

A PRELIMINARY INVESTIGATION OF THE PSYCHOMETRIC CHARACTERISTICS OF
TRANSFORMATIONAL BOSSES' PERFORMANCE EVALUATIONS

by

SEAN P. BALDWIN

(Under the Direction of Brian J. Hoffman)

ABSTRACT

Despite substantial attention focusing on the nature and consequences of transformational leadership, extant research has not examined transformational leaders in the context of their administrative responsibilities. The current study takes a step in this direction by investigating the psychometric properties of 105 transformational bosses' evaluations of their subordinates' performance. Results suggest that transformational leaders' ratings tend to be elevated and in agreement with other sources, yet these effects only held for their ratings of interpersonal performance dimensions. These results are discussed in the context of the information processing of transformational leaders and future directions of research are offered.

INDEX WORDS: Transformational Leadership, Rating Characteristics, Performance Appraisal, Classic Rating Errors, Leadership

A PRELIMINARY INVESTIGATION OF THE PSYCHOMETRIC CHARACTERISTICS OF
TRANSFORMATIONAL BOSSES' PERFORMANCE EVALUATIONS

by

SEAN PATRICK BALDWIN

B.S. University of Georgia, 2007

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

MASTER OF SCIENCE

ATHENS, GEORGIA

2009

© 2009

Sean Patrick Baldwin

All Rights Reserved

A PRELIMINARY INVESTIGATION OF THE PSYCHOMETRIC CHARACTERISTICS OF
TRANSFORMATIONAL BOSSES' PERFORMANCE EVALUATIONS

by

SEAN PATRICK BALDWIN

Major Professor: Brian J. Hoffman

Committee: Charles E. Lance
Lillian T. Eby

Electronic Version Approved:

Maureen Grasso
Dean of the Graduate School
The University of Georgia
December 2009

DEDICATION

I dedicate this document to my wife and my family. There is no doubt in my mind that this project would not be what it is without Calene. During the months of data management, analysis, writing, revising, and editing, you have always been with me to offer support, insight, and patience. Said another way, you have read every version of this document and heard every presentation. Thank you so much. I also thank my Mom, my Dad, Justin, and Aimee for your encouragement, understanding, and love. From my perspective, no accomplishment exists without family. Thanks you guys!

ACKNOWLEDGEMENTS

First and foremost, I want to acknowledge Brian Hoffman for the incredible amount of guidance and support he provided throughout this project. His ambition, work ethic, and mentorship have greatly influenced my professional and personal development and for that I am beyond grateful. Second, I would like to thank my committee members Charles Lance and Lillian Eby for their insightful comments at each stage of the project. Their sage advice, whether disclosed during casual conversations or during their respective doctoral seminars, has had an enormous impact on the final version of this document. Finally, I want to acknowledge my colleagues and my friends: my cohort. Sara Curtis, Bryan Dawson, Taylor Sparks, Lindsay Johnson, and Mia Tran have together acted as a network of knowledge, friendship, and support, without which this document would have been impossible.

TABLE OF CONTENTS

	Page
ACKNOWLEDGEMENTS.....	v
LIST OF TABLES.....	viii
LIST OF FIGURES.....	ix
CHAPTER	
1 INTRODUCTION	1
2 LITERATURE REVIEW AND HYPOTHESES	4
Transformational Leadership	4
Psychometric Rating Characteristics	5
Individual Difference Correlates of Rating Characteristics	8
Transformational Bosses as Performance Evaluators	11
3 METHOD	15
Participants and Procedure	15
Measures	16
Psychometric Characteristics of Ratings	17
4 RESULTS	21
Factor Analysis	21
Primary Analyses	22
5 DISCUSSION.....	27
Main Findings	27
Implications and Avenues for Future Research	29

Limitations	32
Conclusion	34
Footnotes.....	35
REFERENCES	36

LIST OF TABLES

	Page
Table 4.1: Confirmatory Factor Analysis Results	25
Table 4.2: Correlation Matrix for Variables of Interest	26

LIST OF FIGURES

	Page
Figure 3.1: Flow Chart of Data Structure	20

CHAPTER 1

INTRODUCTION

Bass' (1985) translation of Burns' (1978) transformational leadership theory has become one of the most researched leadership paradigms in recent years. A keyword search for transformational leadership in the PSYCINFO database reveals that as of 4 September 2009, over 1200 related documents had been published on the topic, with nearly 80 percent published in the last ten years. Toward this end, numerous studies document the positive effects of transformational leaders on their workgroups. For example, their employees engage in higher task performance (Judge & Piccolo, 2004; Piccolo & Colquitt, 2006), increased organizational citizenship behaviors (Piccolo & Colquitt, 2006; Podsakoff, MacKenzie, Moorman, & Fetter, 1990), and are intrinsically, rather than extrinsically motivated to work (Dvir, Eden, Avolio, & Shamir, 2002; Piccolo & Colquitt, 2006). Furthermore, transformational leadership is associated with higher levels of group and organization performance (Judge & Piccolo, 2004).

