
| Supervisor rating | .746* | 8.44* | | |
| Squared Difference | | | | |
| Self rating | .014 | .162 | .102* | .005 |
| Supervisor rating | .769* | 8.59* | | |
| Self x Self | .003 | .882 | | |
| Supervisor x Supervisor | -.007 | -1.87 | | |
| Self x Supervisor | .001 | .356 | | |
| | | | | |
| Self / Peer Ratings of *Production* | | | | |
| Algebraic Difference | | | | |
| Self rating | .072 | .803 | .024* | |
| Peer rating | .435* | 3.45* | | |
| Squared Difference | | | | |
| Self rating | .084 | .924 | .026* | .002 |
| Peer rating | .433* | 3.41* | | |
| Self x Self | .003 | .802 | | |
| Peer x Peer | .000 | -.088 | | |
| Self x Peer | -.004 | -.736 | | |
| | | | | |
| Self / Direct Report Ratings of *Production* | | | | |
| Algebraic Difference | | | | |
| Self rating | -.001 | -.006 | .046* | |
| Direct report rating | .665* | 5.35* | | |
| Squared Difference | | | | |
| Self rating | .014 | .159 | .049* | .003 |
| Direct report rating | .679* | 5.39* | | |
| Self x Self | .003 | .737 | | |
| Direct report x Direct report | -.007 | -1.35 | | |
| Self x Direct report | .000 | .030 | | |

*Note.* LE = Leader Effectiveness. $b$ = unstandardized regression coefficients.

*p < 0.05.

Table 12

*Regressions of Supervisor Ratings of LE on Self-Other Ratings of Feedback*

| Model and LEA Dimension | $b$ | $t$ | $R^2$ | $\Delta R^2$ |
|---|---|---|---|---|
| **Self / Supervisor Ratings of *Feedback*** | | | | |
| Algebraic Difference | | | | |
| Self rating | -.027 | -.304 | .001 | |
| Supervisor rating | .089 | .981 | | |
| Squared Difference | | | | |
| Self rating | -.038 | -.432 | .009 | .008 |
| Supervisor rating | .111 | 1.21 | | |
| Self x Self | .003 | .928 | | |
| Supervisor x Supervisor | -.005 | -1.29 | | |
| Self x Supervisor | .005 | 1.59 | | |
| **Self / Peer Ratings of *Feedback*** | | | | |
| Algebraic Difference | | | | |
| Self rating | .047 | .506 | .001 | |
| Peer rating | -.113 | -.924 | | |
| Squared Difference | | | | |
| Self rating | .052 | .566 | .005 | .004 |
| Peer rating | -.119 | -.977 | | |
| Self x Self | .005 | 1.48 | | |
| Peer x Peer | .002 | .305 | | |
| Self x Peer | -.001 | -.282 | | |
| **Self / Direct Report Ratings of *Feedback*** | | | | |
| Algebraic Difference | | | | |
| Self rating | .033 | .367 | .001 | |
| Direct report rating | -.088 | -.712 | | |
| Squared Difference | | | | |
| Self rating | .037 | .410 | .006 | .005 |
| Direct report rating | -.086 | -.700 | | |
| Self x Self | .004 | 1.03 | | |
| Direct report x Direct report | .002 | .447 | | |
| Self x Direct report | .003 | .688 | | |

*Note.* LE = Leader Effectiveness. $b$ = unstandardized regression coefficients.

*$p < 0.05$.

Agreement between Self and Supervisor Ratings of Managing Others

Behaviors

Hypothesis 3a stated that a significant relationship would exist

between self-supervisor agreement (on ratings of behaviors related to

*managing others*) and leader effectiveness. The resulting unstandardized

regression coefficients are reported in Table 10 (management focus), Table 11

(production), and Table 12 (feedback).

For self-supervisor agreement on ratings of *management focus*

behaviors (Table 10), results indicated that the squared difference model

accounted for significantly more variance, $\Delta R^2 = .013$, $F(3, 726) = 3.49$, $p <$

.05, beyond the simpler algebraic difference model. Following Edwards'

(1994) recommendations, a three-dimensional surface plot was created, as

shown in Figure 8. The surface in Figure 8 shows a pattern of significant

curvature, which is consistent with the significant interaction term ($b_4 = .007$,

$p < .05$). In order to analyze the complex relationship between agreement and

leader effectiveness, the lines of perfect agreement (y = x) and disagreement

(y = -x) were examined. Along the y = x line, self and supervisor ratings (on

management focus behaviors) are equivalent. The y = x line has been isolated

in Figure 9, where it suggests that as self and supervisor ratings (of

management focus behaviors) simultaneously increase, leader effectiveness

also increases. In other words, when self and supervisor ratings are in-

agreement and *high*, leader effectiveness is maximized.

The y = -x line (shown in Figure 10) represents the points at which supervisor ratings are equal to self-ratings of the opposite sign.  This line shows a concave surface which indicates that when leaders and their supervisor disagree, ratings of leader effectiveness decline.  Further, as seen in Figure 10, the degree of decline for *overestimators* (i.e., higher self than supervisor ratings) was much greater than for *underestimators* (i.e., lower self than supervisor ratings).  Therefore, in general, self and supervisor ratings (of management focus behaviors) that were more aligned were related to higher leader effectiveness ratings compared to self and supervisor ratings that were not aligned.  Further, when self and supervisor ratings were aligned, *high* levels of management focus behaviors resulted in the highest levels of leader effectiveness, which was consistent with Hypothesis 3a.  Less agreement was generally related to lower leader effectiveness; and further, disagreement in terms of *overestimation* was found to be more detrimental to leader effectiveness compared to *underestimation*, which was also consistent with Hypothesis 3a.

Figure 8

*Leader Effectiveness Ratings Relative to Self and Supervisor Ratings of*
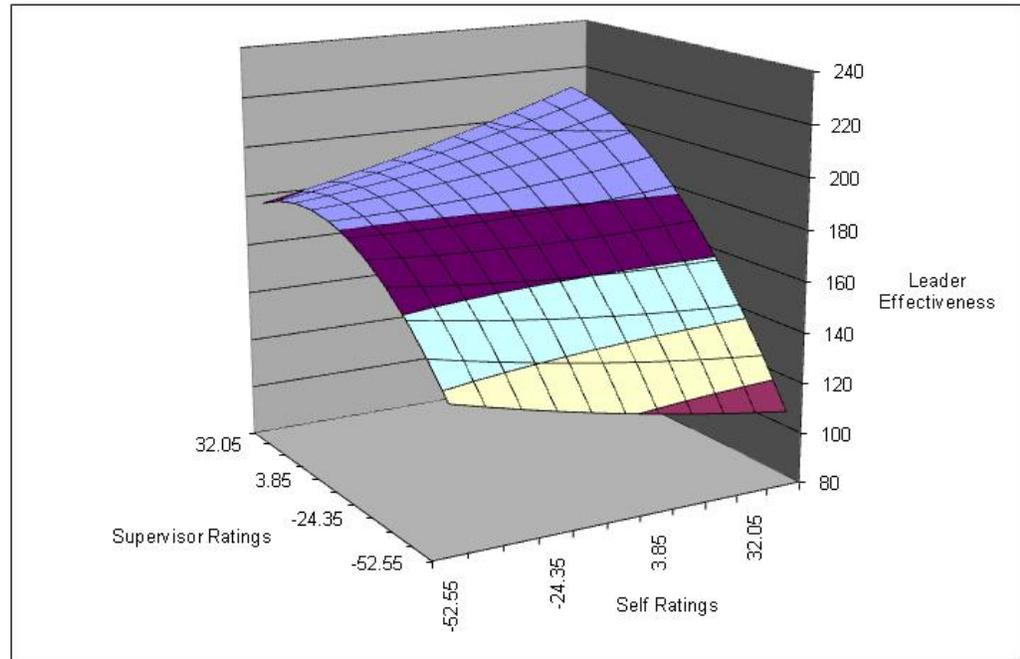
*Management Focus*

