

THE EYE OF THE BEHOLDER: A META-ANALYTIC EXAMINATION OF THE  
CONVERGENCE BETWEEN LEADER AND OBSERVER PERCEPTIONS OF  
LEADERSHIP

BY

ANGELA LEE

THESIS

Submitted in partial fulfillment of the requirements  
for the degree of Master of Arts in Psychology  
in the Graduate College of the  
University of Illinois at Urbana-Champaign, 2016

Urbana, Illinois

Master's Committee:

Assistant Professor Nichelle C. Carpenter, Chair  
Associate Professor Daniel A. Newman

## ABSTRACT

The convergence between a leader's assessment of his/her leadership behaviors and assessments from the leader's subordinates, peers, and superiors—also known as “leader insight”—is linked to important organizational and leader outcomes. Unfortunately, many questions remain regarding the extent to which leaders have insight into their leadership behaviors. This study examines whether leaders' perceptions of their leadership behaviors are similar to or different from observers' perceptions. Importantly, we investigate whether leader-observer agreement is influenced by type of observer and type of leadership. First, we meta-analyzed the relationship (i.e., correlation) between leader- and observer-ratings along several dimensions of leadership (e.g., initiating structure, consideration, transactional, and transformational leadership). We found that leader-observer agreement was moderate overall but was stronger for task-oriented leadership behaviors (e.g., transactional) than for relationship-oriented leadership behaviors (e.g., transformational). Our findings also demonstrated that a leader's subordinates, peers, and superiors had similar views of the leader's behaviors. To better understand leader-observer agreement, we also meta-analyzed the differences in leader and observer mean-level reporting (i.e., Cohen's *d*). We found that leaders generally under-reported task-oriented leadership but over-reported relationship-oriented behaviors relative to observers. Last, our results indicated that sampling method and scale measure moderated leader-observer convergence. Implications of these findings for research, theory, and practice are discussed.

## TABLE OF CONTENTS

CHAPTER 1 INTRODUCTION .....	1
1.1 Types of Leadership .....	3
1.1.1 Task-oriented Leadership .....	4
1.1.2 Relation-oriented Leadership .....	4
1.2 Leader- and Observer-Perceptions of Leadership Behaviors .....	5
1.3 Leader Over-Reporting or Under-Reporting Relative to Observers .....	8
1.4 Moderating Role of Leadership Type .....	9
1.5 Moderating Role of Type of Observer .....	10
CHAPTER 2 METHOD .....	13
2.1 Literature Search .....	13
2.2 Procedure .....	13
2.3 Meta-Analyses .....	15
CHAPTER 3 RESULTS .....	17
3.1 Correlations between Leader- and Observer-Ratings .....	17
3.2 Mean Differences between Leader- and Observer-Ratings .....	19
3.3 Moderator Analyses .....	21
CHAPTER 4 DISCUSSION .....	24
4.1 Implications for Theory and Practice .....	28
4.2 Limitations and Future Research Directions .....	30
TABLES AND FIGURES .....	34
REFERENCES .....	49

## CHAPTER 1 INTRODUCTION

Much leadership scholarship has focused on the extent to which leaders and their subordinates, peers, and superiors have converging perceptions of the leader's behavior, which has been referred to as *leader insight* (e.g., Atwater & Yammarino, 1992, 1997; Fleenor, Smither, Atwater, Braddy, & Sturm, 2010; Halverson, Tonidandel, Barlow, & Dipboye, 2005). Leader insight is a critical predictor of leader effectiveness as, for example, leaders who overestimate their leader behaviors relative to subordinates (i.e., low insight/agreement) may make ineffective decisions or decline training and development opportunities (Bass & Yammarino, 1991), whereas leaders with more insight into their good performance (e.g., high convergence) are more likely to be successful and respond appropriately to constructive feedback (Atwater & Yammarino, 1997). Thus, knowledge of leader-observer agreement is important for understanding and predicting outcomes for employees, managers, and organizations.

Unfortunately, despite the importance of understanding the relationship between leader- and observer-ratings of leadership, there are at least two critical gaps in the literature. First, extant research on leader-observer agreement typically focuses on only two perspectives: the leader and subordinate. However, given that leadership is a construct that encompasses the leader's interactions with numerous parties, such as subordinates, peers, and supervisors (Riggio, Chaleff, & Lipman-Bllumen, 2008), focusing exclusively on convergence between leaders and subordinates yields an incomplete picture. Indeed, observers may be influenced differently by particular leader characteristics and behaviors; thus, restricting observer-ratings to subordinate ratings alone likely provides a rather limited view of convergence (Hiller, DeChurch, Murase & Doty, 2011).

The second issue is that it is currently unknown whether leader-observer agreement is influenced by the type or conceptualization of leadership behavior being examined. Sin, Nahrgang, and Morgeson's (2009) influential meta-analytic review of the relationship between leader- and observer-ratings centered on leader-member exchange (LMX), but leadership is not a monolithic or one-dimensional construct. Rather, leadership has been defined and understood as a variety of behaviors and dimensions, including transformational and transactional leadership (Bass & Avolio, 1995), consideration and initiating structure (Bass, 1990; Fleishman, 1973; Judge, Piccolo, & Ilies, 2004); as well as ethical (Brown, Treviño, & Harrison, 2005), laissez-faire (Bass & Avolio, 1995), and servant (Greenleaf, 1977; Barbuto & Wheeler, 2006) leadership. Although there are similarities between some of these behaviors (Bass & Avolio, 1995; Brown et al., 2005; Treviño & Brown, 2004), there nevertheless exist important conceptual differences between leadership constructs that could influence the convergence between leaders and observers. Thus, investigations that fail to account for multiple forms of leadership provide an incomplete understanding of the extent to which leaders may have more convergence with observers on some behaviors relative to others.

The current study aims to address these gaps in the literature and make three contributions to research and theory on leader insight. First, we conduct a meta-analysis of the relationship (i.e., correlation) between leader-ratings and subordinate-, peer-, and superior-ratings ("observer-ratings") of leadership behaviors. Although recent qualitative reviews have provided important summaries of the numerous factors influencing the convergence of ratings of leadership (e.g., Fleenor et al., 2010), a quantitative summary is essential for understanding the precise levels of convergence between leader- and observer-ratings. In particular, we examine potential differences in convergence between leaders and their subordinates, peers, and superiors.

Second, we evaluate the extent to which the relationship between leader and observers depends on the type of leadership behavior. Finally, the third contribution of the current study is to evaluate leader-observer mean differences—assessing whether leaders under-report or over-report their leader behaviors relative to observers. Specifically, we meta-analyze the mean levels of leadership reported by each source and determine whether leaders report levels of behaviors that are similar to some observer roles but different from other observer roles. We also examine whether these mean-difference patterns are moderated by the type of leadership behavior. In sum, the current meta-analysis investigates the relationship between leader- and observer-ratings, while considering different types of observers, different types of leader behaviors, and both inter-rater correlations and mean-level consensus. It is our hope that this work will inform the selection of rater sources that may be most suitable for measuring particular dimensions of leadership.

### 1.1 Types of Leadership

The leadership domain has been plagued by construct proliferation (DeRue et al., 2011), and according to Bass (1990, p.11), “there are almost as many definitions of leadership as there are persons who have attempted to define this concept.” Nevertheless, scholars continue to demonstrate both conceptually and empirically that most dimensions can fall into two dimensions: task- or relation-oriented dimensions (see Fleishman, 1953; Humphrey, 2002; Judge, Piccolo, & Ilies, 2004). In fact, this two-factor model has garnered renewed popularity (Avolio, 2013; Judge et al., 2004) and is supported in other domains as well. For example, according to social and personality psychology theories, individuals are driven by two types of motivation, getting ahead or getting along (Hogan & Shelton, 1998), which can be translated into task-oriented or relation-oriented leader behavior. Research in managerial performance also advocates

a two-factor model: *task performance*, or structuring work and focusing on getting the work done, and *contextual performance*, or facilitating the psychological and social contexts of work and getting along with others. (Oh & Berry, 2009; Van Scotter & Motowidlo, 1996). Conceptualizing leadership into task- vs. relation-oriented behaviors not only captures a broad range of leadership concepts, but also provides a clear and parsimonious model for understanding leadership. Therefore, we utilize the two-factor model to guide the research aims in the present paper.

1.1.1 Task-oriented leadership. We define task-oriented leadership as behaviors that contribute to the completion of tasks by organizing and directing the work of others. The importance of task-oriented behaviors is evidenced by researchers focusing on production-centered (Blake, Mouton, & Bidwell, 1962; Judge et al., 2004) or task-oriented leader behaviors (Bales, 1950). An early conceptualization of leadership is *initiating structure*, behaviors designed to organize and structure group activities and reach task goals, such as assigning tasks to subordinates, or emphasizing the importance of meeting deadlines (Bass, 1990; Halpin & Winer, 1957; Stogdill, 1950). A relatively more contemporary example of task-oriented leadership is *transactional leadership*, which focuses on contingent rewards, in which specific behaviors are rewarded, as well as management by exception, in which the leader intervenes only when needed (Bass & Avolio, 1990).

1.1.2 Relation-oriented leadership. Relation-oriented leadership reflects behaviors that strive to maintain positive interpersonal interactions among group members. Empirical work emphasizing employee-centered leadership (Blake et al., 1962; Judge et al., 2004) and person-oriented leader behaviors (Bales, 1950) support the idea of relation-oriented leadership. Early leadership research focused on the *consideration* factor, or the leadership behaviors that show interpersonal warmth and sensitivity, open communication, and mutual trust and respect (Halpin

& Winer, 1957; Fleishman, 1973; Stogdill, 1950). Contemporary leadership scholars also examine *transformational leadership*, which is characterized by motivating followers beyond what is expected, by raising consciousness about the value of specific and idealized goals, transcending self-interest for the good of the organization, and addressing higher-level needs (Bass & Avolio, 1990). More recently, researchers have considered *servant leadership*, in which the servant leader's chief goal is to serve and meet the needs of others (Greenleaf, 1977; Russell & Gregory Stone, 2002; Walumba, Hartnell, & Oke, 2010), and *ethical leadership*, which refers to normatively appropriate conduct through personal actions and interpersonal relations, and the promotion of such conduct to followers through two-way communication, reinforcement, and decision-making (Brown et al., 2005).

## 1.2 Leader- and Observer-Perceptions of Leadership Behaviors

The process of providing leadership ratings is fundamentally a performance rating behavior, which means that well-known issues in the performance rating literature such as rater bias and error in self-ratings, superior-, and/or coworker-ratings of performance (e.g., Murphy & Cleveland, 1995) likely apply to a leader's self-ratings and ratings of the leader made by observers. For instance, the concerns about leader-observer agreement parallel concerns of agreement between employees and their superiors and coworkers as part of 360-degree performance rating systems (Faction & Craig, 2001; Murphy & Cleveland, 1995). Because leaders and observers may differ in their information about leadership behavior, as well as in their personal motives to accurately report leadership information, it is not immediately clear whether we should expect considerable agreement between leader- and observer-ratings of leadership. Therefore, we draw on important theories regarding the processes underlying leader-

and observer-ratings in order to provide a clear, integrative understanding of leader-observer agreement.

There are several factors that might lead to high convergence between leader- and observer-ratings. Observers who interact frequently with and work in close proximity to leaders are likely to demonstrate high agreement with leaders, due to enhanced opportunity to observe the rated behavior (cf. Rothstein, 1990). Compared with superiors, subordinates and peers may be more familiar with and have more knowledge of leader behaviors, given that: (a) subordinates are the direct recipients of leader behaviors, and (b) peers have the same status as leaders in the organizational hierarchy (e.g., similar roles), and often work physically nearby leaders. Furthermore, according to Funder's (1995; 2012) Realistic Accuracy Model (RAM), behaviors that are easy for observers to witness should lead to high leader-observer agreement, but behaviors that are difficult for observers to witness should result in divergent ratings. Finally, because many definitions of leadership exist, the relationship between leader- and observer-ratings should be higher if leaders and observers have similar conceptualizations of leadership than when they have different definitions of what constitutes leadership (e.g., Hooijberg & Choi, 2000). Indeed, scholars have demonstrated that even raters occupying different roles often hold similar understandings of performance (Fecteau & Craig, 2001; Maurer, Raju & Collins, 1998; Smither, London, & Reilly, 2005), which would contribute to high leader-observer agreement. Overall, there are several reasons to expect leader- and observer-ratings to converge due to multiple environmental and personal influences.

Nevertheless, leader- and observer-ratings may also diverge for various reasons. Information processing theory (Salancik & Pfeffer, 1978) suggests that fundamentally different processes underlie leader- versus observer-ratings, in that individuals are influenced by

contextual effects and the consequences of past choices. This implies that employees may be differentially influenced by situational factors and access distinct pieces of social information, which could result in lower rating convergence. Furthermore, leaders are expected to have the most information about their own behaviors (Chan, 2009; see Schwarz & Oyserman, 2001). In contrast, subordinates, peers, and superiors witness only a portion of the leader's full repertoire of behavior (e.g., Allen, Barnard, Rush, & Russell, 2000), and are more likely to only witness and remember the *results* of the leader's behaviors (DeNisi, Cafferty, & Meglino, 1984). Finally, different motivational forces may drive leader- versus observer-ratings. Leaders may be susceptible to self-enhancement bias and, subsequently, inflate their ratings in order to present themselves in a favorable light (Atwater, Ostroff, Yammarino, & Fleenor, 1998; Paulhus, 1986). Observers may be vulnerable to other biasing motives, such as trying to avoid punishment from leaders by providing high ratings of leadership, despite actual perceptions of lackluster performance. In summary, multiple theoretical perspectives also support the possibility of low agreement between leader- and observer-ratings of leadership.

Based on the rationales reviewed above, the extent to which leader- and observer-ratings of leadership should converge is currently unclear. However, the extent to which leader-ratings and observer-ratings agree has important practical and theoretical implications. If there is high convergence between leaders and observers, this means that both parties have similar perspectives of leadership behaviors, which further suggests that the rating sources could be used interchangeably. On the other hand, if leader- and observer-ratings diverge, this means that different raters have unique perspectives of leadership, and that each rating source may be only appropriate for measuring certain types of leadership, or perhaps that there is unique incremental value in the ratings from each of the different sources. Because ratings from leaders and

observers are likely to continue to be used in leadership research, leader development, as well as in leadership evaluation efforts, it is important to comprehensively understand the level of agreement that should be expected across rater roles and behavioral content domains.