Transformational leaders are visionaries whose unorthodox and idealized approaches positively influence the development and productivity of their workgroup (Dvir et al., 2002) and their followers' values (Hoffman, Bynum, Piccolo, & Williams, 2008; Shamir, House, & Arthur, 1993). Consistent with the heroic bias in leadership research (Meindl & Ehrlich, 1987; Yukl, 2006), much of the transformational leadership literature has focused on the extraordinary effects of transformational leaders. However, when not inspiring followers (Bass, 1985), defying the status quo (Conger & Kanungo, 1998), redefining domestic "Living" (Glynn & Dowd, 2008), influencing the political course of nations (Bass, Avolio, & Goodheim, 1987; Shamir, House, &

Arthur, 1993), saving lives (Hendel, Fish, & Galon, 2005), saving souls (Rowold, 2008), and engaging in generally great and admirable pursuits, transformational leaders typically operate in organizational bureaucracies with more mundane administrative responsibilities. Among these administrative responsibilities, the appraisal of direct report performance is a frequently applied and researched task completed by organizational managers. Given that performance appraisal is a task commonly associated with managerial work, it is somewhat surprising that very limited existing research has integrated the leadership and performance rating literatures. To further emphasize the administrative, rather than the visionary-role of transformational leaders in organizations, I will refer this group of leaders as *transformational bosses* for the remainder of the study.

In order to provide the first examination of the psychometric rating characteristics of transformational bosses' ratings of their direct reports, this study integrates both the performance rating and transformational leadership literatures. In particular, I examine the degree to which transformational bosses' evaluations of their followers' performance are characterized by leniency, halo effect, and agreement with other raters. In doing so, this research contributes to the literature not only by answering recent calls in the performance appraisal literature to investigate rater characteristics that systematically impact performance ratings (Murphy, 2008), but further allows for inferences with respect to how transformational bosses interact with their followers and cognitively represent their environment. To reach this end, this paper is organized around four primary topics. First, the nature of transformational leadership is discussed. Then, a variety of psychometric characteristics of ratings are outlined. Next, research linking individual differences to psychometric properties of ratings is summarized. Finally, these separate streams of research are integrated in the development of hypotheses and research questions focusing on

the psychometric properties of transformational bosses' evaluations of their followers' performance.

CHAPTER 2

LITERATURE REVIEW AND HYPOTHESES

Transformational Leadership

Transformational bosses provide followers with a sense of purpose towards everyday tasks that transcends short-term goals, extrinsic exchanges, and individual needs by emphasizing collective values, higher order needs, and long-term growth (Bass, 1985; Hoffman et al., 2008; Shamir et al., 1993). Bass (1985) conceptualized this leadership theory as consisting of four dimensions of behavior, including individualized consideration, inspirational motivation, intellectual stimulation, and idealized influence. *Individualized consideration* refers to managers' willingness to devote time and energy to the professional development of their employees. *Inspirational motivation* is a behavioral dimension associated with the articulation of shared experiences and values. These leaders emphasize the importance of working toward a common set of goals. Leaders who challenge their followers to view problems through new or unorthodox perspectives exemplify the behavioral dimension *intellectual stimulation*. Lastly, *idealized influence* (also referred to as charisma) is conceptualized as both a behavioral and attributional phenomenon that is evidenced by leaders projecting confidence in their groups and arousing strong emotions in their followers. By articulating a compelling, value-laden vision of the future, demonstrating confidence in followers' ability to achieve this vision, developing followers to their maximum potential, and challenging them to consider new perspectives to old problems, transformational bosses have a powerful motivational impact on their followers (Fuller,

Patterson, Hester, & Stringer, 1996), positively influence workgroup attitudes (Kirkpatrick & Locke, 1996), and ultimately lead more effective workgroups (Geyer & Steyrer, 1998; Hoffman et al., 2008; Judge & Piccolo, 2004). The positive impact of transformational leadership is documented in a variety of settings, including the military, education, and business (Judge & Piccolo, 2004).

In the clamor and excitement surrounding the extraordinary effects and characteristics of transformational bosses, research has overlooked other administrative roles and tasks common to the leadership role. For instance, research has not considered the way in which transformational bosses schedule and focus their work, develop strategies and contingencies for future events, network with colleagues, complete the tasks associated with their profession, and evaluate their followers' performance. Because performance appraisals are commonplace for the practicing manager, a necessary factor to organizational functioning and success (Ghorpade & Chen, 1995), and a natural extension of the leadership role, an integration of the transformational leadership and performance appraisal literatures can offer a variety of insights with respect to the transformational leadership process. Accordingly, this study uses transformational bosses' appraisals of their followers' performance as a means to view the leadership process.

Psychometric Rating Characteristics

Since psychometric rating characteristics were first documented in subjective measures over a century ago (Wells, 1907), much attention has been paid to their impact on subjective evaluations (Hoffman, Lance, Bynum, & Gentry, in press). For instance, Thorndike (1920) coined the term halo to describe unrealistically strong inter-dimension correlations. Building on Thorndike's (1920) work, Kingsbury (1922) examined three psychometric rating characteristics: halo, leniency, and the fear of making distinctions between rates (i.e., a precursor to range

restriction). Operating under the assumption that halo, leniency, and range restriction were inversely related to accuracy, these idiosyncratic *rater biases* were the most frequently used criteria when evaluating performance ratings for the better part of the 20th century (Austin & Villanova, 1992). Although idiosyncratic rater variance was originally viewed as bias, more recently it has been recognized that idiosyncratic rater effects may not represent bias at all (Hoffman & Woehr, in press; Lance, Hoffman, Gentry, & Baranik, 2008); rather, building evidence suggests that they reflect substantively meaningful variance and an important component of understanding performance evaluations. Bias or not, idiosyncratic rating patterns are important foci for continued study (Cardy & Dobbins, 1994; Kane, Bernardin, Villanova, Peyrefitte, 1995; Murphy & Anhalt, 1992; Tziner, Murphy, & Cleveland, 2002).