Figure 9

*Leader Effectiveness Ratings Where Self and Supervisor Ratings of*
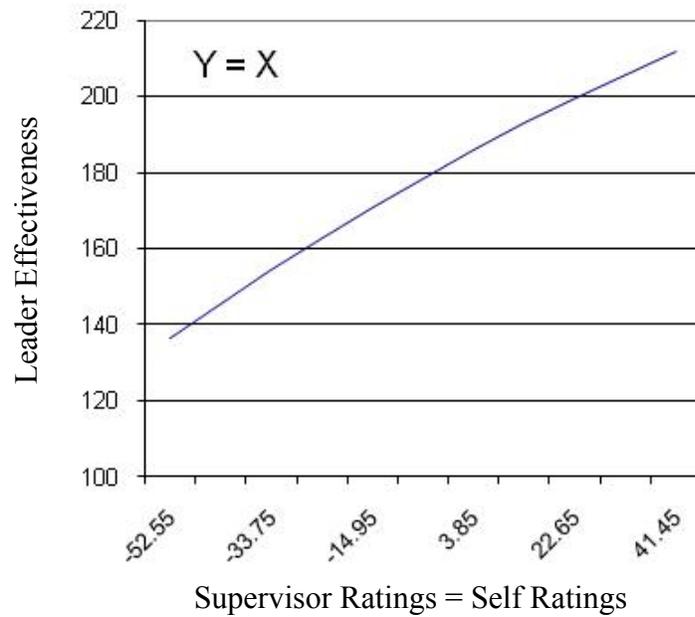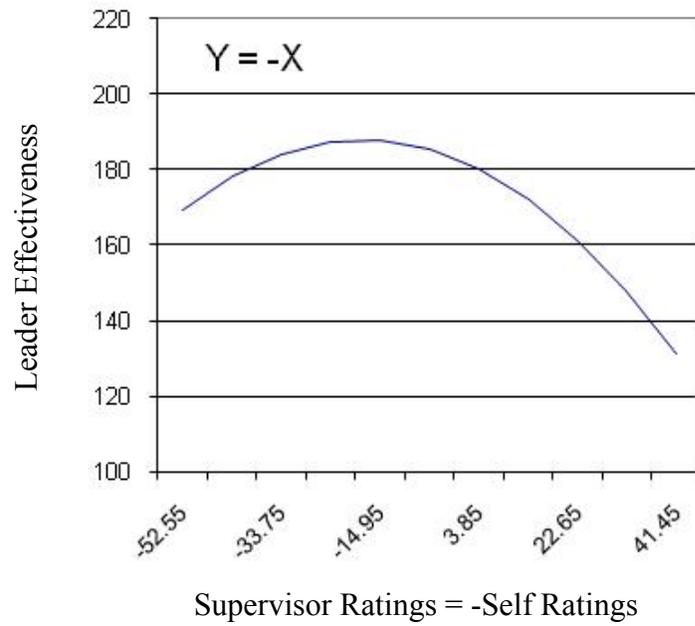
*Management Focus are Equivalent*

Figure 10

*Leader Effectiveness Ratings Where Self and Supervisor Ratings of*

*Management Focus are at Extreme Disagreement*



Supervisor Ratings = -Self Ratings

For self-supervisor agreement on ratings of *production* behaviors

(Table 11), results indicated that the squared difference model did not account

for more variance, $\Delta R^2 = .005$, $F(3, 726) = 1.45$, $p = .23$, beyond the simpler

algebraic difference model. The simpler algebraic difference model

accounted for a significant amount of variance, $R^2 = .097$, $F(2, 729) = 39.04$,

$p < .01$. However, an inspection of the unstandardized coefficients revealed

that only supervisor ratings of production behaviors ($b_2 = .746$, $p < .01$) were a

significant predictor of leader effectiveness. Self-ratings of production

behaviors ($b_1 = -.004$, $p = .96$) did not significantly predict leader

effectiveness. In other words, the regression coefficients suggest that the

higher the supervisor ratings of production behaviors (regardless of the level

of self-ratings), the higher the ratings of leader effectiveness.

For self-supervisor agreement on ratings of *feedback* behaviors (Table

12), results indicated that the squared difference model did not account for

more variance, $\Delta R^2 = .008$, $F(3, 726) = 1.79$, $p = .15$, beyond the simpler

algebraic difference model. The simpler algebraic difference model was also

non-significant, $R^2 = .001$, $F(2, 729) = 0.48$, $p = .62$. Thus, the data did not

support the squared difference model or the simpler algebraic model.

Agreement between Self and Peer Ratings of Managing Others Behaviors

Hypothesis 3b stated that a significant relationship would exist

between self-peer agreement (on ratings of behaviors related to *managing*

*others*) and leader effectiveness. The resulting unstandardized regression

coefficients are reported in Table 10 (management focus), Table 11 (production), and Table 12 (feedback).

For self-peer agreement on *management focus* behaviors (Table 10), results indicated that the squared difference model did not account for more variance, $\Delta R^2 = .004$, $F(3, 726) = 1.01$, $p = .39$, beyond the simpler algebraic difference model. The simpler algebraic difference model accounted for a significant amount of variance, $R^2 = .036$, $F(2, 729) = 13.54$, $p < .01$. However, an inspection of the unstandardized coefficients revealed that only peer ratings on management focus behaviors ($b_2 = .464$, $p < .01$) were a significant predictor of leader effectiveness. Self-ratings on management focus behaviors ($b_1 = .155$, $p = .07$) did not significantly predict leader effectiveness. In other words, the regression coefficients suggest that the higher the peer ratings of management focus behaviors (regardless of the level of self-ratings), the higher the ratings of leader effectiveness.

For self-peer agreement on *production* behaviors (Table 11), results indicated that the squared difference model did not account for more variance, $\Delta R^2 = .002$, $F(3, 726) = 0.35$, $p = .79$, beyond the simpler algebraic difference model. The simpler algebraic difference model accounted for a significant amount of variance, $R^2 = .024$, $F(2, 729) = 9.07$, $p < .01$. However, an inspection of the unstandardized coefficients revealed that only peer ratings on production behaviors ($b_2 = .435$, $p < .01$) were a significant predictor of leader effectiveness. Self-ratings on production behaviors ($b_1 = .072$, $p = .42$) did not significantly predict leader effectiveness. In other words, the regression

coefficients suggest that the higher the peer ratings of production behaviors (regardless of the level of self-ratings), the higher the ratings of leader effectiveness.

For self-peer agreement on ratings of *feedback* behaviors (Table 12), results indicated that the squared difference model did not account for more variance, $\Delta R^2 = .004$, $F$ (3, 726) = 0.83, $p = .48$, beyond the simpler algebraic difference model. The simpler algebraic difference model was also non-significant, $R^2 = .001$, $F$ (2, 729) = 0.43, $p = .65$. Thus, the data did not support the squared difference model or the simpler algebraic model.

Agreement between Self and Direct Report Ratings of Managing Others
Behaviors

Hypothesis 3c stated that a significant relationship would exist between self-direct report agreement (on ratings of behaviors related to *managing others*) and leader effectiveness. The resulting unstandardized regression coefficients are reported in Table 10 (management focus), Table 11 (production), and Table 12 (feedback).

For self-direct report agreement on *management focus* behaviors (Table 10), results indicated that the squared difference model did not account for more variance, $\Delta R^2 = .005$, $F$ (3, 726) = 1.27, $p = .28$, beyond the simpler algebraic difference model. The simpler algebraic difference model accounted for a significant amount of variance, $R^2 = .039$, $F$ (2, 729) = 14.81, $p < .01$. However, an inspection of the unstandardized coefficients revealed that only direct report ratings of management focus behaviors ($b_2 = .491$, $p <$

.01) were a significant predictor of leader effectiveness. Self-ratings of management focus behaviors ($b_1$ = .154, $p$ = .07) did not significantly predict leader effectiveness. In other words, the regression coefficients suggest that the higher the direct report ratings of management focus behaviors (regardless of the level of self-ratings), the higher the ratings of leader effectiveness.