We expect that leader- and observer-ratings will converge somewhat but that this convergence will be considerably less than unity (i.e., correlation will be less than 1.0; see Viswesvaran, Ones, & Schmidt, 1996). In interpreting the magnitudes of correlations, we use Bosco, Aguinis, Singh, Field, and Pierce's (2014) correlational benchmarks (specifically the medium range for all effect sizes), as well as Edwards and Berry's (2010) guidelines for forming hypotheses. Correlations greater than .26 would indicate strong agreement, correlations ranging between .09 and .26 would reflect moderate agreement, and correlations below .09 would demonstrate low agreement. Because observers (i.e., peers, subordinates, and superiors) may have different interpretations of the leader's behavior and also because they are not likely to witness all aspects of the leader's behavior, we expect that leader- and observer-ratings will each contain some unique information and, therefore, be only moderately related. Therefore, we expect to find moderate agreement between leader- and observer-perceptions of leadership.

*Hypothesis 1:* Leader- and observer-ratings of leadership behavior will be positively and moderately correlated.

### 1.3 Leader Over-Reporting or Under-Reporting Relative to Observers

We also examine the mean differences between leader- and observer-ratings of leadership to determine whether leaders (a) under-report or (b) over-report their leader behaviors, relative to observers. If leaders provide higher average ratings of their leader behaviors relative to observers, this suggests leaders may suffer from a "mirage," as they perceive themselves to have leadership qualities that may not really be there (or are not seen by observers). In this case, leaders may be

perceived as overestimating their abilities, which leads to over-reporting leader behavior (Burriss, Detert, Romney, 2013; Paulhus, 1986; Yammarino & Atwater, 1997). However, if leaders under-report their own leadership relative to observers, this suggests “blindness” on the part of leaders, as they do not see the leader qualities that observers see in them. These leaders may be overly modest and avoid taking credit for their work (Burriss et al., 2013; Paulhus, 1986; Yammarino & Atwater, 1997). Indeed, both of these cases are associated with poor individual and organizational outcomes (e.g., Fleenor et al., 2010). Therefore, the meta-analysis of mean differences provides a more precise diagnosis of leader insight into their behavior.

We generally expect leader-ratings to be inflated relative to observer-ratings, for two reasons. First, we expect that observers only witness a subset of the leader’s behaviors, and second, we expect that leaders, as self-raters, are at least somewhat motivated to present themselves favorably and as an effective leader, leading to inflation in leader-ratings. Using Bosco et al.’s (2014) correlational benchmarks, mean differences ( $d$ ) greater than .54 indicate large discrepancies, mean differences ranging from .18 to .54 indicate moderate discrepancies, and mean differences less than .18 indicate small discrepancies. Specifically, we expect that leaders will engage in moderate levels of over-reporting compared to observers.

*Hypothesis 2:* Leaders’ mean-level ratings of their leader behaviors will be greater than observers’ mean-level ratings.

#### 1.4 Moderating Role of Leadership Type

The level of agreement (i.e., correlation and mean differences) between leader- and observer-ratings of leadership likely depends on the type of leadership behavior assessed, because some forms of behavior may be more easily observed and subsequently reported than others. For example, task-oriented leadership includes behaviors such as clearly expressing

performance expectations to subordinates, and appropriately giving rewards for completion. Because task-oriented statements are typically “factual and neutral in emotional tone,” (Humphrey, 2002, p. 496), perceptions of task-oriented leadership may be less subjective and less vulnerable to bias, and thus rated more consistently by different rating sources. In contrast, relationship-oriented behaviors comprise behaviors that are more ambiguous and reflect the leader’s interpersonal relationships with organizational parties (Brown, 2003). Observers in different hierarchical levels have varying skill sets and experiences, which may influence their expectations and perceptions of relation-oriented behaviors. For example, subordinates tend to need guidance and may want to receive attention from leaders, and consequently could be more aware of the enactment of servant leadership. On the other hand, superiors tend to focus more on “getting ahead” or task-oriented behaviors (Conway, Lombardo & Sanders, 2001; Hogan & Shelton, 1998), and as a result may be less aware of servant leadership, as an example. As such, we expect that leader-observer agreement (i.e., correlations and mean differences) will be higher for task-oriented leader behaviors and lower for relation-oriented leader behaviors.

*Hypothesis 3(a, b, c):* Leader-observer agreement will be stronger for (a) task-oriented leadership and its dimensions, (b) initiating structure and (c) transactional leadership; in comparison to relation-oriented leadership dimensions (i.e., consideration, servant, ethical, and transformational).

### 1.5 Moderating Role of Type of Observer

We also expect differences between leader- and observer-perceptions (i.e., correlation and mean differences) to depend on the type of observer who is rating the leader. Specifically, the leader’s subordinates, peers, and superiors each have unique relationships with the leader, which may result in differences in information that observers have about leader behavior and

observers' motivation to provide an accurate rating. Also, empirical findings show that leaders behave differently with subordinates than with other individuals, such as peers and superiors (Yukl, 2010).

Considering that leader behavior involves and is mostly directed toward subordinates (Hansbrough et al., 2014), the modal source for ratings of leadership from observers is subordinate-ratings (Hiller et al., 2011). Indeed, subordinates have many chances to observe their leader in a leadership role (Conway et al., 2001) and as a result, should have a thorough understanding of relevant leader behaviors. Compared to the leader's peers and superiors, subordinates should have the most opportunity to report on outcomes such as motivation and direction provided by leaders, quality of relationship with the leader, and perceptions of many leader behaviors and styles (Hiller et al., 2011). Therefore, we expect that compared to other observers, subordinates will show the highest level of agreement with leaders.

Ratings of leadership from the leader's peers are underrepresented in the literature (Hiller et al., 2011). Relative to subordinates, the leader's peers likely have less opportunity to observe leader behavior (i.e., because they are themselves leaders), and consequently may have an incomplete sense of the leader's leadership behaviors. On the other hand, leader-peer agreement could be enhanced because peers may be more knowledgeable about characteristics of a leader's behavior related to alignment, strategy, positioning, and boundary spanning (Hiller et al., 2011). However, peers may not be motivated to provide accurate ratings, because they may be friends with leaders and may be biased by liking (Stang, 1973) or, alternatively, be influenced by the friendship bias in which peer-rating is just a "popularity contest" (Love, 1981). Therefore, though we generally expect leader-peer agreement to be weaker compared to leader-subordinate agreement, we acknowledge that motivational factors could increase leader-peer agreement.

Finally, the leader's superior likely has even fewer occasions to witness leader behavior, relative to subordinates and peers (Pollack & Pollack, 1996). Superiors are usually hierarchically and physically distant from leaders, which could reduce chances to observe and obtain comprehensive information about the leader's behavior (Pollack & Pollack, 1996). Superior-ratings also appear vulnerable to motivational biases. For example, a superior may inflate ratings of an incompetent leader if locating a successor is inconvenient, or alternatively, give low ratings for an adept leader in order to justify the leader's removal. In contrast to this view, leaders and superiors may alternatively demonstrate high agreement due to superiors' likely increased age and experience with leadership, as opposed to that of subordinates (Harris & Kuhnert, 2008). Nonetheless, we predict that of all the leader-observer relationships, the weakest will be leader-supervisor agreement. In summary, we hypothesize that leader-observer agreement (i.e., correlations, mean differences) is strongest for subordinates, followed by peers, then by superiors.

*Hypothesis 4:* Leader-observer agreement will be moderated by type of observer, such that leaders will have the strongest agreement with subordinates, followed by peers, followed by superiors.

## CHAPTER 2 METHOD

### 2.1 Literature Search

In order to identify relevant studies, we conducted a keyword search for papers through 2014 using the *PsycINFO* and *ProQuest Dissertations* databases. We also searched through the Society of Industrial Organizational Psychology (SIOP) and the Academy of Management (AOM) conferences from 2010 through 2014. As displayed in Table 1, keywords were leadership, consideration, contingent reward, ethical, initiating structure, laissez-faire, non-leadership, servant, transactional, transformational, 360, agree, agreement, consensus, convergence, disagree, disagreement, and self-other. We obtained 61 studies and 65 independent samples, of which there were 25 published and 36 unpublished studies.

### 2.2 Procedure

To be included, a study had to report either (a) a correlation between leader- and observer-ratings of a dimension of leadership behavior or (b) means and standard deviations for the ratings from both leaders and observers (to permit calculation of Cohen's *d*). Observer-ratings were obtained from peers, subordinates, or superiors. Leadership dimensions included consideration, ethical, initiating structure, servant, transactional, and transformational. We divided these six styles into two general categories of leader behavior: task-oriented (i.e., initiating structure, transactional) and relation-oriented (i.e., consideration, transformational, servant, ethical). Table 2 displays this categorization. Additional leadership dimensions that did not fit into these categories included "laissez-faire," defined as the absence of leadership, and "other," a miscellaneous category. Table 2 lists all the leadership styles that appeared in the studies and how they were assigned into our six categories. For studies that reported multiple effect sizes for a given category of leadership, linear composites were calculated to maintain

statistical independence (Ghiselli, Campbell, & Zedeck, 1981). For example, the dimensions *initiating structure* and *transactional leadership* were analyzed separately, but were also composited to form an overall effect size contributing to the *task-oriented leadership* category.

From the literature search and subsequent categorization, we obtained 21 studies that measured task-oriented leadership and 51 studies that measured relation-oriented leadership. Several studies measured more than one type of leadership dimension. Specifically, there were 8 studies that measured initiating structure, 15 transactional, 13 consideration, 3 ethical, 15 servant, and 34 measuring transformational leadership studies. Additionally, there were 8 studies that measured laissez-faire and 12 studies that measured “other” leadership styles (e.g., *style 2 coaching* and *style 4 delegating* [Leadership Behavior Analysis II, Blanchard, Hambleton, Zigmarmi, Forsyth, 1985], and *strategic leadership* [Leadership Effectiveness Analysis, Management Research Group, 1992]).

We coded the leader-observer correlations, means and standard deviations for each source, reliabilities of leadership dimensions, sample sizes, and type of observer. The first and second authors initially coded a subset of studies (15% of final sample), and any discrepancies were discussed until 100% agreement was reached. The remaining samples were coded by the first author. Next, we computed standardized mean differences from the means and standard deviations (i.e., Cohen’s *d*). For studies reporting multiple correlations for sub-facets of a leadership dimension, linear composites were calculated to estimate relationships to avoid violating statistical independence (Ghiselli et al., 1981). For example, the sub-facets of transactional leadership (a) *contingent reward*, (b) *management by exception – active*, and (c) *management by exception – passive* were used to calculate an overall composite for transactional leadership. Additionally, for studies reporting multiple correlations for more than one type of

observer, we also calculated linear composites (Ghiselli et al., 1981) to compute an effect size between leaders and observers. For example, subordinate- and superior-ratings were composited into an effect size for overall observer.

Finally, we coded several methodological variables we posited would serve as potential moderators of the leader-observer relationship. Specifically, we coded the: (a) country in which the study was conducted (i.e., US or non-US), (b) leader level (i.e., level of employment occupied by the leader in the overall organization; “upper” referred to high-level executives and “lower” referred to all other managers), (c) study purpose (i.e., leadership development program or not), (d) leadership scale (i.e., the name of the measure on which raters responded), (e) sampling method (i.e., how observers were selected to provide leadership ratings; “leader didn’t select” referred to observers selected by researchers or randomly selected and “leader selected” referred to observers selected by leaders), and (g) publication status of the study (i.e., published, unpublished). We note that the available information about leadership scale was sufficient only for transformational leadership (i.e.,  $k \geq 3$  for each group). We also attempted to code for additional moderators such as supervisory span (i.e., the number of subordinates per supervisor), time with leader, job tenure, and nomological network correlates (e.g., personality traits, job satisfaction, and attitudes toward feedback), but this information was inconsistently reported in primary studies.

### 2.3 Meta-Analyses

We used Schmidt and Hunter’s (2014; also Hunter & Schmidt, 2004) artifact distribution meta-analysis methods to estimate the mean correlations between leader- and observer-rated leadership and the variability of these estimated correlations. Artifact distribution methods were used because not all primary studies reported reliability information. To correct for unreliability,

we used the obtained studies to calculate average internal consistency reliabilities for leader- and observer-ratings of leadership, shown in Table 3.

First, we conducted two broad meta-analyses to examine correlations and mean differences between leader- and observer-perceptions. For both sets of analyses, we meta-analyzed the leader-observer relationship for task- and relation-oriented leadership. (Results regarding the dimensions in the additional categories, *laissez-faire* and *other*, are not discussed, but are presented in the tables). Next, we meta-analyzed leader-observer relationships for each dimension of task- and relation-oriented leadership. We examined relationships between leaders and all observers combined, and when possible we separated observers into subordinates, peers, and superiors. Lastly, we tested potential moderating effects of several variables. In order to test whether there were significant differences between task- and relation-oriented leadership, between different types of observer roles, and between moderator conditions, we used formulas from Raju and Brand (2003), such that  $z_s > 1.96$  are statistically significant at the significance level of  $p < .05$ .

## CHAPTER 3 RESULTS

### 3.1 Correlations between Leader- and Observer-Ratings

The first purpose of the current study was to meta-analytically evaluate the correlations between leader- and observer-perceptions of leadership, displayed in Table 4. Overall, our results indicated that leader- and observer-ratings of leadership behavior were positively and at least moderately correlated (with the exception of one correlation—*laissez-faire*), and were all significantly different from zero, supporting Hypothesis 1.

Second, we compared leader-observer agreement on task- versus relation-oriented leadership dimensions, with results displayed in Table 4. The corrected correlation between leader- and observer-ratings for task-oriented behavior ( $\rho = .27$ ) was similar in magnitude to, but statistically significantly lower than, the correlation for relation-oriented behavior ( $\rho = .32$ ,  $z = 2.35$ ). This suggests that overall leader-observer agreement for task-oriented leadership is slightly weaker than for relation-oriented behavior, which contradicts Hypothesis 3a. However, we note that the magnitude of this difference may not be practically meaningful. Nonetheless, both correlations fall above Bosco et al.'s (2014) benchmark for medium effect sizes.

Next, we examined the separate leadership dimensions. Surprisingly, the leader-observer correlation for initiating structure ( $\rho = .25$ ) was at least slightly weaker than the relation-oriented dimensions, including consideration ( $\rho = .40$ ), servant ( $\rho = .28$ ), and transformational ( $\rho = .26$ ) leadership, which did not support Hypothesis 3b. However, the leader-observer correlation for transactional leadership ( $\rho = .57$ ) was significantly stronger than each of the relation-oriented dimensions ( $zs > 1.96$ ), supporting Hypothesis 3c.

Next, we investigated whether leader-observer agreement was moderated by type of observer (i.e., subordinate, peer, superior), with results displayed in Table 5 and Figure 1. For task-oriented leadership, the corrected correlations between leader- and subordinate-ratings ( $\rho = .28$ ), peer-ratings ( $\rho = .26$ ), and superior-ratings ( $\rho = .24$ ) were not significantly different from each other ( $z_s < 1.96$ ). Similarly, for initiating structure, correlations for leader-subordinate ( $\rho = .20$ ), leader-peer ( $\rho = .18$ ), and leader-superior ( $\rho = .17$ ) were similar in magnitude ( $z_s < 1.96$ ). (There were not enough primary studies to individually examine each observer role for transactional leadership).