The psychometric rating characteristics most commonly used to describe ratings include distributional rating characteristics, halo, and agreement. Two commonly used distributional rating characteristics are *leniency* and *severity*. These were traditionally conceptualized as higher or lower ratings than warranted (Saal & Landy, 1977) and, therefore, viewed as a form of error. However, in the absence of a performance true score for comparison, it is impossible to determine whether observed leniency/severity effects reflect bias or an accurate depiction of ratee performance (Cardy & Dobbins, 1994). Given that true scores are impractical in field settings, performance evaluation research has adapted the definition of leniency and severity to reflect higher or lower ratings relative to some set point on the scale (e.g., the midpoint, the mean of others in the sample, or the mean of other raters of the same target), and refer to these patterns of ratings as rating elevation and rating depression, respectively (Bernardin, Cooke, & Villanova, 2000). Because the current study could only be conducted in a field setting, negating the possibility of establishing performance true scores, the terms rating elevation and depression

were adopted to describe my use of set points on a scale (rather than performance true scores) to investigate the distributional properties of ratings¹. Halo effect is among the more pervasive of the idiosyncratic rater effects, both in terms of suspected impact on ratings and proportional attention in the performance rating literature. Although operationalized using a variety of different approaches, conceptually, *halo effect* refers to the persistent finding of low variance in a single rater's ratings of multiple, ostensibly distinct dimensions (Cooper, 1981). In contrast to the distributional rating characteristics previously described, halo is not a function of the ratings' level; instead, it is a function of the amount of variance in ratings of different dimensions. As with distributional rating characteristics, the lack of performance true scores precludes one's ability to ascertain the extent to which observed halo effects are a function of true score level covariance among constructs (true halo) or bias (illusory halo; Cooper, 1981).

Previous research has attempted to identify the explanatory mechanisms of halo effect (Lance, LaPointe, & Stewart, 1994) and also developed a variety of approaches to calculate it (Cooper, 1981; Fisicaro, 1988; Saal, Downey, & Lahey, 1980; Solomonson & Lance, 1997; Viswesvaran, Schmidt, & Ones, 2005). Of the variety of mechanisms forwarded to explain halo effect (e.g., rater general impression, salient dimensions of the ratee, and raters' inadequate discrimination amongst performance dimensions), prior research suggests that a rater's general impression is the most appropriate conceptualization, even in research specifically designed to elicit other explanatory mechanisms (Lance et al., 1994; Lance, Foster, Gentry, Thoresen, 2004). Halo effect has been operationalized via (a) the examination of intercategory correlations (Thorndike, 1920), (b) the examination of intraratee variance across dimensions, (c) a single factor in confirmatory factor analysis, and (d) a significant interaction effect between ratee, rater, and dimension, evidenced by analysis of variance (Cooper, 1981). Regardless of the specific

explanatory model one prescribes to or the calculation used, conceptually, halo involves the degree to which ratings of ostensibly distinct dimensions co-vary.

In addition to elevation and halo effect, the rating literature has relied on interrater agreement as criteria by which to evaluate the quality of subjective measures. *Interrater agreement* is conceptualized as the degree to which multiple raters provide the same or similar ratings of an individual's performance (Lahey, Downey, & Saal, 1983). Therefore, interrater agreement is a function of agreement on the level of ratings, while interrater reliability is a function of the correspondence in the pattern of ratings (Lahey et al., 1983). The current study focuses on interrater agreement, rather than reliability, because my concern is the extent to which the absolute level of ratings (not rank-order) corresponds across raters (Lahey et al., 1983).

Individual Difference Correlates of Rating Characteristics

Although the desire to reduce the impact of elevation and halo while increasing cross-rater agreement spawned decades of research attempting to “fix” performance appraisal systems through the use of different rating scales and rater training, these interventions were shown to have a modest impact on rating accuracy (Austin & Villanova, 1992; Feldman, 1981; Hedge & Kavanagh, 1988). One proposed reason for the unimpressive results associated with design fixes is that psychometric characteristics of ratings are impacted as much by stable rater characteristics as the qualities of the appraisal system (Murphy, 2008). Toward this end, Hoffman et al. (in press) found that idiosyncratic rater effects account for an average of 55% of the variance in performance ratings across two large samples. In addition, mounting evidence suggests that these pervasive rater effects are partly attributable to relatively stable rater characteristics.

Providing preliminary evidence for the role that individual differences play in the psychometric qualities of ratings, Kane and colleagues (1995) found that within raters, rating

elevation is a stable characteristic over time, supporting the possibility that elevation is caused by rater individual differences. In recognition of the role that the rater plays in the performance evaluation process, some research has focused on investigating the degree to which specific rater individual differences predict rating elevation. For instance, both Antonioni and Park (2001) and Isen, Shalcker, Clark, and Karp (1978) found that raters with high levels of positive affect tend to provide more elevated ratings relative to raters with lower levels of positive affect. Other research has proposed more complex relationships. For instance, Bernardin and colleagues (2000) hypothesized a relationship between agreeableness and rating elevation, moderated by conscientiousness. Their results supported the hypothesis, such that the most elevated ratings were given by agreeable raters with low levels of conscientiousness. Together, prior research linking individual differences to systematic variation in elevation in observed ratings underscores the role that systematic rater characteristics play in rating elevation.