For self-direct report agreement on *production* behaviors (Table 11), results indicated that the squared difference model did not account for more variance, $\Delta R^2$ = .003, $F$ (3, 726) = 0.87, $p$ = .46, beyond the simpler algebraic difference model. The simpler algebraic difference model accounted for a significant amount of variance, $R^2$ = .046, $F$ (2, 729) = 17.52, $p$ < .01. However, an inspection of the unstandardized coefficients revealed that only direct report ratings of production behaviors ($b_2$ = .665, $p$ < .01) were a significant predictor of leader effectiveness. Self-ratings of production behaviors ($b_1$ = -.001, $p$ = .99) did not significantly predict leader effectiveness. In other words, the regression coefficients suggest that the higher the direct report ratings of production behaviors (regardless of the level of self-ratings), the higher the ratings of leader effectiveness.

For self-direct report agreement on *feedback* behaviors (Table 12), results indicated that the squared difference model did not account for more variance, $\Delta R^2$ = .005, $F$ (3, 726) = 1.22, $p$ = .30, beyond the simpler algebraic difference model. The simpler algebraic difference model was also non-significant, $R^2$ = .001, $F$ (2, 729) = 0.26, $p$ = .77. Thus, the data did not support the squared difference model or the simpler algebraic model.

Differences across Rater Sources on Managing Others Behaviors

Hypothesis 3d sought to examine the differences (on ratings of behaviors related to *managing others*) across rater sources, stating that the relationship between self-direct report agreement and leader effectiveness would be significantly stronger compared to self-supervisor and self-peer agreement. However, because the squared difference model (used to test agreement between rater sources) was not supported for all of the self-other group comparisons, this hypothesis could not be tested.

CHAPTER IV

DISCUSSION

Multi-source feedback (MSF) has been referred to as the most noteworthy management innovation of the 1990s (Waldman & Atwater, 1998). Organizations have embraced the use of MSF, consultants commonly recommend its implementation for leadership development purposes, and many employees now recognize the value in receiving feedback from multiple sources beyond their supervisor. Currently, thousands of employees have been part of a MSF process within their organization (Waldman & Atwater, 1998). Among a number of potential benefits, the most obvious purpose of MSF is to enhance self-awareness by receiving feedback on the way one is perceived by others, with the goal of maximizing skill development, self-enrichment, and leadership performance (Morgeson et al., 2005).

Despite its popularity, largely due to the detailed feedback it provides, MSF has its drawbacks. It can present an overwhelming amount of information to the recipient, making it difficult to identify, process, and interpret the primary findings based on the feedback (DiNisi & Griffin, 2001). It is also a fairly complicated tool, often requiring the assistance of a facilitator or coach in order to make sense of the data and create specific, action-oriented goals. For this reason, researchers recommend that MSF be reviewed with a qualified specialist or consultant given the high likelihood of misinterpreting the MSF results if the recipient is left to interpret the feedback on his/her own (Antonioni, 1996). This potential overload of information

(some of which includes negative or destructive feedback) could lead MSF recipients to feel confused, overwhelmed or frustrated. At its worst, MSF results could lead to tension or dysfunctional relationships among team members. Thus, given the prevalence of MSF, it is critical that organizations, participants, and facilitators have a clear understanding of how to make the best use of MSF: by interpreting potentially conflicting or confusing results, focusing in on key themes, facilitating a conversation to uncover the unique context in which the leader operates, and utilizing the rich feedback gathered from multiple sources to create developmental goals and priorities.

In an effort to develop a greater understanding of MSF, this dissertation explored one key component – the degree of similarity (or agreement) between self and observer ratings, and the degree to which this agreement predicts perceived leader effectiveness. MSF recipients are often advised to pay close attention to large discrepancies between their self-ratings and others' observations of their behavior (Antonioni, 1996). Research has shown that when managers receive lower ratings from others (i.e., compared to their self-ratings), they are motivated to reduce this discrepancy (Johnson & Ferstl, 1999). For example, if a leader believes that she frequently provides feedback to her team, but her direct reports rate her relatively low on this behavior, this discrepancy could motivate the leader to make critical behavioral changes.

Although examining discrepancies is a useful starting point, the feedback may reveal multiple discrepancies in different areas (i.e., depending

on the behavior and rater source examined), leaving the MSF recipient overwhelmed or confused, particularly if the implied behavioral changes seem to conflict with one another. Thus, this dissertation sought to identify the specific *behaviors* and *sources* for which self-other agreement is most important when determining leader effectiveness in an effort to minimize the potential for feelings of confusion and information overload when leaders receive MSF.

<p align="center">Examining Self-Other Agreement</p>

Inconsistent with the study's predictions and previous research, the results indicate that self-other agreement may not be an important predictor of leader effectiveness. In fact, self-other agreement only predicted effectiveness for two leadership behaviors: management focus (i.e., for self-supervisor agreement) and innovative (i.e., for self-peer agreement; see Table 13). Furthermore, even for these significant results, the actual size of the effect was small, indicating that agreement may not have much of an impact on perceived effectiveness.

Interestingly, behavioral ratings from observers (i.e., supervisors, peers, and direct reports) were stronger predictors of leader effectiveness, compared to self-other agreement as well as self-ratings of behavior. These results suggest that observer ratings of leadership behaviors are the most powerful predictors of leader effectiveness, and that self-ratings and agreement (between self-other ratings) are not highly important when it comes to predicting leader effectiveness. One potential explanation for this

surprising result could be that previous studies collapsed all leadership behaviors into one broad dimension instead of examining relationships between self-other agreement for each specific behavior. For example, Atwater et al. (1998) used a MSF instrument which included 16 scales of managerial behaviors; however, for the purposes of their study, they averaged the 16 scales into a single measure of 'overall managerial performance.' It is possible that upon examining each specific leadership behavior within their scale (i.e., individually rather than combined), self-other agreement may be important only for a few select behaviors among specific rater groups, as the current research has revealed.

Despite the fact that in most cases, self-other agreement did not predict leader effectiveness, the results have several implications for the use and interpretation of MSF. The following sections describe several potential uses for self-other agreement (i.e., beyond predicting leader effectiveness) as well as recommendations for practitioners, coaches, and recipients of MSF.

Table 13

*Summary of Results*

| Implication of Findings | Leadership Behavior | Observer Group |
|---|---|---|
| *More* is better: | | |
| Positive relationship with leader effectiveness | Strategic | All Groups* |
| | Management Focus | Peers, Direct Reports |
| | Production | All Groups |
| *Less* is better: | | |
| Negative relationship with leader effectiveness | Cooperation | Peers, Direct Reports |
| | Empathy | Peers, Direct Reports |
| *Moderate* is better: | | |
| Curvilinear relationship with leader effectiveness | Cooperation | Supervisors |
| | Consensual | Supervisors, Peers |
| | Empathy | Supervisors |
| *Agreement* is better: | | |
| Self-other agreement predicts leader effectiveness | Innovation | Peers |
| | Management Focus | Supervisors |

*Indicates that the finding was observed for supervisors, peers, and direct reports.

Implications of Findings

Self-Other Agreement: How Should It Be Used?

Given the finding that self-other agreement may not always be a significant predictor of leader effectiveness, there are several important implications for MSF recipients and practitioners (i.e., facilitators or executive coaches). For the MSF recipient, he/she should not be immediately alarmed or disappointed by a lack of agreement between his/her self and observer ratings. In fact, it is common for various rater groups to provide different ratings of the same individual (Harris & Schaubroeck, 1988). Based on these results, a lack of agreement may not necessarily indicate that the leader is viewed as ineffective by others. Therefore, recipients of MSF should be advised not to make this assumption if discrepancies exist, and rather, should discuss the implications of the discrepancies with the facilitator.

For the coach or facilitator, he/she should also be careful not to imply that complete alignment is the ultimate goal of MSF or that lack of alignment implies that the leader is ineffective. Instead, agreement (or lack thereof) should be used as a discussion point when reviewing MSF results. Although agreement may not be a strong predictor of effectiveness, there may still be value in examining the extent to which rater sources agree or disagree as a way of uncovering potential "blind spots" (i.e., areas where the individual is unaware of the way he/she is perceived by others) and initiating behavioral change. Openly discussing the feedback and uncovering the reasons behind discrepancies in self-other ratings could lead to important self-realizations.

For example, a leader might believe that he is highly strategic, but his supervisor might give him low ratings as a strategic thinker. The coach could then facilitate a conversation regarding what it means to *display* strategic thinking. It may be that the leader thinks in strategic ways; however, this capability may not be effectively communicated or revealed to others. If this is the case, the coach could then provide guidance on ways to more effectively display this behavior to others.