Likewise, for relation-oriented leadership, the corrected correlations between leader- and subordinate-ratings ( $\rho = .30$ ), peer-ratings ( $\rho = .29$ ), and superior-ratings ( $\rho = .27$ ) were not significantly different from each other ( $z_s < 1.96$ ). However, transformational leadership was an exception to this pattern as the leader-subordinate correlation ( $\rho = .45$ ), was significantly larger than the leader-peer correlation ( $\rho = .27$ ,  $z = 5.26$ ), and leader-superior correlation ( $\rho = .35$ ,  $z = 3.06$ ), and leader-superior agreement was significantly higher than leader-peer agreement ( $z = 2.05$ ). For consideration, the corrected correlations between leader- and subordinate-ratings ( $\rho = .33$ ), peer-ratings ( $\rho = .33$ ), and superior-ratings ( $\rho = .33$ ) were the same ( $z_s = 0$ ). For servant leadership, leader-subordinate agreement ( $\rho = .31$ ), leader-peer agreement ( $\rho = .35$ ), and leader-superior agreement ( $\rho = .32$ ) were not significantly different ( $z_s < 1.96$ ). Contrary to our expectations, with the exception of high leader-subordinate agreement in transformational leadership, type of observer did not influence the relationship between leader- and observer-ratings, providing minimal support for Hypothesis 4.

### 3.2 Mean Differences between Leader- and Observer-Ratings

The second purpose of this study was to investigate whether the moderate magnitude of leader-observer correlations is due to the extent to which leaders tend to engage in over-reporting or under-reporting relative to observers. The meta-analytic estimates of mean differences between leader- and observer-rated leadership are displayed in Table 6. Positive values indicate leader over-reporting (i.e., mean leader-ratings are higher than mean observer-ratings), whereas negative values indicate leader under-reporting (i.e., mean leader-ratings are lower than mean observer-ratings). Compared to all observers, leaders under-reported task-oriented ( $\delta = -.09$ ) and over-reported relation-oriented leadership ( $\delta = .10, z = 4.26$ ), although we note that the magnitudes of mean differences are small. This pattern generally held when specific leadership dimensions were considered. In particular, although the effect size is small, leaders under-reported initiating structure ( $\delta = -.13$ ) but reported similar mean levels of transactional leadership ( $\delta = .06$ ), relative to observers.

We found much larger effect sizes for the relation-oriented dimensions, which indicates bigger differences in mean-level reporting. Specifically, leaders over-reported servant ( $\delta = .32$ ) and transformational leadership ( $\delta = .28$ ), but under-reported consideration ( $\delta = -.23$ ). Our results partially support Hypothesis 2 in that, relative to observers, leaders over-reported only relation-oriented leadership but reported equal or lower mean levels of task-oriented leadership.

Next, we evaluated results for subordinates, peers, and superiors separately, and compared their mean-level ratings with leader's mean ratings of task- and relation-oriented leadership, displayed in Table 7 and Figure 2. For task-oriented leadership, the leader-superior mean difference ( $\delta = .03$ ; 95% CI: [-.05, .11]) was significantly smaller than both the mean differences for leader-subordinate ( $\delta = -.23, z = -4.47$ ) and leader-peer ( $\delta = -.27, z = -4.83$ ), and

the latter two mean differences were not significantly different from each other ( $z = .63$ ). This means that leaders and superiors reported similar levels of task-oriented leadership, whereas subordinates and peers provided higher mean levels of task behaviors than leaders. Similarly, when examining the specific dimensions, for initiating structure, the leader-superior mean difference ( $\delta = .04$ ; 95% CI:  $[-.03, .12]$ ) was significantly smaller than both the leader-subordinate ( $\delta = -.27$ ,  $z = -5.60$ ) and leader-peer ( $\delta = -.23$ ,  $z = -4.87$ ) mean differences, and the latter two mean differences were not significantly different from each other ( $z = .59$ ). However, for transactional leadership, the leader-subordinate mean difference ( $\delta = .07$ ) indicated equal reporting compared to the leader-superior mean difference ( $\delta = -.20$ ,  $z = 2.29$ ), which indicated superiors over-reported transactional leadership. Thus, leaders reported either similar or lower mean levels of task-oriented behaviors relative to observers, although the patterns depend on the behavior.

For relation-oriented leadership, the mean difference between leaders and subordinates ( $\delta = .05$ ; 95% CI:  $[0, .10]$ ) indicated nearly equal mean levels of reporting, whereas peers ( $\delta = -.18$ ,  $z = -2.02$ ) and superiors ( $\delta = -.09$ ,  $z = 2.57$ ) had more inflated ratings relative to leaders. For the consideration dimension, the leader-subordinate mean difference ( $\delta = -.36$ ) was significantly larger than the leader-peer difference ( $\delta = -.23$ ,  $z = 2.14$ ), which in turn, was larger than the leader-superior difference ( $\delta = -.05$ ; 95% CI:  $[-.13, .02]$ ,  $z = -3.15$ ).

For servant leadership, the difference between leader and subordinate average perceptions ( $\delta = .47$ ) was significantly larger than the difference between leaders and peers ( $\delta = .08$ , 95% CI:  $[-.04, .20]$ ,  $z = -4.60$ ) and leaders and superiors ( $\delta = .14$ , 95% CI:  $[0, .29]$ ,  $z = 3.42$ ). Our results provide initial evidence that leaders reported inflated levels of ethical leadership compared to subordinates ( $\delta = 1.29$ ). For transformational leadership, leaders inflated

their ratings relative to subordinates ( $\delta = .24$ ), but, interestingly, leaders under-reported transformational leadership relative to both peers ( $\delta = -.14$ ,  $z = -5.89$ ) and superiors ( $\delta = -.22$ ,  $z = 7.44$ ) (leader-peer and leader-superior relationships were not significantly different,  $z = 1.08$ ). Thus, leaders were likely to over-report their levels of servant leadership, transformational (only relative to subordinates) leadership, and ethical leadership behaviors; however, for the other behaviors (i.e., task-oriented and transformational relative to peers and superiors), leaders reported nearly equal or lower levels of behaviors relative to observers.

### 3.3 Moderator Analyses

Our results revealed multiple instances of meta-analytic relationships in which the credibility interval included zero along with a small percentage of variance explained, which justifies the testing of moderating variables (Schmidt & Hunter, 2014). Therefore, we examined six variables we hypothesized could act as potential moderators of the relationship between leader- and observer-perceptions. We conducted moderator analyses on both the correlations and mean differences between leader- and observer-ratings of task- and relation-oriented leadership.

Moderator results for the corrected correlations between leader- and observer-ratings are presented in Table 8 and Figure 3. In general, we found that the leader-observer correlations were not strongly influenced by moderators. We did find evidence that leader-observer correlations were somewhat dependent on the purpose of the study. Specifically, for task-oriented behaviors, the leader-observer correlation was slightly weaker when ratings were collected for leader development purposes ( $\rho = .22$ ) than for other purposes ( $\rho = .29$ ,  $z = -2.02$ ). For relation-oriented behaviors, we found the opposite pattern, though the differences are rather small for both sets of behaviors. . We also found that for task-oriented behaviors (but not relation-oriented), the leader-observer correlation was significantly stronger in US samples ( $\rho$

= .37) than in non-US samples ( $\rho = .14, z = 6.82$ ). We also found that publication status was a moderator of task-oriented relationships, as published studies ( $\rho = .42$ ) had stronger correlations than unpublished studies ( $\rho = .19, z = 6.23$ ). However, factors such as the level of the leader and the sampling method did not moderate leader-observer correlations.

Table 9 and Figure 4 display the moderator results for the corrected mean differences between leader- and observer-ratings. First, country moderated the magnitude of leader-observer mean differences such that observers over-reported (relative to leader) task- ( $\delta = -.31$ ) and relation-oriented ( $\delta = -.15$ ) behaviors in non-US countries, whereas leaders over-reported (relative to observers) relation-oriented behaviors ( $\delta = .22$ ) and reported more equal levels of task-oriented behaviors ( $\delta = .03$ ) in the U.S.

Second, study purpose was also a moderator, as leaders appear to be more truthful (i.e., less over-reporting) in their ratings for developmental purposes. Specifically, leaders under-reported (relative to observers) both their task-oriented ( $\delta = -.29$ ) and relation-oriented behaviors ( $\delta = -.21$ ) for leader development purposes; however, for non-developmental purposes, leaders inflated their ratings (relative to observers) of task-oriented ( $\delta = .17$ ) and relation-oriented behaviors ( $\delta = .24$ ).

Third, the sampling method, or the way in which observer-raters are selected was an important moderator of leader-observer mean differences. Specifically, when the leader selects the observer that will rate the leader's behavior, the observer inflated their ratings (relative to the leader) of both task-oriented ( $\delta = -.20$ ) and relation-oriented ( $\delta = -.13$ ) behaviors. However, when the leader did not select the observer, the leader inflated their ratings of relation-oriented behavior ( $\delta = .73$ ) and reported nearly equal levels of task-oriented behaviors ( $\delta = .14$ ), relative to observers.

Fourth, we found that the transformational leadership scale influenced leader-observer mean differences, such that leader ratings were significantly more inflated on the Multifactor Leadership Questionnaire (MLQ; Bass & Avolio, 1990;  $\delta = .77$ ) versus the Leadership Practices Inventory (LPI; Kouzes & Posner, 2002;  $\delta = .46$ ,  $z = -2.75$ ).

## CHAPTER 4 DISCUSSION

Despite extensive primary research on the factors influencing the agreement between leader- and observer-perceptions of leader behavior (i.e., leader insight), four critical gaps regarding the relationship between leader- and observer-ratings remained. First, there was no comprehensive understanding of the actual convergence between leader-perceptions and observer-perceptions of leadership behaviors. Second, it was unclear whether leader-observer agreement depended on the type of observer (i.e., subordinate, peer, superior). Third, it was also uncertain how the type of leadership dimension influenced the level of agreement between leader-ratings and observer-ratings. Finally, there was no cumulative understanding of how important factors regarding the rating context or process influenced leader-observer relationships. The present study filled in these gaps, while also providing several new insights.

The relationships (i.e., correlation and mean differences) between leader- and observer-ratings of behaviors in the multidimensional leadership domain had not been meta-analytically estimated until now. We found positive and moderate corrected correlations in the .30 range across the different leadership dimensions. Interestingly, this is consistent with existing meta-analytic estimates of leader-subordinate convergence for LMX (Sin et al., 2009;  $\rho = .37$ ), as well as with estimates of self-observer agreement for employee work behaviors, such as performance ( $\rho$ 's ranging from .35 – .36; Harris & Schaubroeck, 1988), organizational citizenship behavior ( $\rho = .26$ ; Carpenter, Berry, & Houston, 2014), and counterproductive work behavior ( $\rho = .38$ ; Berry, Carpenter, & Barratt, 2012). Thus, an important question arising from these parallel findings for leadership is whether correlations in the .30 range are simply what should be expected as metrics of inter-rater reliability regardless of construct, or if there exist methodological factors that can drastically increase the convergence between rating sources.

Our findings for the mean differences between leader- and observer-ratings were surprising – leaders (i.e., self-raters) did not always over-report their leadership behaviors relative to observers. Although leaders did provide inflated ratings of some types of relation-oriented behaviors, they reported ratings for task-oriented behaviors that were equal to or even lower than the ratings provided by observers. Thus, our findings suggest that leader-raters do not provide substantially inflated ratings of their leadership for all behaviors, which further indicates that self-enhancement bias (Atwater et al., 1998; Paulhus, 1986) may not be a big concern as assumed, at least for some of the leadership dimensions.

Another insight from our findings is that the type of leadership dimension matters for understanding the convergence between leader- and observer-perceptions of leadership. Contrary to prior work, our findings showed that for relation-oriented dimensions such as transformational, servant, and ethical leadership, leaders may be vulnerable to a social desirability bias (see Densten & Sarro, 2012; Lievens, Van Geit, & Coetsier, 1997) given that these were the only dimensions on which leaders inflated their ratings relative to observers (subordinates, specifically, for transformational leadership). However, social desirability is not likely to be a factor influencing leader-ratings across other dimensions, as leaders actually reported equal or *lower* mean levels (e.g., initiating structure, consideration) relative to observers. This means that for some dimensions, observer-raters may actually be the rating source that is relatively more vulnerable to biases (e.g., halo) that cause them to provide inflated ratings relative to the leader. In any case, our findings demonstrate that it is important to consider the leadership dimension of interest when interpreting the relationships between leader- and observer-ratings.

Fourth, our findings confirm that the type of observer is essential to understanding patterns of rater underreporting and overreporting. Specifically, the patterns of observer-ratings

were not entirely redundant across source. Superiors and peers provided inflated ratings of transformational leadership relative to leaders, whereas subordinates rated leaders lower on transformational leadership than what leaders reported themselves. This suggests that subordinates likely have perceptions and interpretations of transformational leadership behaviors that may largely differ from those of the leader and the leader's peers and superiors. Similarly, for both consideration and initiating structure, subordinates and peers provided higher mean ratings relative to leaders, but superior-ratings were equal to leader-ratings on both dimensions. This implies that observers with different positions relative to the leader are likely to have unique perspectives of the leader's behavior – importantly, our findings show that the leader-observer correlation is rather similar across sources but that patterns of leader underreporting and overreporting depend on the observer type.

Our examination of moderators of leader-observer correlations and mean differences provided several insights into how contextual and sampling choices influence agreement. For example, obtaining leader- and observer-ratings in the US versus other (non-US) countries influenced agreement such that leader-observer correlations on task-oriented behaviors were significantly stronger in the US, observers over-reported both task- and relation-oriented behaviors relative to leaders in non US countries, and leaders over-reported (or provided equal ratings) their behavior in the US. This suggests the need to examine further how cultural contexts and characteristics influence rating patterns, as this is likely to reveal critical guidelines for interpreting ratings in different cultural contexts. These findings also suggest the need to evaluate the measurement invariance of the different leadership dimension operationalizations across different cultural and national contexts.

Another noteworthy insight pertained to the moderating role of the study purpose. Specifically, we found that leaders underreported (i.e., observers overreported) their behaviors when ratings were obtained for leadership development purposes, but leaders overreported (i.e., observers underreported) for all other purposes. Thus, leaders may provide more honest ratings for developmental purposes, but be influenced by self-enhancement concerns for other purposes (e.g., Murphy & Cleveland, 1995). It is possible that leaders attending development programs gain greater awareness of or are more honest about their shortcomings, because they expect to be called upon to discuss the discrepancies between their own and others' views. It may even be the case that leaders deliberately provide lower ratings for themselves in developmental settings, to create the appearance of humility rather than bravado. In any case, it is important to note the context in which ratings were provided, particularly for evaluating leader ratings.