Although limited in scope and volume, existing research has also substantiated the impact of rater individual differences on the amount of halo in performance ratings. Borman (1977) examined the consistency of psychometric rating characteristics (i.e., leniency, halo, and range restriction) in college students' ratings of different performance incidents. Of the psychometric characteristics examined, halo effect was the most consistent within raters across different performance incidents. To identify the role of rater individual differences in haloed ratings, Tziner and colleagues (2002) examined the interaction between raters' orientations toward performance appraisals and conscientiousness in predicting the degree of halo in their performance ratings. Their results suggested a main effect of a rater's conscientiousness and orientation toward performance appraisal on the degree of halo in ratings, with more conscientious raters and those with a higher orientation toward performance appraisal providing

less haloed ratings. Together, these findings substantiate the proposition that stable rater characteristics play a role in the degree of halo in performance ratings.

Finally, some research has evidenced the role that rater individual differences play in delivering ratings that are in agreement with others'. Vingoe and Antonoff (1968) examined the impact of rater personality on interrater agreement using dormitory roommates' ratings of other students' personalities and demonstrated that well-adjusted, tolerant, self-controlled and introverted raters tended to provide ratings that were in a higher level of agreement with their peers. Hjelle (1969) functionally replicated these findings. Finally, Edison and Adams (1992) studied the agreement of depressed and non-depressed individuals with trained raters' ratings. In this experiment, the participants met for the first time, engaged each other in conversation, and then judged their own social-performance as well as their partner's. Results showed that higher agreement with the objective ratings was provided by depressed rather than non-depressed individuals.

In sum, prior research consistently reveals that idiosyncratic rater effects are pervasive in performance ratings (Hoffman et al., in press; Scullen, Mount, & Goff, 2000) and to some degree are stable within raters (Borman, 1977; Kane et al., 1995). By extension of these findings, others have substantiated the impact of rater individual differences on the psychometric characteristics of ratings. Given that those providing performance ratings are typically organizational leaders, leadership variables may also relate to the psychometric characteristics of ratings. Toward this end, the current study addresses this gap in the literature by examining the degree to which transformational bosses' ratings of their subordinates are characterized by rating elevation, halo effect, and interrater agreement.

Transformational Bosses as Performance Evaluators

Previous leadership literature suggests that bosses' evaluations of follower performance might be characterized by various stable trends. Although this idea has never been investigated in the transformational leadership literature, it has been examined in the context of Fiedler's (1971) Least Preferred Co-Worker (LPC) Scale. In a review of the LPC Scale, Rice (1978) examined the main effects of LPC orientation on performance ratings of other group members and found that high-LPC persons² generally rate higher than low-LPC persons. More specifically, 90% of the published effects that were reviewed were in this direction. Although this question has never been examined in the context of modern leadership theories, these results imply that leadership constructs can have an impact on performance ratings.

A close look at transformational leadership literature suggests that transformational bosses will provide elevated ratings. For instance, transformational bosses articulate optimistic visions (Bass & Avolio, 1995) and tend to view risks favorably (Conger & Kanungo, 1998), implying a tendency to evaluate other entities more positively. Indeed, past research suggests that extraversion, a construct closely aligned with positive affect (Lucas, Le, & Dyrenforth, 2008), is one of the strongest, most consistent personality predictors of transformational leadership (Bass, 1985; Bono & Judge, 2004). Next, the behaviors associated with transformational leadership are thought to influence follower self-confidence, in turn increasing the employee's intrinsic motivation. Shamir and colleagues (1993) specifically proposed that charismatic bosses "increase self-efficacy and collective efficacy through expressing *positive evaluations*...showing confidence in followers ability to meet such expectations" (p.584, italics added). Shamir and colleagues' (1993) proposition is indirectly supported by evidence that transformational bosses tend to be less likely to provide negative feedback, as suggested by the

consistently observed inverse relationship between transformational leadership and management by exception (i.e., intervening to adjust performance deficiencies; Bycio, Hackett, & Allen, 1995). Similarly, the importance of having confidence in one's followers appears in the most common measure of transformational leadership: the Multifactor Leadership Questionnaire (MLQ; Bass & Avolio, 1995). Together, those who form close relationships with the ratees, view ambiguous situations such as crisis situations more positively, are optimistic about the future, are generally in a good mood, and regularly express confidence in their ratees will likely evaluate others more favorably.

Hypothesis 1: Transformational leadership will be positively related to rating elevation.

The behaviors and individual characteristics associated with transformational leadership also point to a potential relationship between transformational leadership and haloed ratings. Again, these relationships have yet to be examined by the transformational leadership literature, but have been studied in the context of Fiedler's (1971) LPC scale. In a review of the LPC literature, Rice (1978) found that person-oriented leaders are less likely to differentiate ostensibly distinct performance dimensions when evaluating follower performance. Characteristics associated with transformational bosses in the extant research suggest that transformational bosses might provide haloed ratings. For example, research suggests rater-ratee familiarity is positively related to higher levels of halo (Van Scotter, Moustafa, Burnett, & Michael, 2007). To the extent then that transformational bosses form closer relationships with followers (Bass, 1985), it is possible that their ratings will be characterized by higher levels of halo. Alternatively, it is also possible that the extraverted nature of transformational bosses could lead to decreased levels of halo in their ratings. Given their awareness and genuine concern for subordinates' professional development (i.e., individualized consideration), transformational

bosses' ratings should be characterized by more differentiation of follower strengths and weaknesses and, consequently, less halo. Further, extraverted managers are likely to have large, yet close knit networks (Forret & Dougherty, 2001), and as a result have access to more diversified information. With this knowledge, transformational bosses should have comparatively holistic, differentiated views of their employees relative to other managers. Given these conflicting theoretical propositions, the relationship between transformational leadership and halo in ratings was exploratory.