Finally, if attempting to determine a leader's effectiveness for an appraisal of one's work, it could be misleading to examine the degree of alignment (i.e., between self and observer ratings) as an indicator of leader effectiveness. While there may be unique cases or situations where agreement matters, and could in fact predict effectiveness on the job, this is likely the exception, not the rule. Thus, practitioners should be careful if they are using MSF results as an indicator or predictor of leader effectiveness. Instead, it is recommended that leader effectiveness be measured using a separate method (i.e., other than an MSF survey) which is a tested and valid predictor of effectiveness or performance in a leadership role. Otherwise, if selection, promotion, or salary decisions are made based on an un-validated MSF measure, the company's process may not be considered legally sound.

Considering Context and Culture

The specific culture and context of the organization from which the data were collected may provide a potential explanation for the unexpected results (regarding the non-significance of agreement as a predictor of

effectiveness). Dierdorff and Surface (2007) explain the importance of considering the environmental context when examining MSF ratings because certain behaviors are viewed as more or less effective within a given setting or situation. Regarding this particular organization, there are several noteworthy characteristics. First of all, participants were leaders in a large financial institution with a long and stable history. In this organization, leaders are rewarded for executing their thinking with decisiveness and confidence. Because the organization has a skilled and seasoned workforce composed mostly of subject matter experts, decisions are made fairly independently and employees tend to operate within silos (i.e., functional departments). As such, forming close relationships with colleagues, particularly across departments, is not as highly valued as being a shrewd decision-maker who is able to deliver impressive results. As an example, this could be an explanation for the negative and curvilinear relationships that were observed for several behaviors related to collaboration and teamwork (which will be discussed in the following sections).

As an example of the importance of organizational context in MSF results, the amount of feedback provided by the leader was not a significant predictor of perceived effectiveness. This finding can be understood by considering the unique characteristics of this particular work environment. Because leaders in this organization represent a highly skilled and mature workforce, there is not a significant need to provide feedback or have a strong inclination towards developing others. Instead, this organization tends to

attract experts in the field, and as such, these individuals are already considered to be experienced, capable, and credible in their roles. Leaders in this organization do not frequently solicit feedback because they often view themselves as specialists who are expected to be confident in their knowledge and capabilities. Also, because the organization relies heavily on following historical precedence, employees are less likely to 'rock the boat' by questioning others or providing constructive feedback to their colleagues. Possibly as a result of these factors, giving direct feedback is not a highly expected or rewarded leadership behavior in this organization. This example illustrates a potential explanation of these results and encourages future researchers to examine the extent to which the organizational culture and climate have an impact on the degree of effectiveness associated with certain leadership behaviors.

Organizational culture and norms may also influence the degree to which certain behaviors are related to effective leadership; an additional consideration when interpreting MSF results. Schein (1992) defines the culture of an organization as the shared, underlying assumptions and beliefs of its members. One of the commonly observed norms within this particular organization is that feedback is not openly shared among colleagues, but that it is primarily given during one-on-one formal performance reviews. Thus, only two rater groups (i.e., self and supervisors) would typically observe behaviors related to feedback, which is what the data revealed. Specifically, across all leadership behaviors, peers and direct reports rated feedback as the

least frequently observed behavior. When discussing MSF results within this organization, the facilitator might want to spend relatively less time discussing ratings in the area of feedback, given that this behavior does not appear to impact one's perceived effectiveness. However, other organizations that are more customer service oriented, family-owned, or smaller in size may place a higher priority on giving open and honest feedback – and thus, this behavior could then be important to perceived effectiveness.

<div align="center">Behaviors for Which "<em>Moderate</em> is better"</div>

When interpreting MSF results, some recipients or facilitators might assume that "more is better" in terms of the leadership behaviors displayed. However, these results indicate that this assumption might not be universally true. In a few cases, particularly when it comes to collaborative behaviors (i.e., cooperative, consensual, empathy), curvilinear relationships exist. In other words, for collaborative behaviors, leader effectiveness was highest when observer ratings were *moderate*, and leader effectiveness decreased as observer ratings of these behaviors became either high or low. One potential explanation is that moderate levels of certain behaviors are actually viewed as most effective among leaders. For example, a leader who is rated highly in terms of empathy is likely to be sensitive and supportive of others; however, this person might actually be perceived as too concerned with others' reactions or may struggle to be objective when handling sensitive personnel issues. Another possibility is that certain situations may require more or less of the behavior and that consistently high or low levels of the behavior may

indicate that the leader is inflexible or insensitive to the demands of the situation.

The curvilinear relationship observed for consensual behaviors provides an example of a situation in which "more does not always imply better." A leader who displays consensual behaviors is one who values and solicits the opinions of others as part of the decision-making process. These leaders often encourage others to share ideas and tend to seek consensus before taking action. The organizational culture may shed light on these results, given that leaders in this organization are seen as subject matter experts who are known for their ability to make smart and quick decisions. If a leader displays *low* levels of consensual behaviors, he may be perceived as ignoring the expertise of others or unresponsive to others' ideas. On the other hand, if a leader displays *high* levels of consensual behaviors, he may be perceived as spending too much time gathering input or belaboring decisions. At their worst, highly consensual leaders could be seen as indecisive or unwilling to take a stand on critical issues. Thus, at both ends of the scale, consensual behaviors have potential drawbacks, which may explain why moderate levels of consensual behaviors are related to higher levels of effectiveness. Overall, these findings point to the importance of conducting an organizational analysis to uncover the unique dynamics and desirable behaviors within the organization.

Research on participative leadership could also help to explain the non-linear relationship observed for consensual behaviors. Participative

leadership involves group decision-making and consultation of employees at all levels (Yukl, 2006). Although this type of leadership has been shown to increase performance and satisfaction among followers, some research reveals that participative leadership may not have a significant impact on performance-related outcomes (Yukl, 2006). These inconsistent results can be explained by the idea that different types of participation may be more or less effective depending on the demands and requirements of the situation.

Contingency theories of leadership recognize the fact that the situation may determine the effectiveness of various approaches to leading others (Yukl, 2006), including the amount of participation or collaboration that is most appropriate. Vroom and Yetton (1973) developed the Normative Decision Model, which outlines the decision-making procedure that is believed to be most effective in specific situations. Situational variables that warrant consideration include the amount of information possessed by the leader and followers, the likelihood that followers will accept a non-participative decision, and the extent to which the decision requires creative problem-solving. Yukl (2006) builds upon Vroom and Yetton's model by proposing guidelines for participative leadership. For example, in time-pressured or crisis situations, a leader who takes charge by making an autocratic decision is often viewed as more effective than one who involves all team members in the decision. Research is still needed to test the efficacy of Yukl's participative leadership guidelines. However, results of this

dissertation provide initial support for the idea that consensual behaviors may not have a linear relationship with leadership effectiveness.

<p style="text-align:center">Behaviors for Which "*Less* is better"</p>

The results also revealed another interesting finding: for some behaviors, *lower* ratings were actually related to *higher* leadership effectiveness.  According to ratings from peers and direct reports, lower levels of cooperative and empathetic behaviors were predictive of greater effectiveness.  At first this might seem counterintuitive, given that both of these behaviors are related to working well with others and building strong relationships.  Cooperative leaders tend to be viewed as accommodating, helpful, and willing to compromise.  Similarly, empathetic leaders are typically seen as caring, sensitive, and able to form close and supportive relationships with their colleagues.  Again, in this organizational culture, leaders are typically promoted for being smart, decisive, and results-oriented.  They are unlikely to be promoted based on their ability to develop trusting and open relationships with their team members.  It may be that highly cooperative and empathetic leaders are viewed as spending too much time caring about others' opinions or feelings, and not enough time making decisions, achieving goals, or delivering tangible results.