Finally, we found that the manner in which observers were selected moderated leader-observer agreement. When leaders selected the observer-rater, the observer-rater provided higher mean ratings relative to the observer; however, when the leader did not designate the observer-rater (e.g., the researchers randomly selected observer-raters), the observer-rater provided lower (or equal) mean ratings than the leader. On one hand, this could mean that leaders may deliberately pick observers who possess a lot of relevant information about the leader's behavior, which may explain observer over-reporting. It is also plausible that the leader selects observers with whom there may be a friendship, indicating that the observer-ratings in this approach may be inflated due to interpersonal liking. However, it also seems likely that observer-raters who are selected by the leader may provide inflated ratings because they feel obligated to present their leader in a good light or because ratings are not entirely anonymous. Thus, it is important to examine the way in which observers were invited to rate their leaders, as there are clear effects

on leader-observer agreement and over-reporting bias. Interestingly, we note in Sin et al.'s (2009) meta-analysis on LMX agreement, this moderator did not influence the leader-observer correlation.

#### 4.1 Implications for Theory and Practice

Our findings for leader-observer correlations indicate that both leader-ratings and observer-ratings are important measures of leadership behaviors. Additionally, our results of leader-observer mean differences provide important guidelines on how ratings from leaders and observers should be obtained and interpreted in practice. As noted above, we found that leaders may show a “mirage” bias for ethical, servant, and transformational leadership, given that leaders inflated their ratings of these behaviors compared to what observers reported. This suggests that it is likely inappropriate to rely solely on leaders' ratings of these behaviors. For example, considering only the leader's (inflated) perception of their ethical leadership behaviors is potentially legally and practically dangerous, and researchers and practitioners should obtain subordinates' and other observers' perspectives of the extent to which the leader displays desired ethical behaviors. However, for task-oriented leadership behaviors as well as consideration, leader-ratings were not typically inflated relative to observers, indicating that relying on leader ratings alone is likely suitable for these behaviors.

Our study also has important implications for tests of theories pertaining to the leadership dimensions. When a leadership dimension is an outcome of interest, it is crucial that researchers consider whose leadership perspective is most relevant – our findings demonstrate that rating sources are not necessarily interchangeable. In particular, leaders' self-ratings or different types of observer-ratings may be inflated or deflated depending on the dimension of interest. The

inflation or deflation from either rating source could produce misleading conclusions regarding the antecedents and consequences of different leadership dimensions.

Thus, it appears that leadership dimensions may vary in the extent to which they are adequately represented by leader-ratings versus observer-ratings. As a result, researchers must decide whose leadership perspective is important for the study, as different rating sources may influence empirical relationships tied to hypotheses. For example, if theoretical expectations involve transformational leadership, it is important to consider that ratings provided by leaders are likely to overestimate leadership relative to subordinates, though the leader's ratings may also be similar to peers and superiors. Although this may indicate that leader-, peer-, and superior-ratings may contain abundant information about transformational leadership (relative to subordinates), it also could mean that leaders, peers, and superiors are more lenient judges of transformational leadership, while subordinates are harsher. Thus, it may also be that subordinate ratings represent more "honest", conservative measures of transformational leadership relative to other sources. Either way, our findings show that researchers and practitioners should be aware that for different leadership dimensions, ratings from different sources have different patterns and empirical consequences.

Our findings have implications for several leadership theoretical models, including transactional models (Hollander, 1992) and the multilevel framework of transformational leadership (Avolio & Bass, 1995). For example, transactional models of leadership, which focus on social exchange between leaders and subordinates, also recognize leadership as a two-way process. Thus, it is not surprising that both the leader and subordinate's perspective of leadership are considered in such reciprocal models. However, our findings suggest that the differences between leader and subordinate ratings may not translate to broken exchange patterns, but rather,

may be due to respective rater errors and/or biases. Thus, it is imperative that researchers and practitioners who are interested in both leader and subordinate (or other observer) perceptions of phenomena consider the “typical” meta-analytic convergence between the two sources shown here, as this provides a foundation for understanding and interpreting any differences between leader and observer perceptions.

Additionally, it is important to note that the multilevel theory regarding the diffusion of transformational leadership across the individual, group, and organizational levels of analysis (Avolio & Bass, 1995) does *not* specify whose perspective of leadership is important for diffusion. However, our findings suggest that this is a question that needs to be empirically addressed. Although ratings from different sources may provide important and potentially valid information about transformational leadership, for example, it cannot be ignored that the sources are also vulnerable to different biases that may influence the understanding of leadership spread across levels of analysis.

#### 4.2 Limitations and Future Research Directions

Although the present study makes a contribution to leadership practice and theory, there are some limitations that provide directions for future research. Some of the moderator relationships we evaluated in this study were based on a small number of samples, which means that these relationships should be interpreted with caution. In particular, studies that obtained observer-ratings from peers and superiors were limited for certain leadership dimensions (e.g., consideration, initiating structure, servant leadership), which reveals an important need for future research. Although these parties may have different observations of the leader’s behavior, these perspectives are still important for obtaining a comprehensive understanding of the leader. As such, more research that includes multiple raters is likely to be important.

Despite this meta-analysis being comprehensive in focusing on multiple leadership dimensions and different rating sources, a consequence was that there were some dimensions for which we were unable to obtain enough samples for a meta-analysis. In particular, for ethical leadership, there were too few primary studies (i.e.,  $k < 3$ ) to estimate the meta-analytic correlation between leader- and observer-ratings (although we were able to estimate the leader-observer mean difference). Also, there were not enough samples to separately meta-analyze mean differences between leaders and each type of observer (e.g., leader-peer, leader-superior) for ethical leadership. Moreover, we were unable to examine other important leadership dimensions such as shared leadership (Pearce & Conger, 2002) and dimensions from a four-factor model consisting of supportive, directive, participative, and achievement-oriented leadership (House, 1977; Sims & Manz, 1996). This indicates that additional research is needed that provides the multisource correlations for these dimensions.

It is important to note that many primary studies obtained multiple ratings for a given observer (e.g., five of a leader's subordinates provided ratings). This means that in a primary study, observer-ratings may have been aggregated prior to evaluating the leader-observer relationship. There are many ways to aggregate multisource ratings (e.g., Arensberg, Schiller, Vivian, Johnson, & Strasser, 1996; Beck, 2014), but primary studies did not consistently report how ratings were combined. Additionally, interrater agreement should be examined before proceeding with aggregation, but primary studies did not always report agreement indices to justify aggregation. Therefore, an important question we were unable to answer in this study is whether the aggregation method or level of agreement in observer-ratings influenced the relationship between leader-ratings and observer-ratings. We recommend that future researchers or practitioners who aggregate multiple observer-ratings (e.g., from subordinates) report how

aggregation was conducted and report agreement indices, such as intraclass correlation coefficients (ICC; e.g., Atwater et al., 2005),  $r_{wg}$  (e.g., Berson & Sosik, 2007), or  $r^*_{wg(j)}$  (e.g., Braddy, Gooty, Fleenor, & Yammarino, 2014).

Although the current study provided important information about the level of convergence between leader and observer ratings of different leadership dimensions, our findings do not show the extent to which leader and observer ratings are valid measures of leadership. One way to disentangle this is to use multitrait-multimethod (MTMM; Campbell & Fiske, 1959) analysis to evaluate whether leader- and observer-ratings are equivalent measures of the leadership dimensions. Specifically, if leader-ratings and observer-ratings reflect more trait (i.e., leadership dimension) variance than method (i.e., rater-specific) variance, this provides important evidence that ratings from both leaders and observers are justified. However, it is also plausible that not all rating sources will reflect more trait variance than method variance, which will reveal whether some rating sources are more contaminated by errors and bias, relative to others sources. As a result, future research that further evaluates different rating sources for measuring leadership will be imperative for better understanding their relative strengths and weaknesses.

Finally, an essential evaluation of the uniqueness and validity of different rating sources should include an empirical examination of whether nomological correlations depend on the source that is used to measure leadership. For example, self- and observer-perceptions of leadership could be differentially related to variables such as personality traits or job attitudes. If these relationships have similar patterns and magnitudes regardless of whether leadership is measured via leader-, subordinate-, peer-, or superior-ratings, this would suggest that despite the different leader-observer relationships (i.e., method variance), different rating sources could

appropriately be viewed as representing the same construct. On the other hand, if there were significant differences in the relationships between transformational leadership and other correlates based on leadership rating source, this would indicate that the rating sources were not interchangeable and perhaps did not represent the same construct. We were unable to locate enough samples with the same set of common correlates across rating sources, which prevented us from examining the nomological relationships. However, this suggests a need for researchers to measure (and report in text) theoretically relevant correlates of different leadership dimensions so that enough studies can accumulate to conduct this analysis.

Although leadership dimensions are often measured with ratings from both leaders and their observers, the extent to which these different perspectives of leadership converged was previously unclear. We conducted the first meta-analysis of leader-observer relationships, and provided important insights about the measurement of leadership dimensions. Specifically, leader-observer correlations are generally of similar magnitude regardless of the type of observer-rater or leadership dimension. However, these factors do affect whether leaders inflate or underreport their leadership behaviors relative to observers, as do important contextual variables. This study serves as an important call to first consider whose perspective of leadership is relevant (i.e., leader, subordinate, superior, peer), and then to understand how such choices may affect the understanding or interpretation of leadership behaviors.

## TABLES AND FIGURES

Table 1

*List of Search Terms Used in the Literature Search for Primary Studies*

---

Search terms	
leadership	transformational
consideration	360
contingent reward	agree
ethical	agreement
initiating structure	consensus
laissez-faire	convergence
non-leadership	disagree
servant	disagreement
transactional	self-other

---

Table 2

*Leadership Styles Sorted into Task- or Relation-Oriented Behaviors*

<b>Task or Relation</b>	<b>Style</b>	<b>Definition</b>
task	initiating structure	organize and structure group activities to achieve task goals
task	transactional	based on the exchange process in which the leader provides rewards in return for the subordinate's effort
relation	consideration	emphasizes interpersonal warmth and sensitivity, open communication, and mutual trust and respect
relation	transformational	motivates subordinates to move beyond self-interests through inspiration, charisma, and intellectual stimulation
relation	servant	places needs of subordinates before own and helps subordinates reach their maximum potential
relation	ethical	displays ethical behavior through personal actions and interpersonal relationships

Table 3  
*Reliability (Alpha) Artifact Distributions*

<b>Leadership Dimension</b>	<i>r<sub>xx</sub></i>	<i>SD</i>	<i>N</i>	<i>k</i>
Relation-oriented (self-rated)	.81	.11	9,991	46
Relation-oriented (observer-rated)	.83	.12	9,991	46
Task-oriented (self-rated)	.78	.12	4,664	17
Task-oriented (observer-rated)	.80	.13	4,664	17
Consideration (self-rated)	.79	.05	3,833	7
Consideration (observer-rated)	.84	.09	4,680	8
Ethical (self-rated)	.86	.06	270	3
Ethical (observer-rated)	.92	.02	270	3
Initiating Structure (self-rated)	.78	.07	3,833	7
Initiating Structure (observer-rated)	.82	.11	3,833	7
Laissez-faire (self-rated)	.75	.11	465	5
Laissez-faire (observer-rated)	.74	.11	465	5
Other (self-rated)	.81	.14	10,710	25
Other (observer-rated)	.83	.14	10,710	25
Servant (self-rated)	.86	.10	1,817	14
Servant (observer-rated)	.87	.09	1,817	14
Transactional (self-rated)	.71	.09	831	10
Transactional (observer-rated)	.72	.12	831	10
Transformational (self-rated)	.85	.09	3,224	21
Transformational (observer-rated)	.87	.09	3,224	21
Overall (self-rated)	.81	.11	25,830	93
Overall (observer-rated)	.83	.12	25,830	93

Note. All artifact distributions were calculated in the present study. *r<sub>xx</sub>* = reliability artifact distribution mean; *SD* = reliability artifact distribution standard deviation; *N* = reliability artifact distribution sample size; *k* = number of samples contributing to artifact distributions.

Table 4

*Meta-analytic Results: Leader-Observer Correlations*

<b>Leadership Style</b>	<b>N</b>	<b><i>k</i></b>	<b><i>r<sub>m</sub></i></b>	<b><i>SD<sub>r</sub></i></b>	<b><math>\rho</math></b>	<b><i>SD<sub><math>\rho</math></sub></i></b>	<b>% Var</b>	<b><i>CV<sub>10</sub></i></b>	<b><i>CV<sub>90</sub></i></b>	<b><i>CI<sub>L</sub></i></b>	<b><i>CI<sub>U</sub></i></b>	<b>Z test</b>
Relation-Oriented <sup>b</sup>	6,768	33	.27	.15	.32	.16	18.25	.11	.52	.26	.38	<sup>ab</sup> -2.27*
Transformational	3,201	24	.23	.21	.26	.23	14.86	-.02	.55	.17	.36	
Consideration	4,859	9	.33	.12	.40	.14	9.95	.23	.57	.31	.50	
Servant	1,974	8	.23	.09	.28	.08	42.92	.18	.39	.21	.36	
Task-Oriented <sup>a</sup>	5,400	15	.20	.15	.27	.19	11.20	.03	.51	.17	.37	<sup>ca</sup> -2.32*
Initiating Structure	4,940	10	.21	.16	.25	.19	7.14	.01	.49	.13	.37	
Transactional	1,431	7	.41	.19	.57	.26	9.04	.24	.90	.37	.77	
Additional dimensions												
Other	12,252	14	.26	.11	.34	.13	8.50	.17	.51	.27	.41	
Laissez-faire <sup>c</sup>	420	3	.09	.14	.14	.16	38.47	-.07	.34	-.09	.37	<sup>cb</sup> -2.80*

*Note.* Ratings from subordinates, peers, and superiors were combined into overall observer-ratings.  $r_m$  = mean sample size-weighted correlation;  $SD_r$  = sample size-weighted observed standard deviation of correlations;  $\rho$  = mean sample size-weighted correlation corrected for internal consistency reliability;  $SD_\rho$  = corrected standard deviation of corrected correlations; % var. = percentage of variance attributable to statistical artifacts;  $CV_{10}$  and  $CV_{90}$  = 10% and 90% credibility values, respectively;  $CI_L$  and  $CI_U$  = lower and upper bounds, respectively, of the 95% confidence interval around the corrected mean correlation.