Research Question 1: Will transformational bosses' ratings be characterized by halo?

Next, prior leadership research has indirect implications for a possible relationship between transformational leadership and interrater agreement. Transformational bosses' ability to effectively motivate their employees is at the heart of transformational theory (Bass, 1985; Yukl, 2006). Given that performance appraisals are commonly used as motivational tools (Longenecker et al, 1987; Murphy, 2008), it is possible that transformational bosses will use performance ratings to motivate followers- intentionally inflating or deflating ratings. Performance ratings that are intentionally distorted are unlikely to be in agreement with other raters' ratings, which would imply an inverse relationship between transformational leadership and interrater agreement.

Alternatively, prior research implies a positive relationship between transformational leadership and interrater agreement. A key component of transformational leadership theory is a leader's recognition of opportunities and the development and articulation of a strategic vision that capitalizes on these opportunities (Bass, 1985; Conger & Kanungo, 1998; Kirkpatrick & Locke, 1996). The fact that strategic visions are a cornerstone to transformational leadership theory speaks strongly to transformational bosses' keen sense of situational awareness

(Waldman, Ramirez, House, and Puranam, 2001). These are bosses who are both intelligent and environmentally aware and, therefore, are particularly effective at diagnosing their environment (Hoffman & Frost, 2006). It is conceivable then that this high degree of situational awareness will translate to the evaluation of their followers' strengths and weaknesses and ultimately result in agreement with other raters. This idea is further supported by early transformational leadership theory (Bass, 1985), which characterized these bosses as individually considerate individuals who are aware of follower abilities and work to further develop deficiencies. Together, previous research is unclear about the relationship between transformational leadership and interrater agreement. Therefore, I will address the relationship with a research question.

Research Question 2: Will transformational bosses' performance ratings be in agreement or disagreement with others' ratings?

The concept of performance is described in various content areas with various definitions, yet across domains and occupations, the distinction between two broad performance domains corresponding roughly to task and interpersonal performance consistently emerges (Borman & Motowidlo, 1997; Fleishman, 1957; Hoffman, Blair, Woehr, & Meriac, 2007; Kram, 1985; Noe, 1988; Shore, Thornton, and Shore, 1990). Accordingly, this study: (a) conceptualizes performance as a multidimensional construct space, encompassing both interpersonal and task performance, and (b) investigates differences in rating characteristics based on the particular performance dimension being rated.

Research Question 3: Does the performance dimension being rated impact the relationship between transformational leadership and the psychometric characteristics of ratings?

CHAPTER 3

METHOD

Participants & Procedure

The sample was drawn from (N=105) managers enrolled in an Executive Masters of Business Administration (EMBA) program at a large southeastern university. Before entering the EMBA program, participants were mailed a 360-degree feedback form and instructed to distribute them to their peers (N=389), supervisor (N=105), and subordinates (N=386). Figure 3.1 depicts the data collected and specifies the raters and ratees of each relationship. All coworkers (peers, higher level managers, and subordinates) were asked to rate the EMBA manager on a variety of behavioral performance dimensions- that is, the EMBA managers were the target of performance ratings from their boss, peers, and subordinates. EMBA managers also rated their supervisor's transformational leadership. These ratings were used to operationalize transformational leadership. In turn, the supervisor's downward ratings of their follower (the target EMBA manager) were used to examine the qualities of transformational bosses' ratings. Finally, to construct a referent with which to calculate rating elevation and agreement, others' (peer and subordinate) ratings of the target EMBA manager were collapsed to form a single referent indicator. To ensure that aggregating ratings between peers and subordinates was appropriate for the current dataset, I calculated the average interrater agreement across sources ($r^*_{wg(J)}=.95$) and tested for metric invariance between sources ($\Delta\chi^2 = 4.54, ns$). Results suggested that the ratings across sources were in high agreement and that the sources conceptualized

performance similarly (see Table 4.1). The decision to collapse each of the individual raters into broad groups corresponding to the raters' relationship with the target was based on the desire to maximize the reliability of my referent indicator and because differences in rating levels across sources were not central to this study³. There was an average of 7.4 other ratings per target.

Measures

Transformational Leadership. A modified version of the MLQ (Bass & Avolio, 1995) was used to measure the supervisor's transformational leadership. The MLQ assesses four dimensions associated with transformational leadership, including: individualized consideration, idealized influence, intellectual stimulation, and inspirational motivation. Each of the dimensions was assessed with four items rated on a 5-point scale ranging from 1 (*strongly disagree*) to 5 (*strongly agree*). Because prior research has consistently supported a single factor structure of transformational leadership (Judge & Piccolo, 2004; Piccolo & Colquitt, 2006; Tajeda, Scandura, & Pillai, 2001), each of the transformational scales was expected to load on a single latent transformational factor.