The finding that "less is better" when it comes to cooperative and empathetic behaviors could also be explained with the concept of "need for affiliation" (Yukl, 2006).  Similar to empathetic behaviors, leaders with a high need for affiliation enjoy being liked by others and work hard to develop close

relationships with their co-workers.  Yukl (2006) suggests that leaders with a high need for affiliation may be perceived as more concerned with building relationships than performing tasks.  These leaders may also avoid making unpopular decisions, have a tendency to steer clear of conflict, or show favoritism toward close friends.  Thus, given the potential undesirable consequences of high levels of need for affiliation, it is more easily understood why lower levels of empathy would be related to greater effectiveness in a results-focused organization.

<div align="center">Behaviors for Which "<em>More</em> is better"</div>

There are also certain behaviors for which *higher* levels are related to more effective leadership.  For these three behaviors (i.e., strategic, production, and management focus), all observer groups similarly rated leaders who display these behaviors as more effective.  In other words, when leaders display more of these behaviors during work interactions (regardless of the group with whom they are interacting), they are seen as more effective leaders.  A common theme among these behaviors is that they are all related to meeting tangible business goals, which is a highly valued ability within this organization.  Strategic behaviors require an understanding of the long-term direction of the organization, while management focus and production behaviors are focused on taking charge, directing others, and pushing others to achieve objectives.  These three behaviors are fairly concrete and tangible compared to team-playing or relationship-based "soft skills," such as showing concern and sensitivity toward others (i.e., empathy).  Because leader

effectiveness is often related to, or even defined as one's ability to perform (e.g., delivering tangible results), this could explain why behaviors related to strategic decision-making and ability to achieve outcomes are related to greater perceived effectiveness.

The positive relationship for strategic behaviors can also be understood through transformational leadership theory (Bass, 1985). Leaders described as "transformational" focus their efforts on establishing long-term goals, developing a vision, and inspiring followers to pursue the vision. Similarly, strategic behaviors (as measured in this study) are related to demonstrating a longer term, broad perspective and creating an orientation toward the future, which is similar to the behaviors exhibited by transformational leaders. Transformational leaders are often described as inspirational by their followers and measures of transformational leadership have been linked to key organizational outcomes, such as higher business-unit performance (Howell & Avolio, 1993); thus, it is consistent with the theory of transformational leadership that higher ratings of strategic behaviors would be related to more effective leadership.

The significant findings for management focus and production behaviors can also be explained by examining a prominent leadership theory. Transactional leadership theory (Bass, 1985) includes a component called "contingent reward leadership," which involves a series of exchanges between leaders and followers in which followers are rewarded or recognized for accomplishing mutually agreed-upon goals. Contingent reward leadership is

similar to management focus and production behaviors – leaders who display these behaviors often take charge, direct others' efforts, keep others focused on results, and create an achievement atmosphere. Contingent reward leadership is positively related to follower performance and job satisfaction (e.g., Podsakoff, Todor, Grover, & Huber, 1984); thus, it is logical that management focus and production behaviors are also related to higher levels of perceived leader effectiveness.

In sum, a consideration of situational factors embedded in the organizational culture may provide for a better understanding of the relationship between multi-source ratings of leadership behaviors and perceived effectiveness. When interpreting MSF results, recipients and facilitators should keep in mind that there are certain behaviors for which low, moderate, or high amounts can be viewed as most desirable or effective (according to their observers). Furthermore, agreement may not always occur between self and observer ratings and this does not necessarily imply that the leader is ineffective. Therefore, it may be important to consider the context when interpreting MSF results, moving away from absolute conclusions (e.g., "more is better, "alignment is ideal") and toward a more tailored conversation that considers the specific needs and values of leaders within the organization.

### Limitations

This study should be considered in light of a few limitations. Most notably, there are several weaknesses in the way the dependent variable (i.e., leader effectiveness) was measured. First, leader effectiveness ratings were

gathered at the same point in time as the leadership behavior ratings (i.e., the

independent variable). Thus, the current study does not capture behavioral

changes over time as a result of MSF. Instead, this study answers the question

of whether or not self-other agreement is related to *current* perceived leader

effectiveness. Kluger and DeNisi (1996) propose that discrepancies in self-

other ratings may signal the need for behavioral adjustments and may

motivate MSF recipients to change their behavior in the future. Thus, future

research should also measure changes in behavior over time to determine

whether or not self-other agreement has an impact on behavioral change or

performance improvement.

Second, measurement equivalence was not supported for the leader

effectiveness scale. In other words, different rater groups (i.e., supervisors,

peers, direct reports) may have used the leader effectiveness scale differently,

implying that direct comparisons cannot be made across groups. For the

purposes of this study, supervisory ratings were used to measure leader

effectiveness because supervisors' ratings of performance are considered most

critical in pay and promotional decisions (Atwater et al., 1998; Fleenor, 1996).

Because it was not the focus of the current study, the reasons behind the

failure of measurement equivalence were not fully explored. Researchers

have suggested that different rater groups may have unique perceptions of

what constitutes effective performance in a particular job (Campbell & Lee,

1988) and that rater groups may differ in their opportunities to observe

specific work behaviors, which could result in divergent ratings of

effectiveness (Murphy & Cleveland, 1995). Future research should continue

to investigate the assumption that ratings of leader effectiveness are equivalent

across sources, given that leader effectiveness may have been interpreted

differently across rater groups included in this study.

Additionally, it is possible that the four items used to measure leader

effectiveness do not provide the best fit to the data. As previously described,

the RMSEA value was above the preferred cutoff point of .08 (for the

supervisor group), which is often inflated when degrees of freedom are small

(Kenny, 2008). In this case, because there are only two degrees of freedom,

this could be artificially inflating RMSEA. On the other hand, it could also

indicate a weakness in the measurement of leader effectiveness that should be

considered when interpreting results.

Third, results of this study are susceptible to same source bias.

Specifically, for some of the analyses, supervisors completed ratings of

leadership behaviors as well as ratings of perceived leader effectiveness.

Thus, it is possible that the relationship between supervisor ratings of

behaviors and leader effectiveness is often stronger (i.e., compared to peers

and direct reports) because the same source is rating both variables. However,

despite the likelihood of same source bias, there were a few behaviors for

which self-supervisor agreement was not the strongest predictor of

effectiveness (e.g., self-peer agreement on innovative behaviors was most

strongly related to effectiveness). Also, despite the fact that including

outcome variables from other sources would have strengthened the findings of

this study, supervisors' ratings of perceived leader effectiveness were utilized given their use in organizational decision-making such as salary and promotion decisions. Thus, when considering the point of view of the individual receiving the feedback, he/she is likely to care most about his/her supervisor's perception of overall effectiveness as a leader.

Fourth, the results generally reveal small to moderate effect sizes. Across all regression analyses, the largest R-squared value was .203 (for strategic behaviors), which Cohen (1992) describes as a small to medium effect size. Further, many of the relatively larger effect sizes were observed when the supervisor provided ratings of both variables, implying that some of this effect may have been due to same source bias (as previously discussed). Again, this suggests that self-other agreement on ratings of leadership behaviors may not be a powerful predictor of leader effectiveness.

Lastly, participants in this study were from a single organization which was a large financial institution. Some of the findings may be explained by the nature of the organization's culture (e.g., conservative, risk averse, and individualistic versus team-oriented). While organizational context provides for a potential interpretation of the results, the role of context was not examined in the study. Future research may consider the impact of culture on the relationship between MSF ratings and perceived effectiveness by examining the relative importance placed on specific leadership dimensions across other organizations and industries.

Suggestions for Future Research

One of the most surprising results was that self-other agreement was not a significant predictor of leader effectiveness for most of the leadership behaviors examined. Future research should continue to explore the relationship between agreement and leader effectiveness to provide greater clarity toward this discrepant finding. One potential reason for the non-significant findings is that this study examined *specific* leadership behaviors among specific rater groups. It is possible that a different relationship is revealed when collapsing leader behaviors and/or rater sources, as previous research has done (e.g., Atwater et al., 1998). It is also possible that a different relationship would be discovered if agreement was measured in a different way. For example, agreement could be a stronger predictor of effectiveness when self-ratings are high and *all* observer ratings are low and clustered together. As such, future research could examine both the degree to which observer ratings cluster together as well as the degree of agreement between self and aggregated observer ratings (i.e., instead of examining each separate observer group). It may be that when leaders overestimate themselves and all of their observers are in-agreement *and* provide low ratings, this could be a significant 'blind spot' which limits leader effectiveness.