Table 5

*Meta-analytic Results: Leader-Observer Correlations (separated by type of observer)*

<b>Leadership Style</b>	<b>N</b>	<b>k</b>	<b><math>r_m</math></b>	<b><math>SD_r</math></b>	<b><math>\rho</math></b>	<b><math>SD_\rho</math></b>	<b>% Var</b>	<b>CV<sub>10</sub></b>	<b>CV<sub>90</sub></b>	<b>CI<sub>L</sub></b>	<b>CI<sub>U</sub></b>	<b>Z test</b>
Relation-Oriented												
Subordinate <sup>a</sup>	6,127	29	.26	.15	.30	.15	19.50	.11	.50	.24	.36	<sup>ba</sup> .46
Peer <sup>b</sup>	4,343	6	.25	.08	.29	.08	21.20	.19	.39	.22	.36	<sup>bc</sup> .86
Superior <sup>c</sup>	4,813	9	.23	.09	.27	.09	23.25	.16	.38	.20	.33	<sup>ac</sup> 1.41
Transformational												
Subordinate <sup>a</sup>	2,688	19	.39	.31	.45	.35	5.33	.00	.89	.28	.61	<sup>ba</sup> -5.26*
Peer <sup>b</sup>	1,490	3	.22	.20	.27	.24	4.53	-.03	.58	-.01	.55	<sup>bc</sup> -2.05*
Superior <sup>c</sup>	1,564	5	.30	.28	.35	.32	3.34	-.06	.75	.06	.63	<sup>ac</sup> 3.06*
Consideration												
Subordinate <sup>a</sup>	4,463	8	.27	.12	.33	.13	10.81	.16	.50	.23	.43	<sup>ba</sup> 0
Peer <sup>b</sup>	4,078	4	.28	.11	.33	.12	7.34	.18	.49	.21	.46	<sup>bc</sup> 0
Superior <sup>c</sup>	4,474	5	.27	.11	.33	.12	8.69	.17	.48	.21	.44	<sup>ac</sup> 0
Servant												
Subordinate <sup>a</sup>	1,974	8	.26	.12	.31	.13	23.18	.15	.48	.21	.42	<sup>ba</sup> 1.05
Peer <sup>b</sup>	1,515	3	.29	.13	.35	.14	10.18	.17	.54	.18	.53	<sup>bc</sup> .74
Superior <sup>c</sup>	1,515	3	.27	.11	.32	.13	12.98	.16	.48	.16	.48	<sup>ac</sup> -.26
Task-Oriented												
Subordinate <sup>a</sup>	4,769	12	.21	.18	.28	.23	6.91	-.02	.58	.14	.42	<sup>ba</sup> .76
Peer <sup>b</sup>	4,078	4	.20	.22	.26	.28	1.82	-.10	.62	-.02	.54	<sup>bc</sup> .74
Superior <sup>c</sup>	4,501	6	.18	.17	.24	.21	4.30	-.03	.51	.06	.41	<sup>ac</sup> 1.55
Initiating Structure												
Subordinate <sup>a</sup>	4,544	9	.16	.10	.20	.11	18.94	.06	.34	.12	.28	<sup>ba</sup> -.78
Peer <sup>b</sup>	4,078	4	.15	.12	.18	.14	6.25	.00	.36	.03	.32	<sup>bc</sup> .39

Table 5 (cont.)

Superior <sup>c</sup>	4,474	5	.14	.10	.17	.11	11.51	.03	.31	.07	.27	ac	1.20
Transactional													
Subordinate	1,072	4	.36	.11	.51	.13	23.23	.34	.68	.36	.66		
Additional dimensions													
Other													
Subordinate <sup>a</sup>	11,738	12	.24	.10	.31	.12	8.76	.15	.47	.24	.39	ba	0
Peer <sup>b</sup>	9,705	8	.24	.12	.31	.15	5.24	.12	.50	.20	.42	bc	-.59
Superior <sup>c</sup>	10,605	8	.25	.10	.32	.13	6.31	.16	.48	.23	.41	ac	-.62

*Note.*  $r_m$  = mean sample size-weighted correlation;  $SD_r$  = sample size-weighted observed standard deviation of correlations;  $\rho$  = mean sample size-weighted correlation corrected for internal consistency reliability;  $SD_\rho$  = corrected standard deviation of corrected correlations; % var. = percentage of variance attributable to statistical artifacts;  $CV_{10}$  and  $CV_{90}$  = 10% and 90% credibility values, respectively;  $CI_L$  and  $CI_U$  = lower and upper bounds, respectively, of the 95% confidence interval around the corrected mean correlation

Table 6

*Meta-analytic Results: Leader-Observer Mean Differences*

<b>Leadership Style</b>	<b>N</b>	<b><i>k</i></b>	<b><i>d<sub>m</sub></i></b>	<b><i>SD<sub>d</sub></i></b>	<b><math>\delta</math></b>	<b><i>SD<sub>d</sub></i></b>	<b>% Var</b>	<b>CV<sub>10</sub></b>	<b>CV<sub>90</sub></b>	<b>CI<sub>L</sub></b>	<b>CI<sub>U</sub></b>	<b>Z test</b>
Relation-Oriented <sup>b</sup>	8,764	56	.11	.53	.10	.51	12.04	-.56	.76	.05	.15	<sup>ab</sup> -4.26*
Transformational	4,299	37	.28	.58	.28	.57	13.55	-.45	1.00	.21	.35	
Consideration	4,071	6	-.19	.17	-.23	.19	20.36	-.47	.01	-.31	-.15	
Servant	2,298	17	.30	.50	.32	.47	16.77	-.29	.92	.22	.42	
Task-Oriented <sup>a</sup>	4,948	21	-.07	.26	-.09	.28	27.75	-.46	.27	-.17	-.02	<sup>ca</sup> -2.35*
Initiating Structure	4,296	8	-.10	.24	-.13	.27	13.65	-.48	.22	-.20	-.05	
Transactional	1,643	15	.05	.24	.06	.19	68.29	-.17	.30	-.08	.20	
Additional dimensions												
Other	12,638	14	-.11	.25	-.12	.29	7.64	-.49	.24	-.17	-.08	
Laissez-faire <sup>c</sup>	463	9	-.31	.46	-.40	.48	40.47	-1.01	.21	-.67	-.15	<sup>cb</sup> -3.97*

*Note.* Ratings from subordinates, peers, and superiors were combined into overall observer-ratings.  $d_m$  = mean sample size-weighted correlation;  $SD_d$  = sample size-weighted observed standard deviation of correlations;  $\delta$  = mean sample size-weighted correlation corrected for internal consistency reliability;  $SD_\delta$  = corrected standard deviation of corrected correlations; % var. = percentage of variance attributable to statistical artifacts; CV<sub>10</sub> and CV<sub>90</sub> = 10% and 90% credibility values, respectively; CI<sub>L</sub> and CI<sub>U</sub> = lower and upper bounds, respectively, of the 95% confidence interval around the corrected mean correlation.

Table 7

*Meta-analytic Results: Leader-Observer Mean Differences (separated by type of observer)*

<b>Leadership Style</b>	<b>N</b>	<b>k</b>	<b><math>d_m</math></b>	<b><math>SD_d</math></b>	<b><math>\delta</math></b>	<b><math>SD_d</math></b>	<b>% Var</b>	<b>CV<sub>10</sub></b>	<b>CV<sub>90</sub></b>	<b>CI<sub>L</sub></b>	<b>CI<sub>U</sub></b>	<b>Z test</b>
Relation-Oriented												
Subordinate <sup>a</sup>	8,649	53	.06	.58	.05	.57	9.57	-.69	.78	.00	.10	<sup>ba</sup> 4.74*
Peer <sup>b</sup>	5,267	9	-.16	.23	-.18	.24	14.05	-.49	.12	-.25	-.12	<sup>bc</sup> -2.02*
Superior <sup>c</sup>	5,539	16	-.08	.24	-.09	.22	24.14	-.37	.20	-.15	-.02	<sup>ac</sup> 2.57*
Transformational												
Subordinate <sup>a</sup>	4,225	35	.24	.60	.24	.60	11.85	-.53	1.01	.17	.31	<sup>ba</sup> -5.89*
Peer <sup>b</sup>	1,922	5	-.12	.06	-.14	.00	100.00	-.14	-.14	-.24	-.03	<sup>bc</sup> 1.08
Superior <sup>c</sup>	2,163	10	-.19	.32	-.22	.31	21.83	-.61	.17	-.32	-.12	<sup>ac</sup> 7.44*
Consideration												
Subordinate <sup>a</sup>	3,675	5	-.29	.23	-.36	.27	10.90	-.70	-.02	-.44	-.28	<sup>ba</sup> 2.14*
Peer <sup>b</sup>	3,555	3	-.19	.25	-.23	.29	5.70	-.61	.14	-.31	-.15	<sup>bc</sup> -3.15*
Superior <sup>c</sup>	4,040	5	-.04	.13	-.05	.14	28.69	-.23	.12	-.13	.02	<sup>ac</sup> -5.40*
Ethical												
Subordinate	270	3	1.26	.72	1.29	.78	13.63	.30	2.29	.98	1.69	
Servant												
Subordinate <sup>a</sup>	1,806	16	.44	.54	.47	.53	17.02	-.20	1.14	.36	.59	<sup>ba</sup> -4.60*
Peer <sup>b</sup>	1,484	3	.07	.14	.08	.13	41.49	-.08	.24	-.04	.20	<sup>bc</sup> -.67
Superior <sup>c</sup>	1,030	3	.12	.09	.14	.00	100.00	.14	.14	.00	.29	<sup>ac</sup> 3.42*
Task-Oriented												
Subordinate <sup>a</sup>	4,552	20	-.17	.29	-.23	.34	22.42	-.66	.20	-.31	-.15	<sup>ba</sup> .63
Peer <sup>b</sup>	3,596	4	-.20	.24	-.27	.31	7.69	-.66	.13	-.35	-.18	<sup>bc</sup> -4.83*
Superior <sup>c</sup>	4,092	7	.02	.15	.03	.17	29.96	-.19	.25	-.05	.11	<sup>ac</sup> -4.47*
Initiating Structure												

Table 7 (cont.)

Subordinate <sup>a</sup>	3,900	7	-.22	.28	-.27	.32	9.86	-.69	.15	-.35	-.19	ba .59
Peer <sup>b</sup>	3,555	3	-.19	.26	-.23	.31	5.13	-.64	.17	-.32	-.15	bc -4.87*
Superior <sup>c</sup>	4,040	5	.04	.14	.04	.15	24.99	-.15	.24	-.03	.12	ac -5.60*
Transactional												
Subordinate <sup>a</sup>	1,643	15	.05	.24	.07	.18	69.19	-.16	.30	-.07	.21	ab 2.29*
Superior <sup>b</sup>	899	3	-.14	.07	-.20	.00	100.00	-.20	-.20	-.39	-.01	
Additional dimensions												
Other												
Subordinate <sup>a</sup>	10,632	10	-.06	.21	-.07	.24	9.22	-.37	.24	-.12	-.02	ba -2.25*
Peer <sup>b</sup>	10,008	7	-.13	.29	-.15	.33	3.62	-.57	.28	-.19	-.10	bc .13
Superior <sup>c</sup>	11,268	9	-.13	.23	-.15	.26	6.46	-.49	.19	-.20	-.11	ac .24
Laissez-Faire												
Subordinate	463	9	-.28	.50	-.37	.54	34.78	-1.07	.32	-.64	-.12	

*Note.*  $d_m$  = mean sample size-weighted correlation;  $SD_d$  = sample size-weighted observed standard deviation of correlations;  $\delta$  = mean sample size-weighted correlation corrected for internal consistency reliability;  $SD_\delta$  = corrected standard deviation of corrected correlations; % var. = percentage of variance attributable to statistical artifacts;  $CV_{10}$  and  $CV_{90}$  = 10% and 90% credibility values, respectively;  $CI_L$  and  $CI_U$  = lower and upper bounds, respectively, of the 95% confidence interval around the corrected mean correlation.

Table 8

*Meta-analytic Results: Moderators of Leader-Observer Correlations*

<b>Leadership</b>	<b>N</b>	<b>k</b>	<b><math>r_m</math></b>	<b><math>SD_r</math></b>	<b><math>\rho</math></b>	<b><math>SD_\rho</math></b>	<b>% Var</b>	<b>CV<sub>10</sub></b>	<b>CV<sub>90</sub></b>	<b>CI<sub>L</sub></b>	<b>CI<sub>U</sub></b>	<b>Z test</b>
Relation-Oriented												
Country: US	3,806	22	.26	.15	.31	.15	23.75	.12	.49	.24	.38	.19
Country: Non-US	3,021	10	.26	.09	.30	.09	33.66	.19	.41	.23	.37	
Leader Level: Lower	726	10	.19	.14	.22	.09	67.98	.11	.34	.12	.32	1.78
Leader Level: Upper	285	5	.07	.14	.08	.06	87.13	.01	.16	-.06	.23	
Purpose: Leader dev.	3,167	6	.29	.06	.34	.05	48.21	.27	.40	.28	.39	2.26*
Purpose: Non-leader dev.	3,660	26	.24	.16	.28	.16	25.40	.08	.48	.21	.35	
Sampling: Leader did not select	1,070	14	.23	.29	.27	.30	14.40	-.12	.66	.10	.45	-1.57
Sampling: Leader selected	4,874	12	.28	.11	.33	.11	18.11	.19	.47	.26	.40	
Published	1,953	15	.26	.18	.30	.18	20.86	.06	.53	.19	.40	-.35
Unpublished	4,874	17	.26	.09	.31	.09	34.17	.19	.42	.25	.36	
Task-Oriented												
Country: US	2,699	9	.28	.13	.37	.16	15.97	.17	.58	.26	.49	6.82*
Country: Non-US	2,619	5	.10	.05	.14	.04	70.55	.09	.18	.08	.20	
Purpose: Leader dev.	2,852	3	.17	.06	.22	.07	25.00	.13	.32	.13	.32	-2.02
Purpose: Non-leader dev.	2,466	11	.22	.18	.29	.22	12.14	.01	.58	.15	.44	
Sampling: Leader did not select	279	4	.31	.31	.41	.38	12.42	-.08	.89	.00	.81	1.85
Sampling: Leader selected	4,556	8	.20	.14	.27	.18	8.20	.04	.49	.14	.40	
Published	1,533	8	.32	.16	.42	.19	16.58	.18	.67	.27	.57	6.23*
Unpublished	3,909	7	.14	.08	.19	.09	26.58	.08	.31	.11	.27	

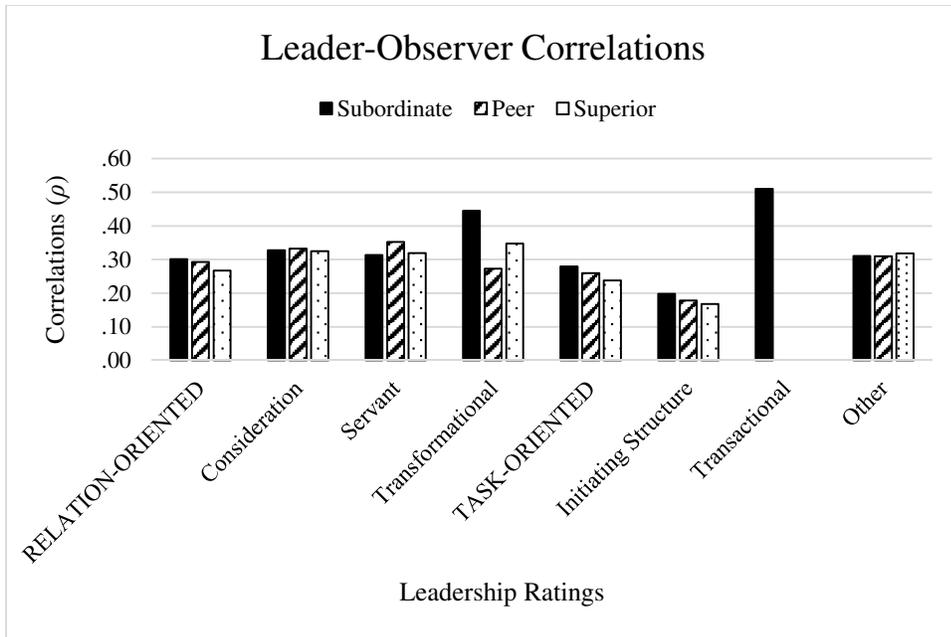
*Note.* "Leader did not select" refers to observer-raters being randomly selected or selected by the researcher. "Leader selected" refers to observer-raters being selected by leaders.