Performance Ratings. Higher level managers' ratings of their follower's (the EMBA participant) performance were used to examine the psychometric properties of bosses' ratings. The EMBA participants' peers' and subordinates' ratings of the EMBA participants' performance were used as a referent with which to examine the psychometric properties of transformational bosses' performance evaluations. Performance dimensions were classified into interpersonal or task-style categories. Task performance was measured with analysis, judgment and decision-making, planning and organizing, public speaking, and initiative, while interpersonal performance was assessed with team-building and confrontation skill. Respondents were asked to make ratings on a 5-point scale ranging from 1 (*strongly disagree*) to 5 (*strongly*

agree). There was also a response option for undecided respondents: 6 (*don't know/unable to observe*).

Psychometric Characteristics of Ratings

Rating Elevation. Since performance true scores are virtually unattainable in field settings, elevation was operationalized using two referents derived from the performance rating data. The first referent used to calculate rating elevation was operationalized as the deviation of the higher level managers' downward ratings from the mean ratings provided by all higher level managers in the sample (Bartels & Doverspike, 1997; Bernardin, Alvares, & Cranny, 1976; Bernardin & Orban, 1990). This is referred to as sample referent elevation and is indicative of how elevated the higher level manager's ratings are in reference to all managers in the sample. To calculate this form of elevation, the higher level manager's downward ratings were subtracted from the mean downward ratings of all managers in the sample. Due to the possibility of negative elevation scores, this variable was standardized using t-scores for more intuitive interpretation.

Although the sample-referent approach to calculating rating elevation is one of the more popular operationalizations of elevation (Saal et al., 1980), this method has two limitations. First, the ratings used in this approach are of completely different targets, which make it difficult to determine if this form of elevation is a function of the rater or the ratee (Kane, et al., 1995). Second, transformational leadership research consistently suggests that transformational bosses have high performing subordinates (Judge & Piccolo, 2004). Thus, the hypothesized positive correlation between elevation and transformational leadership could be a function of transformational bosses actually having more effective employees, as opposed to effects introduced by the rater. In other words, the transformational bosses' ratings would not be

inflated; rather, they would represent the frequently substantiated positive impact of transformational leadership on workgroup productivity. Given these limitations, an additional measure of elevation was used.

The second operationalization of elevation used the mean of subordinate and peer ratings of the same ratee as a referent. By using ratings of the same target as a referent, rater-referent elevation reduces concerns about confounds in traditional measures of elevation (Kane, et al., 1995). Specifically, if the positive relationship between transformational leadership and elevation is due to the impact of transformational leadership on follower performance, one would expect an increase in peer and subordinate ratings of the target manager as well. To the degree that the higher level managers' ratings are higher than other raters' ratings of the exact same target, confidence is given to this effect as introduced by the rater, not the ratee.

Halo effect. Consistent with prior research (Bernardin & Walter, 1977; Borman, 1977; Fisicaro, 1988; Woehr & Lance, 1991), halo was operationalized as the standard deviation of the higher level manager's downward performance ratings of the target EMBA manager on all 7 performance dimensions. This method indexes halo effect as the inverse degree of variability in a single rater's ratings and does not make a direct true score comparison. Using this approach to calculating halo, a lower standard deviation (e.g. less variance) among the higher level manager's ratings evidences an increased halo effect, while a greater standard deviation is indicative of less halo effect in ratings. Halo was only calculated for the overall ratings, as opposed to the two performance dimension-levels, because the nature of halo involves examining the covariance of ratings of distinct domains.

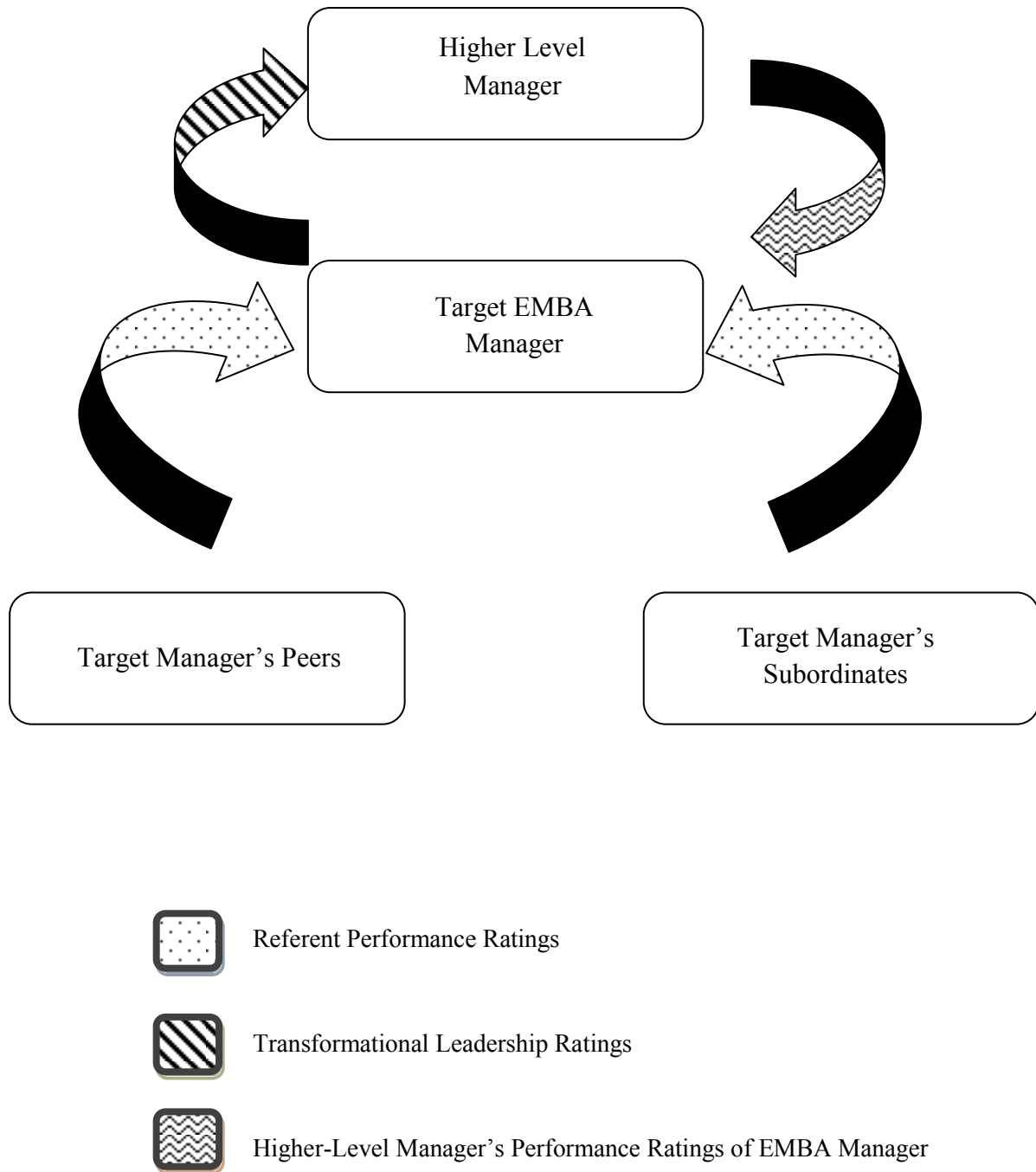
Agreement. James, Demaree, & Wolf's (1984) $r_{wg(J)}$ statistic indexes the level of agreement within a group of individuals on a multidimensional measure and typically ranges