Future research could also continue to explore the outcomes (e.g., effectiveness, performance, behavioral changes) for which self-other agreement is a significant predictor. In this dissertation, the outcome variable

was perceived effectiveness by the supervisor; however, future research could also examine effectiveness ratings from other sources, as well as other measures of performance, such as more objective criteria (e.g., sales, productivity). For more objective measures of leader effectiveness, self-other agreement may also not be as critical; however, not much research has explored this question. Atwater et al. (1998) suggests that self-other agreement might be more relevant for subjective outcomes (e.g., perceptions of effectiveness) and less relevant for objective measures such as performance criteria. Gaining clarification on this issue will help practitioners to better understand whether or not leaders should focus on aligning their self-ratings with observer ratings when faced with discrepancies in their MSF results. The results of this research suggest that agreement may not be as critical for perceived effectiveness as was previously suggested; however, having more objective criteria would help to further understand this relationship.

Future research could also explore one of the clearest findings from this research: ratings of strategic behaviors were the strongest predictors of leader effectiveness. In other words, the extent to which observers perceive the leader to be future-oriented, capable of long-term planning and able to communicate a vision for the future of the organization is predictive of the leader's perceived effectiveness. Future research could explore the concept of strategic leadership. According to House and Aditya (1997), one of the emerging trends within leadership research is strategic leadership. However, much of the previous research on strategic leadership has been based on

qualitative data, such as case studies, and the few studies that have

incorporated quantitative data have involved small sample sizes (Avolio,

Sosik, Jung, & Berson, 2003).  Thus, this study presents initial findings based

on a large, quantitative data set that point to the importance of strategic

leadership.  Future research could also explore the relative importance of

strategic leadership behaviors across environmental conditions (e.g., economic

state) and over time.  For example, is strategic leadership more critical in

times of crisis, change, or uncertainty?  It is possible that strategic behaviors

(e.g., identifying potential risks, opportunities, and challenges; creating a

unified vision for the organization) become even more crucial to leader

effectiveness during an economic downtown when leaders are looked upon to

provide clarity, focus, and direction for their followers.

This research also sheds light on the complexities and challenges

associated with interpreting MSF.  Given that at least 90% of Fortune 1000

firms have used some form of multi-source assessment (Atwater & Waldman,

1998), research should continue to explore ways to get the most out of MSF,

including the most effective techniques for structuring and guiding the

feedback discussion with the MSF recipient.  Organizations would likely

benefit from clear guidelines on ways to make the best use of MSF and avoid

MSF being implemented as a popular practice that may not add significant

value to employees' development.  Some researchers (e.g., Antonioni, 1996;

Atwater & Waldman, 1998; London et al., 1991) have begun to provide useful

recommendations for interpreting MSF results and researchers should

continue to do so as MSF systems become more prevalent and utilized for a variety of purposes, including performance reviews or promotional decisions.

<div align="center">Implications for Practice</div>

One of the goals of this study was to help leaders, as well as executive coaches and facilitators, make the best use of MSF. This dissertation is based on the premise that if MSF recipients have a better understanding of where to focus their attention (i.e., in terms of specific behaviors and rater sources), they will be able to capitalize on the benefits of MSF (i.e., change their behavior in a way that enhances their professional development; Morgeson et al., 2005). With this objective in mind, these results provide a few suggestions for leaders and coaches to help them interpret, discuss, and take action based on the most relevant and useful MSF results.

Because of the potential for information overload or misinterpretation of results, MSF facilitators should help recipients by narrowing their attention to specific rater groups and behaviors that are most critical to their perceived effectiveness. The current findings imply that there are certain behaviors and rater sources for which observer ratings (i.e., regardless of the degree of agreement with self-ratings, or self-ratings on their own) are significant predictors of perceived effectiveness. Practitioners should be aware of these behavioral trends in order to alert MSF recipients to potential "watch outs" or "red flags" when interpreting their MSF results (e.g., very high levels of empathy may not be viewed as effective in this particular organization).

Given the cultural assumptions and norms that are believed to influence individuals' perceptions of "effective leadership," MSF is likely to be most useful when results are discussed between the MSF recipient and a personal coach (consistent with suggestions from Antonioni, 1996). The reason for this is that the coach is able to ask questions aimed at uncovering the underlying reasons behind trends in the data (such as those previously described). The coach can also facilitate a conversation about the behaviors that are viewed as most appropriate given the interpersonal dynamics, political environment, and organizational culture in which the leader operates. For example, if the supervisor provides low ratings of strategic behaviors, the coach could prompt the leader to think about what her supervisor expects in terms of strategic thinking. Does the leader understand what it means to be strategic at that particular organization? Do they have the time and resources needed to plan ahead and formulate a long-term vision for their team? Is it possible that the leader possesses these capabilities, but is not able to *display* her strategic thinking to others because of difficulty communicating, influencing, or standing up to colleagues?

One of the most important findings in this research is that higher ratings on MSF surveys do not always relate to higher levels of perceived effectiveness. In fact, in some cases, lower levels of behaviors are related to more effective leadership. If MSF is used in the performance review process, as an increasing number of organizations have begun to do (London & Smither, 1995), there is potential to assume that higher levels of MSF-rated

behaviors are more desirable or more effective, when in fact; this may depend

on situational variables such as the organizational culture, context, and

perceptions of desirable leadership behaviors.  This underscores the

importance of conducting a thorough job analysis to determine the leadership

behaviors that are significantly related to performance and effectiveness in

order to ensure a legally defensible selection and promotion system.

In summary, there are several relevant implications for organizations

using MSF.  As a starting point, MSF recipients and facilitators should

carefully consider the context in which a leader operates.  What are the

behavioral norms – what leadership behaviors (e.g., collaborating with others,

delivering feedback, displaying empathy) are commonly or rarely displayed?

What leadership behaviors are rewarded and how are employees typically

promoted into leadership roles?  Which rater groups have the best insight into

these critical leadership behaviors?  This information will help to guide a

discussion of MSF results.  Next, because certain behaviors are viewed as

more effective than others, facilitators should discuss specific and definable

behaviors instead of broad, aggregated dimensions.  Displaying more of

certain leadership behaviors may not universally equate to higher

effectiveness, and in fact, some behaviors might be better at moderate or even

low levels.  MSF recipients should be guided to think about the situations in

which more or less of these behaviors may be more effective.  Lastly, MSF

facilitators and recipients should not be discouraged if there are differences

between self and observer ratings and, in fact, gaining insight into divergent

perspectives is one of the advantages of receiving MSF (Morgeson et al., 2005).  Misalignment does not necessarily imply ineffectiveness.  Above all else, this study demonstrates the complexities of leadership – perceptions of effectiveness are likely to depend on a number of factors, including the organizational context and culture, the audience, and the value placed on the specific leadership behavior being rated.

CHAPTER V

SUMMARY

Multi-source feedback (MSF) refers to the process of soliciting feedback from followers, peers, and supervisors in order to provide a comprehensive viewpoint of an individual's leadership style (Nowack, 1993). The underlying premise of MSF is that leadership development can be initiated through an examination of discrepancies between self and observer ratings on a number of behavioral dimensions (Morgeson et al., 2005). Although MSF provides the recipient with rich and detailed feedback, the amount of information could be overwhelming to the recipient or difficult to interpret if multiple discrepancies (i.e., across different behaviors or sources) exist. This information overload could limit the MSF recipients' ability to set specific developmental goals, which is one of the recommended outcomes of MSF (Antonioni, 1996).

In an effort to help MSF recipients interpret their results, this research examined one key component of MSF, the degree to which self and observer ratings are aligned, and the relationship between self-other agreement and perceived leader effectiveness. Research suggests that higher performing leaders tend to be more self-aware, and that self-awareness can be measured through the degree of agreement between self and observer ratings (Church, 1997). Although previous research has examined the relationship between self-other agreement and effectiveness (e.g., Atwater et al., 1998; Tekleab et al., 2008), this study sought to identify the specific *behaviors* and *sources* for

which self-other agreement is most predictive of leader effectiveness. Hypotheses were based on the premise that certain sources may be better suited to provide feedback on specific behaviors because they observe the individual in different settings and have a unique understanding of what it takes to effectively display that behavior (Greguras et al., 2003).