Table 9

*Meta-analytic Results: Moderators of Leader-Observer Mean Differences*

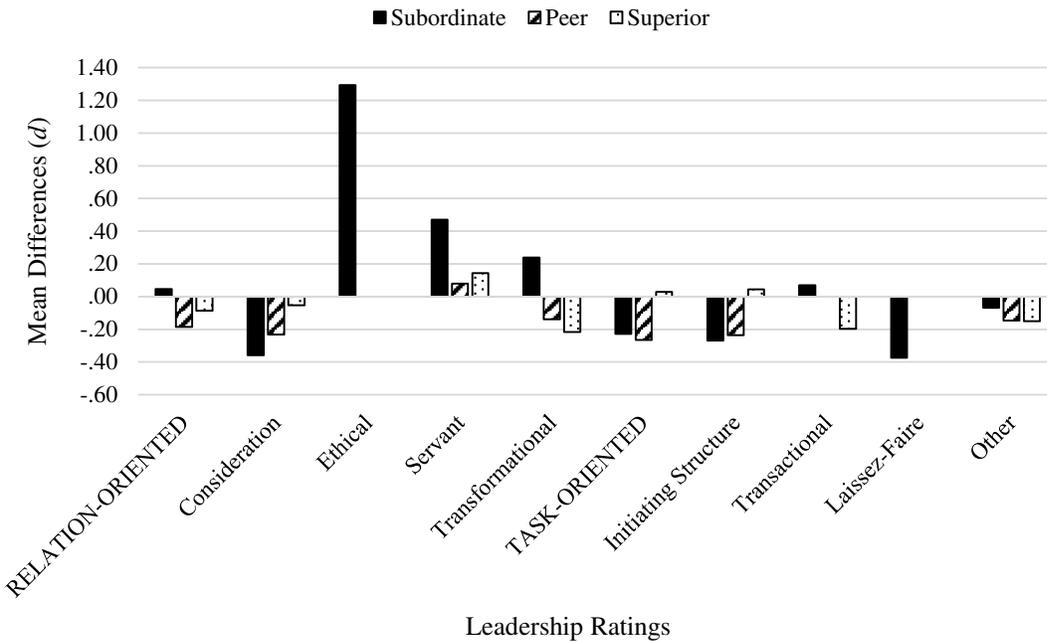
<b>Leadership</b>	<b>N</b>	<b>k</b>	<b><math>d_m</math></b>	<b><math>SD_d</math></b>	<b><math>\delta</math></b>	<b><math>SD_\delta</math></b>	<b>% Var</b>	<b>CV<sub>10</sub></b>	<b>CV<sub>90</sub></b>	<b>CI<sub>L</sub></b>	<b>CI<sub>U</sub></b>	<b>Z test</b>
Relation-Oriented												
Country: US	5,352	40	.21	.48	.22	.46	16.96	-.37	.80	.15	.28	7.18*
Country: Non-US	3,277	14	-.12	.46	-.15	.47	9.86	-.75	.45	-.23	-.07	
Leader Level: Lower	1,064	20	.58	.56	.62	.49	33.85	-.01	1.24	.47	.78	1.37
Leader Level: Upper	730	9	.42	.41	.46	.31	44.51	.06	.85	.28	.64	
Purpose: Leader dev.	3,106	6	-.18	.22	-.21	.23	16.98	-.50	.08	-.29	-.13	-8.66
Purpose: Non-leader dev.	5,523	48	.24	.55	.24	.53	14.90	-.44	.92	.18	.30	
Scale: LPI	650	8	.40	.27	.46	.13	80.83	.28	.63	.27	.66	-2.75
Scale: MLQ	1,604	22	.70	.51	.77	.51	26.89	.12	1.42	.64	.90	
Sampling: Leader did not select	1,437	26	.66	.44	.73	.39	45.74	.24	1.23	.60	.87	13.04*
Sampling: Leader selected	6,012	17	-.11	.29	-.13	.29	15.63	-.50	.24	-.19	-.07	
Published	3,474	30	.19	.56	.18	.52	15.48	-.48	.85	.10	.26	2.74*
Unpublished	5,232	24	.06	.51	.04	.51	8.91	-.61	.70	-.02	.11	
Task-Oriented												
Country: US	3,115	18	.03	.27	.03	.28	34.09	-.33	.40	-.06	.13	4.35*
Country: Non-US	1,860	4	-.23	.10	-.31	.06	83.45	-.38	-.23	-.43	-.18	
Leader Level: Lower	444	8	-.01	.36	-.01	.29	61.30	-.38	.35	-.27	.24	.65
Leader Level: Upper	238	4	-.12	.28	-.15	.13	87.05	-.33	.02	-.51	.19	
Purpose: Leader dev.	2,852	3	-.22	.05	-.29	.00	100.00	-.29	-.29	-.40	-.20	-6.17
Purpose: Non-leader dev.	2,123	19	.14	.28	.17	.25	50.35	-.15	.50	.06	.29	
Sampling: Leader did not select	586	11	.11	.42	.14	.39	47.20	-.36	.65	-.07	.37	2.94*
Sampling: Leader selected	3,754	7	-.15	.12	-.20	.11	52.21	-.35	-.06	-.29	-.12	
Published	1,627	15	-.02	.20	-.03	.06	94.36	-.11	.05	-.16	.10	1.24
Unpublished	3,348	7	-.09	.27	-.13	.33	11.90	-.55	.30	-.22	-.04	

*Note.* "Leader did not select" refers to observer-raters being randomly selected or selected by the researcher. "Leader selected" refers to observer-raters being selected by leaders. LPI = Leadership Practices Inventory (Kouzes & Posner). MLQ = Multifactor Leadership Questionnaire (Bass & Avolio).

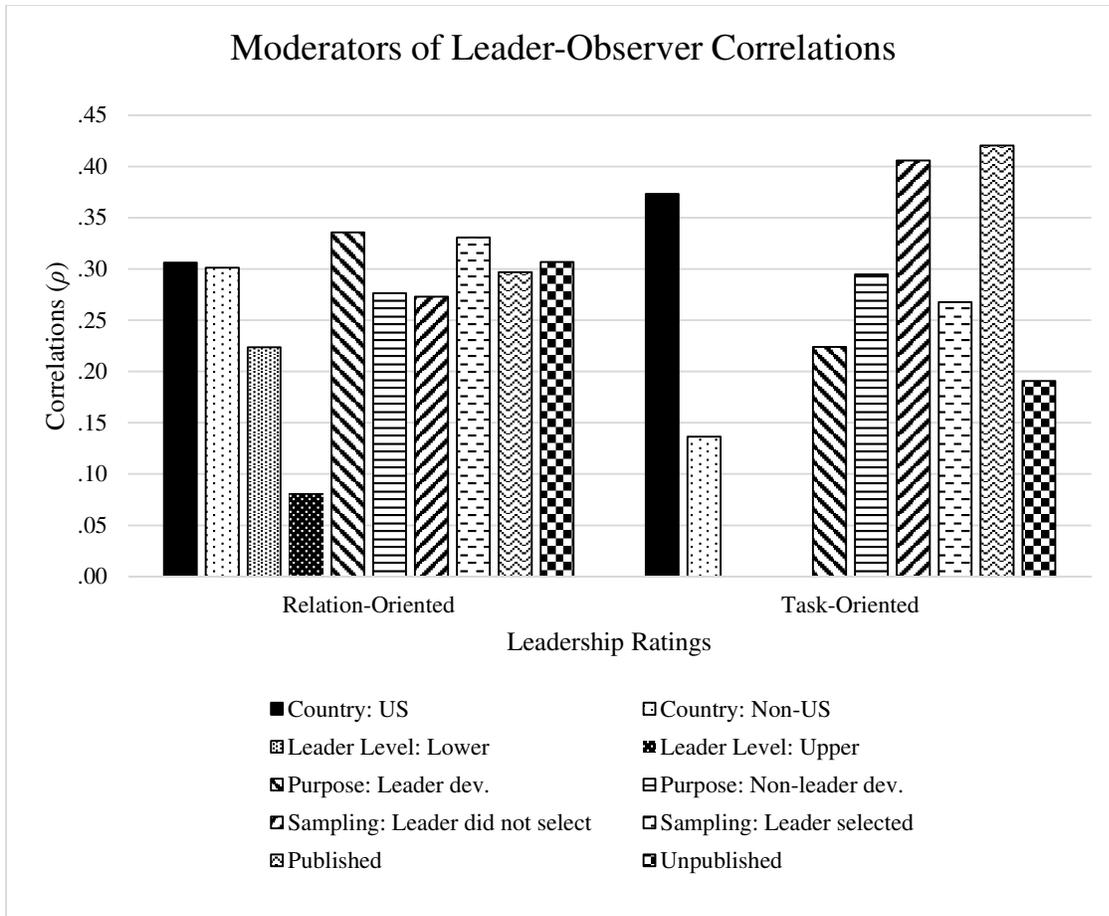


*Figure 1.* Correlations ( $\rho$ ) between leader- and observer-ratings of leadership. Relation-oriented subordinate  $N = 6,127$ ;  $k = 29$ ; peer  $N = 4,343$ ;  $k = 6$ ; superior  $N = 4,813$ ;  $k = 9$ . Consideration subordinate  $N = 4,463$ ;  $k = 8$ ; peer  $N = 4,078$ ;  $k = 4$ ; superior  $N = 4,474$ ;  $k = 5$ . Servant subordinate  $N = 1,974$ ;  $k = 8$ ; peer  $N = 1,515$ ;  $k = 3$ ; superior  $N = 1,515$ ;  $k = 3$ . Transformational subordinate  $N = 2,688$ ;  $k = 19$ ; peer  $N = 1,490$ ;  $k = 3$ ; superior  $N = 1,564$ ;  $k = 5$ . Task-oriented subordinate  $N = 4,769$ ;  $k = 12$ ; peer  $N = 4,078$ ;  $k = 4$ ; superior  $N = 4,501$ ;  $k = 6$ . Initiating structure subordinate  $N = 4,544$ ,  $k = 9$ ; peer  $N = 4,078$ ;  $k = 4$ ; superior  $N = 4,474$ ;  $k = 5$ . Transaccional subordinate  $N = 1,072$ ;  $k = 4$ ; superior  $N = 899$ ;  $k = 3$ . Other subordinate  $N = 11,738$ ;  $k = 12$ ; peer  $N = 9,705$ ;  $k = 8$ ; superior  $N = 10,605$ ;  $k = 8$ .