from 0 to 1. This method of calculating agreement compares observed variance within a group to a hypothetical distribution of the amount of variance that would be expected due to chance alone. In a Monte Carlo study of r_{wg} , Lindell, Brandy, and Whitney (1999) examined four of its derivations and consistently found that $r_{wg(J)}^*$ provided the most accurate representation of the true dataset. Also, because this version does not include any representation of the Pearson-Brown attenuation formula as did earlier versions, it is a more theoretically meaningful model. Accordingly, this operationalization of interrater agreement was adopted. For the current study, it is expected that if ratings were assigned wholly due to chance, each source would be equally likely to choose each of the response options. As such, the hypothetical distribution of variance due to chance will be represented by a uniform, rectangular distribution.

A $r_{wg(J)}^*$ value was calculated to index the level of agreement between each higher level manager and the other raters rating the same target EMBA participant. The resulting index of within-group interrater agreement is indicative of the degree to which the higher level manager's ratings correspond to the rating level evidenced by other raters of the same target EMBA participant.

Figure 3.1

Flow Chart of Data Structure.



CHAPTER 4

RESULTS

Factor Analysis

LISREL 8.7 (Jöreskog & Sörbom, 2004) was used to examine the structure of the study variables. It was expected that task and interpersonal performance would provide the best representation of the performance ratings (Borman & Motowidlo, 1997; Fleishman, 1957; Hoffman, et al., 2007; Kram, 1985; Noe, 1988; Shore, Thornton, and Shore, 1990). I also expected that a single factor, transformational leadership, would most closely approximate the transformational ratings (Bono & Judge, 2004; Tejada, et al., 2001). As expected, confirmatory factor analysis revealed (a) that a two factor performance model including task and interpersonal-related performance for both higher level manager and other raters fit the data adequately in an absolute sense and significantly more closely than a single factor performance model for each set of raters ($\Delta\chi^2(5) = 37.04, p < .001$) and (b) that transformational leadership can be described using a single factor model (Table 4.1).

Given that my questions necessitate a comparison of performance ratings provided by different raters (i.e., supervisors and referent others), it was necessary to ensure that the raters (supervisors and others) conceptualize performance similarly. Accordingly, I investigated the extent to which the performance ratings were invariant to the source providing the ratings following the procedure recommended by Vandenberg and Lance (2000). Based on the acceptable fit for the two performance dimensions for each rater group, configural variance was supported. In addition, the metric invariance model did not yield a significant decrement in

model fit relative to the configural invariance model, providing support for metric invariance ($\Delta\chi^2(5) = 6.64, ns$). Together, consistent with prior research (Woehr, Sheehan, & Bennett, 2005), the invariance analyses suggest that higher level managers and other raters are evaluating the same performance constructs on the same measurement scale, and accordingly, that the ratings are comparable across groups. Thus, I proceeded with the primary analyses.