Hypotheses were tested using archival data which included 847 leaders from a large, financial organization. Participating leaders completed self-assessments of their leadership behaviors using the *Leadership Effectiveness Analysis* (LEA; Management Research Group, 1992) as part of a leadership development program. For each leader, a combination of supervisors, peers, and direct reports anonymously completed the LEA, and supervisors also completed a separate survey to measure leader effectiveness. Polynomial regression was used to test all hypotheses (Edwards, 1994).

Inconsistent with the study's predictions and previous research, results revealed that self-other agreement may not be an important predictor of leader effectiveness. Self-other agreement only predicted effectiveness for two (out of the eight that were examined) leadership behaviors and effect sizes were small, indicating that agreement may not be a strong predictor of leader effectiveness. Instead, results revealed that *observer* ratings of leadership behaviors were the most powerful predictors of leader effectiveness. Furthermore, results indicated that unique relationships exist between leadership behaviors and perceived effectiveness. For certain behaviors,

*higher* ratings were related to greater effectiveness, while for others, *lower* or *moderate* levels of the behavior were viewed as more effective.

These findings reinforce the complexities and challenges associated with interpreting MSF and provide a few implications for practitioners. First, agreement may not always occur between self and observer ratings and this may not necessarily indicate that the individual is ineffective. Second, observer ratings on specific leadership behaviors are most predictive of the leader's perceived effectiveness, and third, there are certain behaviors for which low, moderate, or high amounts are viewed as most effective according to observers. Although not the focus of this study, results may point to the importance of considering the situational factors embedded in the organizational culture to provide a better understanding of the relationship between multi-source ratings of leadership behaviors and perceived leader effectiveness.

REFERENCES

Antonioni, D. (1994). The effects of feedback accountability on upward appraisal ratings. *Personnel Psychology, 47*, 349-356.

Antonioni, D. (1996). Designing an effective 360-degree appraisal feedback process. *Organizational Dynamics, 25*, 24-38.

Ashford, S. (1989). Self-assessments in organizations: A literature review and integrative model. *Research in Organizational Behavior, 11*, 133-174.

Ashford, S. J., & Tsui, A. S. (1991). Self-regulation for managerial effectiveness: The role of active feedback seeking. *Academy of Management Journal, 34*, 251-280.

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

Atwater, L. E., Brett, J. F., & Charles, A. C. (2007). Multisource feedback: Lessons learned and implications for practice. *Human Resource Management, 46*, 285-307.

Atwater, L. E., Ostroff, C., Yammarino, F. J., & Fleenor, J. W. (1998) Self-other agreement: Does it really matter? *Personnel Psychology, 51*, 577-598.

Atwater, L. E., Roush, P., Fischthal, A. (1995). The influence of upward feedback on self and follower ratings of leadership. *Personnel Psychology, 48*, 35-59.

Atwater, L. E., & Waldman, D. (1998). Accountability in 360 degree feedback. *HR Magazine, 43*, 96-104.

Atwater, L. E., Waldman, D., Atwater, D., & Cartier, T. (2000). An upward feedback field experiment. Supervisors' cynicism, follow-up and commitment to subordinates. *Personnel Psychology, 53*, 275-297.

Atwater, L. E., & Yammarino, F. (1992). Does self-other agreement on leadership perceptions moderate the validity of leadership and performance predictions? *Personnel Psychology, 45*, 141-164.

Atwater, L. E., & Yammarino, F. (1997). Self-other rating agreement: A review and model. *Research in Personnel and Human Resource Management, 15*, 121-174.

Austin, J. T., & Villanova, P. (1992). The criterion problem: 1917-1992. *Journal of Applied Psychology, 77*, 836-874.

Avolio, B. J., Sosik, J. J., Jung, D. I., & Berson, Y. (2003). Leadership models, methods, and applications. In Borman, W. C., Ilgen, D. R. Llimoski, R. J., & Weiner, I. B. (Eds.) The Handbook of Psychology. Hoboken, NJ: John Wiley & Sons, Inc.

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

Baumeister, R. F. (1982). A self-presentational view of social phenomena. *Psychological Bulletin, 91*, 3-26.

Borman, W. C. (1974). The rating of individuals in organizations: An alternate approach. *Organizational Behavior and Human Performance, 112*, 105-124.

Bracken, D. W., Timmreck, C. W., & Church, A. H. (Eds.). (2001). *The handbook of multisource feedback: The comprehensive resource for designing and implementing MSF processes*. San Francisco: Jossey-Bass.

Brett, J., & Atwater, L. (2001). 360-degree feedback: Accuracy, reactions and perceptions of usefulness. *Journal of Applied Psychology, 86*, 930-942.

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 136-162). Newbury Park, CA: Sage.

Brutus, S., & Derayeh, M. (2002). Multi-source assessment programs in organizations: An insider's perspective. *Human Resource Development Quarterly, 13*, 187-201.

Byrne, B. (1989). *A primer for LISREL: Basic applications and programming for confirmatory factor analytic models*. New York: Springer-Verlag.

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

Carver, C. S., & Scheier, M. F. (1982). Control theory: A useful conceptual framework for personality, social, clinical, and health psychology. *Psychological Bulletin, 92*, 111-135.

Charan, R., Drotter, S., Noel, J. (2001). *The leadership pipeline: How to build the leadership-powered company*. San Francisco, CA: Jossey Bass.

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

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

Church, A. H. (2000). Do higher performing managers actually receive better ratings? A validation of multirater assessment methodology. *Consulting Psychology Journal, 52,* 99-116.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*, 155-159.

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

Day, D. V. (2001). Leadership development: A review in context. *Leadership Quarterly, 11*, 581-613.

DeNisi, A. S., & Griffin, R. W. (2001). *Human resource management*. Boston: Houghton Mifflin.

DeNisi, A. S., & Kluger, A. N. (2000). Feedback effectiveness: Can 360-degree appraisals be improved? *Academy of Management, 14,* 129-139.

Dierdorff, E. C., & Surface, E. A. (2007). Placing peer ratings in context: Systematic influences beyond rate performance. *Personnel Psychology, 60,* 93-126.

Drasgow, F., & Kanfer, R. (1985). Equivalence of psychological measurement in heterogeneous populations. *Journal of Applied Psychology, 70,* 662-680.

Dweck, C. S. (1986). Motivational processes affecting learning. *American Psychology, 41,* 1040-1048.

Dweck, C. S., & Leggett, E. L. (1988). A social-cognitive approach to motivation and personality, *Psychological Review, 95,* 256-273.

Edwards, J. R. (1994). The study of congruence in organizational behavior research: Critique and proposed alternative. *Organizational Behavior and Human Decision Processes, 58,* 683-689.

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

Fleenor, J., McCauley, C., & Brutus, S. (1996). Self-other rating agreement and leader effectiveness. *Leadership Quarterly, 7,* 487-506.

Fleenor, J., & Prince, J. M. (1997). *Using 360-degree feedback in organizations.* Greensboro, N.C.: Center for Creative Leadership.

Funderburg, S. A., & Levy, P. E. (1997). The influence of individual and contextual variables on 360-degree feedback system attitudes. *Group and Organizational Studies, 22*, 210-230.

Garavan, T. N., & Morley, M. (1997). 360-degree feedback: Its role in employee development. *Journal of Management Development, 16*, 134-158.

Gebelein, S. (1996). Employee development: Multi-rater feedback goes strategic. *HR Focus,* 4-6.

Greguras, G. J., Ford, J. M., & Brutus, S. (2003). Manager attention to multisource feedback. *Journal of Management Development, 22*, 345-361.

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

Hendricks, J. W., & Payne, S. C. (2007). Beyond the Big Five: Leader goal orientation as a predictor of leadership effectiveness. *Human Performance, 20*, 317-343.

Hegarty, W. H. (1974). Using subordinate ratings to elicit behavioral changes in managers. *Journal of Applied Psychology, 59*, 764-766.

House, R. J., & Aditya, R. N. (1997). The social scientific study of leadership: Quo vadis? *Journal of Management, 23*, 409-473.

Howard, A., & Bray, D. W. (1988). *Managerial lives in transition: Advancing age and changing times.* New York: Guilford Press.