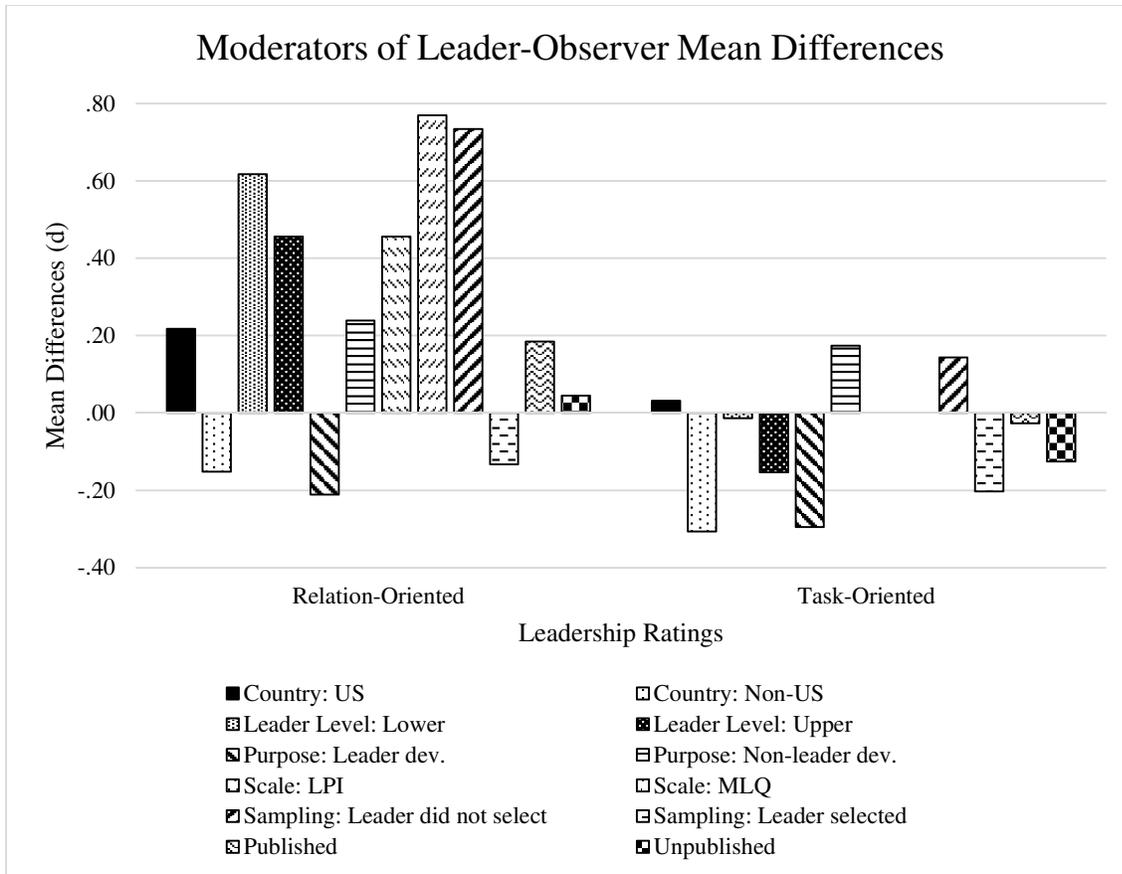
## Leader-Observer Mean Differences



*Figure 2.* Mean differences ( $d$ ) between leader- and observer-ratings in leadership. Positive  $d$ 's indicate leader over-reporting and negative  $d$ 's indicate leader under-reporting. Relation-oriented subordinate  $N = 8,649$ ;  $k = 53$ ; peer  $N = 5,267$ ;  $k = 9$ ; superior  $N = 5,539$ ;  $k = 16$ . Consideration subordinate  $N = 6,675$ ;  $k = 5$ ; peer  $N = 3,555$ ;  $k = 3$ ; superior  $N = 4,040$ ;  $k = 5$ . Ethical subordinate  $N = 270$ ;  $k = 3$ . Servant subordinate  $N = 1,806$ ;  $k = 16$ ; peer  $N = 1,484$ ;  $k = 3$ ; superior  $N = 1,030$ ;  $k = 3$ . Transformational subordinate  $N = 4,225$ ;  $k = 35$ ; peer  $N = 1,922$ ;  $k = 5$ ; superior  $N = 2,163$ ;  $k = 10$ . Task-oriented subordinate  $N = 4,552$ ;  $k = 20$ ; peer  $N = 3,596$ ;  $k = 4$ ; superior  $N = 4,092$ ;  $k = 7$ . Initiating structure subordinate  $N = 3,900$ ;  $k = 7$ ; peer  $N = 3,555$ ;  $k = 3$ ; superior  $N = 4,040$ ;  $k = 5$ . Transactional subordinate  $N = 1,643$ ;  $k = 15$ ; superior  $N = 899$ ;  $k = 3$ . Laissez-faire subordinate  $N = 463$ ;  $k = 9$ . Other subordinate  $N = 10,632$ ;  $k = 10$ ; peer  $N = 10,008$ ;  $k = 7$ ; superior  $N = 11,268$ ;  $k = 9$ .



*Figure 3. Moderators of the correlations ( $\rho$ ) between leader- and observer-ratings of leadership. Relation-Oriented – Country: US  $N = 3,806$ ;  $k = 22$ . Country: Non-US  $N = 3,021$ ;  $k = 10$ . Leader Level: Lower  $N = 726$ ;  $k = 10$ . Leader Level: Upper  $N = 285$ ;  $k = 5$ . Purpose: Leader dev.  $N = 3,167$ ;  $k = 6$ . Purpose: Non-leader dev.  $N = 3,660$ ;  $k = 26$ . Sampling: Leader didn't select  $N = 1,070$ ;  $k = 14$ . Sampling: Leader selected  $N = 4,874$ ;  $k = 12$ . Published  $N = 1,953$ ;  $k = 15$ . Unpublished  $N = 4,874$ ;  $k = 17$ . Task-Oriented – Country: US  $N = 2,699$ ;  $k = 9$ . Country: Non-US  $N = 2,619$ ;  $k = 5$ . Purpose: Leader dev.  $N = 2,852$ ;  $k = 3$ . Purpose: Non-leader dev.  $N = 2,466$ ;  $k = 11$ . Sampling: Leader didn't select  $N = 279$ ;  $k = 4$ . Sampling: Leader selected  $N = 4,556$ ;  $k = 8$ . Published  $N = 1,533$ ;  $k = 8$ . Unpublished  $N = 3,909$ ;  $k = 7$ .*



*Figure 4.* Moderators of the mean differences (*d*) between leader- and observer-ratings of leadership. Relation-Oriented – Country: US  $N = 5,352$ ;  $k = 40$ . Country: Non-US  $N = 3,277$ ;  $k = 14$ . Leader Level: Lower  $N = 1,064$ ;  $k = 20$ . Leader Level: Upper  $N = 730$ ;  $k = 9$ . Purpose: Leader dev.  $N = 3,106$ ;  $k = 6$ . Purpose: Non-leader dev.  $N = 5,523$ ;  $k = 48$ . Scale: LPI  $N = 650$ ;  $k = 8$ . Scale: MLQ  $N = 1,604$ ;  $k = 22$ . Sampling: Leader didn't select  $N = 1,437$ ;  $k = 26$ . Sampling: Leader selected  $N = 6,012$ ;  $k = 17$ . Published  $N = 3,474$ ;  $k = 30$ . Unpublished  $N = 5,232$ ;  $k = 24$ . Task-Oriented – Country: US  $N = 3,155$ ;  $k = 18$ . Country: Non-US  $N = 1,860$ ;  $k = 4$ . Leader Level: Lower  $N = 444$ ;  $k = 8$ . Leader Level: Upper  $N = 238$ ;  $k = 4$ . Purpose: Leader dev.  $N = 2,852$ ;  $k = 3$ . Purpose: Non-leader dev.  $N = 2,123$ ;  $k = 19$ . Sampling: Leader didn't select  $N = 586$ ;  $k = 22$ . Sampling: Leader selected  $N = 3,754$ ;  $k = 7$ . Published  $N = 1,627$ ;  $k = 15$ . Unpublished  $N = 3,348$ ;  $k = 7$ .

## REFERENCES

- \*Abouel-Enin, M. (1995). *The relationship among perceived leadership style, leadership decision-making criteria, and leadership effectiveness in business settings in southern california* (Unpublished doctoral dissertation). United States International University.
- Allen, T. D., Barnard, S., Rush, M. C., & Russell, J. E. A. (2000). Ratings of organizational citizenship behavior: Does the source make a difference? *Human Resource Management Review, 10*, 97-114. doi:10.1016/S1053-4822(99)00041-8
- \*Amundsen, S., & Martinsen, Ø. L. (2014). Self–other agreement in empowering leadership: Relationships with leader effectiveness and subordinates' job satisfaction and turnover intention. *The Leadership Quarterly, 25*, 784-800. doi:http://dx.doi.org/10.1016/j.leaqua.2014.04.007
- \*Arensberg, M. F., Schiller, M. R., Vivian, V. M., Johnson, W. A., & Strasser, S. (1996). Transformational leadership of clinical nutrition managers. *Journal of the American Dietetic Association, 96*, 39-45.
- \*Atalla, A. D. (2010). *Self-assessment vs. employee perceptions of nonprofit leadership* (Unpublished doctoral dissertation). Capella University.
- \*Atwater, L. E., & Brett, J. F. (2005). Antecedents and consequences of reactions to developmental 360° feedback. *Journal of Vocational Behavior, 66*, 532-548. doi:http://dx.doi.org/10.1016/j.jvb.2004.05.003
- 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. A., Roush, P., & Fischthal, A. (1995). The influence of upward feedback on self- and follower ratings of leadership. *Personnel Psychology, 48*, 35–60.
- Atwater, L.E., Waldman, D., Ostroff, C., Robie, C., & Johnson, K. M. (2005). Self-other agreement: Comparing its relationship with performance in the U.S. and europe. *International Journal of Selection and Assessment, 13*, 25-40. doi:http://dx.doi.org/10.1111/j.0965-075X.2005.00297.x
- \*Atwater, L., Wang, M., Smither, J. W., & Fleenor, J. W. (2009). Are cultural characteristics associated with the relationship between self and others' ratings of leadership? *Journal of Applied Psychology, 94*, 876-886. doi:http://dx.doi.org/10.1037/a0014561
- \*Atwater, L. E., & Yammarino, F. J. (1992). Does self-other agreement on leadership perceptions moderate the validity of leadership and performance predictions? *Personnel Psychology, 45*, 141-164.
- \*Baal, K. M. (2011). *Leadership practices of science department chairs in secondary public schools* (Unpublished doctoral dissertation). Loyola University.
- Bales, R. F. (1950). *Interaction process analysis: a method for the study of small groups*. Cambridge, MA: Addison-Wesley

- Barbuto, J. E., & Wheeler, D. W. (2006). Scale development and construct clarification of servant leadership. *Group & Organization Management, 31*, 300-326.
- \*Barbuto, J. E., Jr., Wilmot, M. P., Singh, M., & Story, J. S. P. (2012). Self-other rating agreement and leader-member exchange (LMX): A quasi-replication. *Perceptual and Motor Skills, 114*, 479-484. doi:<http://dx.doi.org/10.2466/01.07.28.PMS.114.2.479-484>.
- Bass, B. M. (1990). *Bass and Stogdill's handbook of leadership*. New York: Free Press.
- \*Bass, B. M., & Yammarino, F. J. (1991). Congruence of self and others' leadership ratings of naval officers for understanding successful performance. *Applied Psychology: An International Review, 40*, 437-454.
- \*Beck, C. D. (2014). Antecedents of Servant Leadership A Mixed Methods Study. *Journal of Leadership & Organizational Studies, 21*, 299-314. doi: 10.1177/1548051814529993
- \*Berson, Y., & Sosik, J. J. (2007). The relationship between self-other rating agreement and influence tactics and organizational processes. *Group & Organization Management, 32*, 675-698.
- Blake, R. R., Mouton, J. S., & Bidwell, A. C. (1962). Managerial grid. *Advanced Management-Office Executive, 1*, 12-15.
- \*Borchert, D. M. (2012). *A meta-model of ethical behavior: An empirical examination of ethical leadership, ethical identity, ethical climate and emotions on unethical work behavior* (Unpublished doctoral dissertation). St. Louis University.
- Bosco, F. A., Aguinis, H., Singh, K., Field, J. G., & Pierce, C. A. (2015). Correlational effect size benchmarks. *Journal of Applied Psychology, 100*, 431-449.
- \*Braddy, P. W., Gooty, J., Fleenor, J. W., & Yammarino, F. J. (2014). Leader behaviors and career derailment potential: A multi-analytic method examination of rating source and self-other agreement. *The Leadership Quarterly, 25*, 373-390. doi:<http://dx.doi.org/10.1016/j.leaqua.2013.10.001>
- \*Bratton, V. K., Dodd, N. G., & Brown, F. W. (2011). The impact of emotional intelligence on accuracy of self-awareness and leadership performance. *Leadership & Organization Development Journal, 32*, 127-149. doi:<http://dx.doi.org/10.1108/01437731111112971>
- Brown, B. B. (2003). *Employees' organizational commitment and their perception of supervisors' relations-oriented and task-oriented leadership behaviors* (Unpublished doctoral dissertation). Virginia Polytechnic Institute and State University.
- Brown, M. E., Treviño, L. K., & Harrison, D. A. (2005). Ethical leadership: A social learning perspective for construct development and testing. *Organizational behavior and human decision processes, 97*, 117-134.
- \*Bugenhagen, M. J. (2006). *Antecedents of transactional, transformational, and servant leadership: A constructive-development theory approach* (Unpublished doctoral dissertation). University of Nebraska at Lincoln.

- Burris, E. R., Detert, J. R., & Romney, A. C. (2013). Speaking up vs. being heard: The disagreement around and outcomes of employee voice. *Organization Science*, 24, 22-38. doi:<http://dx.doi.org/10.1287/orsc.1110.0732>
- Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological bulletin*, 56, 81-105.
- Campbell, J. P., McCloy, R. A., Oppler, S. H., & Sager, C. E. (1993). A theory of performance. In N. Schmitt & W. C. Borman (Eds.), *Personnel selection in organizations* (pp. 35–70). San Francisco, CA: Jossey-Bass.
- \*Castellese, S. R. (2006). *Guatemala health care practitioners' leadership styles: Medical worker perception versus leader self-perception* (Unpublished doctoral dissertation). Capella University.
- Chan, D. (2009). So why ask me? Are self-report data really that bad? In C. E. Lance & R. J. Vandenberg (Eds.), *Statistical and methodological myths and urban legends: Doctrine, verity and fable in the organizational and social sciences* (pp. 311–338). New York: Routledge.
- \*Chun, J. U. (2006). *Close and distant charismatic and contingent reward leadership: Multiple levels-of-management and multiple levels-of-analysis perspectives* (Unpublished doctoral dissertation). State University of New York at Binghamton.
- Conway, J. M., Lombardo, K., & Sanders, K. C. (2001). A meta-analysis of incremental validity and nomological networks for subordinate and peer rating. *Human Performance*, 14, 267-303.
- \*Davis, V. E. (2011). *The perceived leadership style of senior enlisted personnel in the U.S. armed forces* (Unpublished doctoral dissertation). Northcentral University.
- \*de Vries, R. E. (2012). Personality predictors of leadership styles and the self–other agreement problem. *The Leadership Quarterly*, 23, 809-821. doi:<http://dx.doi.org/10.1016/j.leaqua.2012.03.005>
- \*Delmhorst, F. P. (2006). *Self-awareness and leadership development: Examining the impact of self-subordinate agreement on perceptions of leadership* (Unpublished doctoral dissertation). Columbia University.
- \*DeLucia, J. M. (2011). *Barriers and Supports to Implementation of Principal Leadership for School Change* (Unpublished doctoral dissertation). Walden University.
- \*Dennis, C. J. (2009). *The relationship between principals' self-perceptions and teachers' perceptions of high school principals' instructional leadership behaviors in south carolina high schools* (Unpublished doctoral dissertation). University of South Carolina.
- Densten, I. L., & Sarros, J. C. (2012). The impact of organizational culture and social desirability on Australian CEO leadership. *Leadership & Organization Development Journal*, 33, 342-368.
- \*Devos, G., Hulpia, H., Tuytens, M., & Sinnaeve, I. (2013). Self-other agreement as an alternative perspective of school leadership analysis: An exploratory study. *School*

*Effectiveness and School Improvement*, 24, 296-315.  
doi:<http://dx.doi.org/10.1080/09243453.2012.693103>

- \*Eckert, R., Ekelund, B. Z., Gentry, W. A., & Dawson, J. F. (2010). "I don't see me like you see me, but is that a problem?" cultural influences on rating discrepancy in 360-degree feedback instruments. *European Journal of Work and Organizational Psychology*, 19, 259-278. doi:<http://dx.doi.org/10.1080/13594320802678414>
- Edwards, J. R., & Berry, J. W. (2010). The presence of something or the absence of nothing: Increasing theoretical precision in management research. *Organizational Research Methods*, 13, 668-689.
- \*Etheridge, J. C. (2009). *Mirror, mirror... perceptions of principal leadership behavior and the effect on student achievement* (Unpublished doctoral dissertation). University of South Carolina.
- Facteau, J.D. & Craig, S.B. (2001). Are performance appraisal ratings from different rating sources comparable? *Journal of Applied Psychology*, 86, 215–227.
- \*Felfe, J., & Heinitz, K. (2010). The impact of consensus and agreement of leadership perceptions on commitment, organizational citizenship behaviour, and customer satisfaction. *European Journal of Work and Organizational Psychology*, 19, 279-303. doi:<http://dx.doi.org/10.1080/13594320802708070>
- \*Fleenor, J. W., McCauley, C. D., & Brutus, S. (1996). Self-other rating agreement and leader effectiveness. *The Leadership Quarterly*, 7, 487-506.
- Fleenor, J. W., Smither, J. W., Atwater, L. E., Braddy, P. W., & Sturm, R. E. (2010). Self–other rating agreement in leadership: A review. *The Leadership Quarterly*, 21, 1005-1034.
- Fleishman, E. A. (1953). Leadership climate, human relations training, and supervisory behavior. *Personnel Psychology*, 6, 205-222.
- Fleishman, E. A. (1973). Twenty years of consideration and structure. In E. A. Fleishman & J. G. Hunt (Eds.), *Current developments in the study of leadership* (pp. 1–40). Carbondale, IL: Southern Illinois University Press.
- Funder, D. C. (2012). Accurate personality judgment. *Current Directions in Psychological Science*, 21, 177-182.
- \*Gaylord, T. W. (2003). *Feedback acceptance within leadership development: Effects of self-other rating discrepancies, conscientiousness and neuroticism* (Unpublished doctoral dissertation). Kansas State University.
- \*Gentry, W. A., Hannum, K. M., Ekelund, B. Z., & de Jong, A. (2007). A study of the discrepancy between self- and observer-ratings on managerial derailment characteristics of european managers. *European Journal of Work and Organizational Psychology*, 16, 295-325. doi:<http://dx.doi.org/10.1080/13594320701394188>

- \*Godshalk, V. M., & Sosik, J. J. (2000). Does mentor-protégé agreement on mentor leadership behavior influence the quality of a mentoring relationship? *Group & Organization Management, 25*, 291-317.
- \*Goldberg, L. K. (1991). *What is the nurse administrator's perception of his/her leadership as perceived by the nurse middle manager* (Unpublished doctoral dissertation). Columbia University Teachers College.
- \*Grafton, K. S. (2009). *Presidential transformational leadership practices: Analysis of self-perceptions and observers at community colleges in Oklahoma* (Unpublished doctoral dissertation). University of Phoenix.
- Greenleaf, R. K. 1977. *Servant Leadership: A Journey into the Nature of Legitimate Power and Greatness*. New York, NY: Paulist Press.
- Halpin, A. W., & Winer, B. J. (1957). A factorial study of the leader behavior descriptions. *Leader behavior: Its description and measurement*, 39-51.
- Halverson, S. K., Tonidandel, S., Barlow, C. B., & Dipboye, R. L. (2005). Self–other agreement on a 360-degree leadership evaluation. In S. Reddy (Ed.), *Multi-source performance assessment: Perspective and Insights* (pp. 125–144). Hyderabad, India: ICFAI University Press.
- Harris, L. S., & Kuhnert, K. W. (2008). Looking through the lens of leadership: A constructive developmental approach. *Leadership & Organization Development Journal, 29*, 47-67.
- Harris, M.M. & Schaubroeck, J. (1988). A meta-analysis of self-supervisor, self-peer, and peer-supervisor ratings. *Personnel Psychology, 41*, 43-62.
- \*Hartman, E. L. (2005). *The relationship between CEOs traits, empowering leader behaviors and objective and subjective measures of performance* (Unpublished doctoral dissertation). Illinois Institute of Technology.
- \*Henderson, N. P. (2008). *Teacher and principal perceptions of effective instructional leadership: An exploration of guiding practice and personal beliefs* (Unpublished doctoral dissertation). University of Texas at San Antonio.
- \*Higgins, C. C. (1999). *Transactional and transformational leadership: An examination of the relationship between leadership orientation and perceptions of organizational effectiveness* (Unpublished doctoral dissertation). George Washington University.
- \*Hill, S. A. (2001). *Leadership style, behaviors and spheres of influence of university leaders: Moving beyond position* (Unpublished doctoral dissertation). Queen's University.
- Hiller, N. J., DeChurch, L. A., Murase, T., & Doty, D. (2011). Searching for outcomes of leadership: A 25-year review. *Journal of Management, 37*, 1137–1177. doi:10.1177/0149206310393520
- Hooijberg, R., & Choi, J. (2000). Which leadership roles matter to whom? An examination of rater effects on perceptions of effectiveness. *The Leadership Quarterly, 11*, 341-364.

- \*Hough, K. L. (2011). *Superintendents' empowering leadership and district achievement: Does humility produce results* (Unpublished doctoral dissertation). George Washington University.
- House, R. J. (1977). A 1976 theory of charismatic leadership. In J. G. Hunt & L. L. Larsen (Eds.), *Leadership: The cutting edge*. Carbondale, IL: Southern Illinois University Press.
- Humphrey, R. H. (2002). The many faces of emotional leadership. *The Leadership Quarterly*, *13*, 493-504.
- \*Kelley, R. C. (1998). *The relationship between the leadership behavior analysis II for self and other and the staff development and climate assessment questionnaire* (Unpublished doctoral dissertation). University of Nevada at Reno.
- \*Keys, M. R., Sr. (2010). *The relationship between transformational leadership behaviors of middle school principals, the development of learning communities, and student achievement in rural middle schools in the mississippi delta* (Unpublished doctoral dissertation). Union University.
- \*Kopperud, K. H., Martinsen, Ø., & Humborstad, S. I. W. (2014). Engaging leaders in the eyes of the beholder: On the relationship between transformational leadership, work engagement, service climate, and self–other agreement. *Journal of Leadership & Organizational Studies*, *21*, 29-42. doi:http://dx.doi.org/10.1177/1548051813475666
- \*Lichter, J. (2010). *Religiosity and transformational leadership in K-8 catholic school principals* (Unpublished doctoral dissertation). Cardinal Stritch University.
- \*London, M., & Wohlers, A. J. (1991). Agreement between subordinate and self-ratings in upward feedback. *Personnel Psychology*, *44*, 375-390.
- Love, K. G. (1981). Comparison of peer assessment methods: Reliability, validity, friendship bias, and user reaction. *Journal of Applied Psychology*, *66*, 451-457.
- \*Malesh, W. L. (2007). *Leaders of university baccalaureate nursing programs leadership style: Perceptions of faculty* (Unpublished doctoral dissertation). Laurentian University.
- \*Maloney, R. J. (1997). *Leadership behavior of head nurses: Self-perceptions compared with the perceptions of supervisors and of subordinates* (Unpublished doctoral dissertation). University of Southern California.
- Maurer, T.J., Raju N.S., & Collins W.C. (1998). Peer and subordinate performance appraisal measurement equivalence. *Journal of Applied Psychology*, *83*, 693–702.
- \*McElhaney, R. F. (2003). *Perceptions of nurse managers' leadership style by nurse managers and RN staff: Job satisfaction as perceived by RN staff* (Unpublished doctoral dissertation). Louisiana State.
- \*Moshavi, D., Brown, F. W., & Dodd, N. G. (2003). Leader self-awareness and its relationship to subordinate attitudes and performance. *Leadership & Organization Development Journal*, *24*, 407-418. doi:http://dx.doi.org/10.1108/01437730310498622

- \*Parks, D. S. (2004). *A 360-degree analysis of the leadership styles of southern baptist ministers of youth* (Unpublished doctoral dissertation). The Southern Baptist Theological Seminary.
- Paulhus 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.
- Pearce, C. L., & Conger, J. A. (2002). *Shared leadership: Reframing the hows and whys of leadership*. Sage Publications.
- \*Phillips, T. J. (2013). *Servant leadership and instructional literacy coaching* (Unpublished doctoral dissertation). Grand Canyon University.
- Pollack, D.M. & Pollack, L.J. (1996). Using 360 degree feedback in performance appraisal. *Public Personnel Management*, 25, 507-28.
- \*Raines Evers, C. J. (2012). *A relational study of elementary principals' leadership traits, teacher morale, and school performance* (Unpublished doctoral dissertation). University of Southern Mississippi.
- Reilly, R. R., Smither, J. W., & Vasilopoulos, N. L. (1996). A longitudinal study of upward feedback. *Personnel Psychology*, 49, 599–612.
- \*Riester, D. N. (2010). *Self-other agreement and leader effectiveness: An examination of differences across rater sources and leader behaviors* (Unpublished doctoral dissertation). DePaul University.
- Riggio, R.E., Chaleff, I., & Lipman-Blumen, J. (Eds.). (2008). *The art of followership: How great followers create great leaders and organizations*. San Francisco, CA: Jossey-Bass.
- \*Roark, C. S. (2013). *A mixed methods study exploring the relationship between servant leadership and emotional intelligence* (Unpublished doctoral dissertation). Indiana Wesleyan University.
- \*Romanowska, J., Larsson, G., & Theorell, T. (2014). An art-based leadership intervention for enhancement of self-awareness, humility, and leader performance. *Journal of Personnel Psychology*, 13, 97-106. doi:<http://dx.doi.org/10.1027/1866-5888/a000114>
- Rothstein, H. R. (1990). Interrater reliability of job performance ratings: Growth to asymptote level with increasing opportunity to observe. *Journal of Applied Psychology*, 75, 322-327.
- \*Rozeboom, D. J. (2009). *Self-report and direct observer's perceived leadership practices of chief student affairs officers in selected institutions of higher education in the united states* (Unpublished doctoral dissertation). Texas A&M University.
- Russell, R. F., & Gregory Stone, A. (2002). A review of servant leadership attributes: Developing a practical model. *Leadership & Organization Development Journal*, 23, 145-157.
- Salancik, G. R., & Pfeffer, J. (1978). A social information processing approach to job attitudes and task design. *Administrative Science Quarterly*, 23, 224-253.

- \*Schneider, S. K., & George, W. M. (2011). Servant leadership versus transformational leadership in voluntary service organizations. *Leadership & Organization Development Journal*, 32, 60-77. doi:http://dx.doi.org/10.1108/01437731111099283
- Schneier, C.E. (1977). Multiple rater groups and performance appraisal. *Public Personnel Management*, 13-20.
- Schwarz, N., & Oyserman, D. (2001). Asking questions about behavior: Cognition, communication, and questionnaire construction. *American Journal of Evaluation*, 22, 127-160.
- \*Seidler, A. R. (2008). *A self-regulatory approach to 360-degree feedback and motivation for leader development* (Unpublished doctoral dissertation). Northern Illinois University.
- \*Shannon, L. D. (2009). *The relationship between south carolina elementary principals' perceptions and teachers' perceptions of the principals' leadership practices and academic success* (Unpublished doctoral dissertation). University of South Carolina.
- Sims, H. P., & Manz, C. C. (1996). *Company of heroes: Unleashing the power of self-leadership*. John Wiley & Sons.
- \*Smith, J. M. (2008). *The relationship between north carolina elementary principals' and teachers' perceptions of principals' leadership styles and the social organization of the school* (Unpublished doctoral dissertation). Fayetteville State University.
- Smither, J. W., London, M., & Reilly, R. R. (2005). Does performance improve following multisource feedback? A theoretical model, meta-analysis, and review of empirical findings. *Personnel Psychology*, 58, 33-66.
- \*Sosik, J. J. (2001). Self-other agreement on charismatic leadership: Relationships with work attitudes and managerial performance. *Group & Organization Management*, 26, 484-511.
- Stang, D.J. (1973). Effect of interaction rate on ratings of leadership and liking. *Journal of Personality and Social Psychology*, 27, 405-408.
- Stogdill, R. M. (1950). Leadership, membership and organization. *Psychological Bulletin*, 47, 1-14.
- \*Svoboda, S. N. (2008). *A correlational study of servant leadership and elementary principal job satisfaction in ohio public school districts* (Unpublished doctoral dissertation). Northcentral University.
- \*Tang, K. Y., Dai, G., & De Meuse, K. P. (2013). Assessing leadership derailment factors in 360° feedback: Differences across position levels and self-other agreement. *Leadership & Organization Development Journal*, 34, 326-343. doi:http://dx.doi.org/10.1108/LODJ-07-2011-0070.
- Treviño, L. K., & Brown, M. E. (2004). Managing to be ethical: Debunking five business ethics myths. *Academy of Management Executive*, 18, 69-81.
- \*Turkel, C. C. (2008). *Female leaders' 360-degree self-perception accuracy for leadership competencies and skills* (Unpublished doctoral dissertation). Capella University.

- Viswesvaran, C., Ones, D. S., & Schmidt, F. L. (1996). Comparative analysis of the reliability of job performance ratings. *Journal of Applied Psychology, 81*, 557-574.
- Waldman, D. A., & Atwater, L. A. (1998). *The power of 360-degree feedback: How to leverage performance evaluations for top productivity*. Houston, TX: Gulf.
- Walumbwa, F. O., Hartnell, C. A., & Oke, A. (2010). Servant leadership, procedural justice climate, service climate, employee attitudes, and organizational citizenship behavior: a cross-level investigation. *Journal of Applied Psychology, 95*, 517-529.
- \*Werner, K. J. (2013). *An examination of the relationship between emotional intelligence and servant leadership practices of elementary public school principals in washington state* (Unpublished doctoral dissertation). Grand Canyon University.
- \*Wilson, W., Jr. (2014). *An exploration of leadership style among country chief administrative officers in northwest louisiana* (Unpublished doctoral dissertation). Northcentral University.
- \*Wohlers, A. J., & London, M. (1989). Ratings of managerial characteristics: Evaluation difficulty, co-worker agreement, and self-awareness. *Personnel Psychology, 42*, 235-261.
- \*Xu, Z. X. (2010). *The relationship between principal's perceived leadership behaviors and teachers' perceptions* (Unpublished doctoral dissertation). Marian University.
- \*Yammarino, F. J., & Dubinsky, A. J. (1994). Transformational leadership theory: Using levels of analysis to determine boundary conditions. *Personnel Psychology, 47*, 787-811.
- \*Yammarino, F. J., Dubinsky, A. J., Comer, L. B., & Jolson, M. A. (1997). Women and transformational and contingent reward leadership: A multiple-levels-of-analysis perspective. *Academy of Management Journal, 40*, 205-222.
- Yukl, G. A. (2010). *Leadership in organizations*. Upper Saddle River, NJ: Prentice Hall.
- \*Zimmerman, R. D., Mount, M. K., & Goff, Maynard, I., II. (2008). Multisource feedback and leaders' goal performance: Moderating effects of rating purpose, rater perspective, and performance dimension. *International Journal of Selection and Assessment, 16*, 121-133. doi:<http://dx.doi.org/10.1111/j.1468-2389.2008.00417.x>