Howell, J. M., & Avolio, B. J. (1993). Transformational leadership, transactional leadership, locus of control, and support for innovation: Key predictors of consolidated-business-unit performance. *Journal of Applied Psychology, 78*, 891-902.

Johns, G. (2006). The essential impact of context on organizational behavior. *Academy of Management Review, 31*, 386-408.

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.

Kabacoff, R. (1998). *Leadership Effectiveness Analysis: Technical Considerations*. (Available from the Management Research Group, 14 York Street, Portland, ME 04101).

Kenny, D. A. (2008). *Measuring Model Fit* [Online tutorial]. Storrs, CT: University of Connecticut.

Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin, 119*, 254-258.

Korman, A. K. (1976). Hypothesis of work behavior revisited and an extension. *Academy of Management Review, 1*, 50-63.

Kulas, J. T., & Finkelstein, L. M. (2007). Content and reliability of discrepancy-defined self-awareness in multisource feedback. *Organizational Research Methods, 10*, 502-522.

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

Lawler, E. E. (1967). The multitrait-multirater approach to measuring managerial job performance. *Journal of Applied Psychology, 51*, 369-380.

Lepsinger, R., & Lucia, A. D. (1997). *The Art and Science of 360-Degree Feedback*. San Francisco, CA: Jossey-Bass.

Leslie, J. B., & Fleenor, J. W. (1998). *Feedback to managers: A review and comparison of multi-rater instruments for management development*. Greensboro, NC: Center for Creative Leadership.

Locke, E. A., & Latham, G. P. (2002). Building a practically useful theory of goal setting and task motivation. *American Psychologist, 57,* 705-717.

London, M., & Beatty, R. W. (1993). 360-degree feedback as a competitive advantage. *Human Resource Management, 32*, 353-372.

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

Longenecker, C. O., & Gioia, D. A. (1992). The executive appraisal paradox. *Academy of Management Executive, 6*, 18-28.

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

Mahoney, J., & Mahoney, C. (1996). *MRG questionnaire design: A hybridized approach to psychological test development.* Paper presented at the Network & Internal Associates Conference, Chicago, IL.

Management Research Group. (1992). *Leadership Effectiveness Analysis.* Portland, ME.

Manz, C. C., & Sims, H. P. Jr. (1987). Leading workers to lead themselves: The external leadership of self-managing work teams. *Administrative Science Quarterly, 32*, 106-128.

McCall, M. W., Jr., Lombardo, M. M., & Morrison, A. M. (1988). *The lessons of experience: How successful executives develop on the job.* Lexington, MA: Lexington Books.

McClelland, D. C. (1975). *Power: The inner experience.* New York: Irvington.

McGarvey, R., & Smith, S. (1993). When workers rate the boss. *Training*, 31-34.

McGregor, D. (1957). An uneasy look at performance appraisal. *Harvard Business Review, 35*, 89-94.

Miller, G. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological review, 63*, 81-97.

Mitchell, T. R., & Kalb, L. S. (1982). Effects of job experience on supervisor attributions for a subordinate's poor performance. *Journal of Applied Psychology, 67,* 181-188.

Morgeson, F. P., Mumford, T. V., & Campion, M. A. (2005). Coming full

circle: Using research and practice to address 27 questions about 360-

degree feedback programs. *Consulting Psychology Journal: Practice

and Research, 57*, 196-209.

Mount, M. K., Judge, T. A., Scullen, T. E., Sytsma, M. R., & Hezlett, S. A.

(1998). Trait, rater, and level effects in 360-degree performance ratings.

*Personnel Psychology, 51*, 557-576.

Murphy, K. R., & Cleveland, J. N. (1995). *Understanding performance

appraisal: Social organizational and goal-based perspectives*. Thousand

Oaks, CA: Sage Publications.

Nikolaou, I., Vakola, M., & Robertson, I. T. (2006). *360-degree feedback and

leadership development*. In Burke, R. J., & Cooper, C. L. (Eds).

*Inspiring Leaders*. New York, NY: Routledge.

Nowack, K. M. (1993). 360-degree feedback: The whole story. *Training &

Development, 47*, 69-72.

Parker-Gore, S. (1996). Perception is reality: Using 360-degree appraisal

against behavioral competencies to effect organizational change and

improve management performance. *Career Development International,

1*, 24-27.

Paulhaus, D. (1986). Self-deception and impression management in test

responses. In Angleitner A, Wiggins, J (Eds.), *Personality assessment

via questionnaire* (pp. 143-165). New York: Springer.

Podsakoff, P. M., & Organ, D. (1986). Self-reports in organizational

    research: problems and prospects. *Journal of Management, 12*, 531-544.

Podsakoff, P. M., Todor, W. D., Grover, R. A., & Huber, V. L. (1984).

    Situational moderators of leader reward and punishment behavior: Fact

    or fiction?  Organizational Behavior and Human Performance, 34, 21-

    63.

Radar, M. H. (1981). Dealing with information overload. *Personnel Journal,*

    *60*, 373-375.

Schein, E. H. (1992). *Organizational culture and leadership.*  San Francisco,

    CA: Jossey-Bass.

Scullen, S. E., Mount, M. K., & Goff, M. (2000).  Understanding the latent

    structure of job performance ratings. *Journal of Applied Psychology, 85*,

    956-970.

Smither, J. W., London, M., Reilly, R. R., Flautt, R., Vargas, Y., & Kucine, I.

    (2002).  Does discussing multisource feedback with raters enhance

    performance improvement?  Paper presented at the Seventeenth Annual

    Conference of the Society for Industrial/Organizational Psychology,

    Toronto, Canada.

Smither, J. W., London, M., & Richmond, K. R. (2005).  The relationship

    between leaders' personality and their reactions to and use of

    multisource feedback: A longitudinal study. *Group & Organization*

    *Management, 30*, 181-211.

Smither, J. W., London, M., Vasilipoulos, N. L., Reilly, R. R., Millsap, R. E., & Salvemini, N. (1995). An examination of the effects of an upward feedback program over time. *Personnel Psychology, 48*, 1-34.

Smither, J. W., Walker, A. G., & Yap, M. (2004). An examination of the equivalence of web-based versus paper-and-pencil upward feedback ratings: Rater-and-ratee-level analyses. *Education and Psychological Measurement, 64*, 40-61.

Stahl, M. J. (1983). Achievement, power, and managerial motivation: Selecting managerial talent with the job choice exercise. *Personnel Psychology, 36*, 775-789.

Taylor, S., & Brown, J. (1988). Illusion and well-being: A social psychological perspective on mental health. *Psychological Bulletin, 103*, 193-210.

Tekleab, A. G., Sims, H. P., Yun, S., Tesluk, P. E., & Cox, J. (2008). Are we on the same page? Effects of self-awareness of empowering and transformational leadership. *Journal of Leadership & Organizational Studies, 14*, 185-201.

Thornton, G. C. (1968). The relationship between supervisory and self-appraisals of executive performance. *Personnel Psychology, 21*, 441-456.

Tinsley, H. E. A., & Tinsley, D. J. (1987). Uses of factor analysis in counseling psychology research. *Journal of Counseling Psychology, 34*, 414-424.

Tornow, W. W. (1993a). Introduction to special issue on 360-degree feedback. *Human Resource Management, 32,* 211-219.

Tuckman, B., & Oliver, W. (1968). Effectiveness of feedback to teachers as a function of source. *Journal of Educational Psychology, 59,* 297-301.

Vandewalle, D. (1997). Development and validation of a work domain goal orientation instrument. *Educational and Psychological Measurement, 57,* 995-1015.

Vroom, V. H., & Yetton, P. W. (1973). *Leadership and decision making.* Pittsburgh: University of Pittsburgh Press.

Waldman, D. A., & Atwater, L. E. (1998). *The power of 360-degree feedback: How to leverage performance evaluations for top productivity.* Houston, TX: Gulf Publishing Company.

Wiese, D. S., & Buckley, M. R. (1998). The evolution of the performance appraisal process. *Journal of Management History, 4,* 233-249.

Yukl, G. (2006). *Leadership in Organizations.* Upper Saddle River, NJ: Pearson Prentice Hall.