Primary Analyses

Pearson's correlation coefficient was used to test the hypothesis and research questions. Table 4.2 reports the mean of each study variable and correlations amongst them. The upward transformational leadership ratings ranged from 1.5 to 4.92, with an average of 3.69, suggesting that there was sufficient variability in the predictor to proceed with primary analyses. The reliability coefficients of performance ratings from each source were acceptable. Supervisory ratings of task and interpersonal performance yielded alpha coefficients of .88 and .85 respectively, while the others' ratings of task and interpersonal performance yielded alpha coefficient of .92 and .81, respectively. The reliability coefficient of transformational leadership was .89. Hypothesis 1 predicted that transformational bosses would provide elevated ratings of their followers' performance. The first measure, sample-referent elevation, compared higher level managers' performance ratings to the mean level of downward ratings across the entire sample. Results indicated that transformational bosses provide more elevated ratings of their followers' performance (with reference to other bosses' ratings of performance) than less transformational higher level managers on interpersonal-related performance, ($r = .31, p < .01$), but not on task-related performance ($r = .15, ns$). Despite the initial support for the hypothesis, this form of elevation confounds rater elevation with the often supported impact of transformational leadership on follower performance with rating elevation. Accordingly, I also

operationalized elevation as the difference between the downward ratings provided by each boss and ratings of the same ratee provided by other raters. Transformational bosses provided more elevated ratings (with reference to others' ratings of the same target) than their less transformational counterparts when rating interpersonal ($r = .27, p < .01$), but not task-related job performance ($r = .03, p = ns$). To address the possibility that transformational bosses' ratings might differ depending on the performance dimension being rated, I used the methods proposed by Meng, Rosenthal, and Rubin (1992) for comparing dependent correlations. Specifically, the correlation between interpersonal rating elevation and transformational leadership was compared with the correlation between administrative rating elevation and transformational leadership. Results of the test showed that transformational bosses rate their subordinates with significantly higher levels of elevation for interpersonal, compared to task performance for both sample ($t = 2.17, p < .05$) and rater referent elevation ($t = 2.50, p < .05$). Together, Hypothesis 1 was partially supported. The second psychometric characteristic of interest was halo effect. Results indicated a negative, albeit non-significant relationship between transformational leadership and halo effect ($r = -.12, p = ns$). Thus, in response to Research Question 1, transformational leadership is not related to halo effect.

The next research question focused on the degree to which transformational bosses' ratings are in agreement with evaluations of the target by coworkers from other organizational levels. Results supported a positive relationship between transformational leadership and within group agreement for interpersonal ($r = .29, p < .01$), but not task performance ($r = .03, ns$). Given that $r^*_{wg(j)}$ calculations take into account the mean level of ratings when forming the hypothetical variance distribution, there is a concern that interrater agreement and elevation will be confounded- a possibility supported by the moderate relationship between agreement and

elevation. To address this possibility, I examined the correlation between transformational leadership and agreement, holding constant the effects of elevation. The relationship between transformational leadership and agreement on interpersonal performance ratings approached significance when the effect of rater-referent elevation was removed ($r = .17, p = .07$) and transformational leadership remained a non-significant correlate of agreement on task dimensions with the effect of rater-referent elevation removed ($r = .01, ns$). Although the partial correlation between transformational leadership and agreement was greater for interpersonal than for task dimensions, the difference was arguably not statistically significant by conventional standards ($t = 1.65, p = .10$).

Table 4.1

Model Fit Statistics & Invariance Tests.

	$\Delta\chi^2$	df	χ^2	RMSEA	TLI	CFI	SRMR
Models of Performance & Transformational Leadership							
1. Transformational Leadership: One factor		2	.58	<.01	1.00	1.00	.01
2. Performance: One factor per rater		69	161.54	.11	.93	.90	.06
3. Performance: Two factors per rater		64	124.50	.09	.93	.95	.06
Model 4 vs. Model 5	37.04**	(5)					
Peer-Subordinate Invariance							
4. Configural Invariance		64	130.72	.10	.93	.95	.05
5. Metric Invariance		69	135.26	.09	.93	.95	.08
Model 1 vs. Model 2	4.54	(5)					
Boss-Other Invariance							
6. Configural Invariance ¹		64	124.50	.09	.93	.95	.06
7. Metric Invariance		69	131.14	.09	.93	.95	.07
Model 6 vs. Model 7	6.64	(5)					
Note. * p < .05; ** p < .01; ¹ Model 3 was used as the configural invariance model and as such, the baseline model for the metric invariance test RMSEA= Root Mean Square Error of Approximation; TLI= Tucker Lewis Index; CFI = Comparative Fit Index; Value in parentheses is Δ df for that model comparison.							

Table 4.2

Correlations among Study Variables.

Factor	Mean	1	2	3	4	5	6	7
1. Transformational Leadership	3.69	-						
2. Interpersonal Performance Elevation (s)	50.0	.31**	-					
3. Task Performance Elevation (s)	50.0	.15	.71**	-				
4. Interpersonal Performance Elevation (r)	48.20	.27**	.89**	.53**	-			
5. Task Performance Elevation (r)	48.99	.03	.44**	.77**	.50**	-		
6. Halo Effect	.38	-.12	-.46**	-.47**	-.42**	-.39**	-	
7. Interpersonal Performance Agreement	.88	.29*	.57**	.24*	.59**	.21*	-.46**	-
8. Task Performance Agreement	.91	.03	.21*	.37**	.22**	.41**	-.41**	.47**

Note. * denotes $p < .05$; ** denotes $p < .01$; r denotes rater referent; s denotes sample referent

CHAPTER 5

DISCUSSION

Despite the important role that performance evaluation plays in effective leadership, existing research has rarely considered the impact of leadership on the quality of performance ratings. Accordingly, this study contributes to the literature by providing the first investigation of the psychometric characteristics of transformational bosses' ratings of their followers' performance as well as responding to recent suggestions to investigate the role that rater characteristics play in performance ratings (Murphy, 2008). In addition to enhancing the understanding of leader-follower relations by using a behavioral criterion measure (rating behaviors), the current results also provide an initial glimpse at an underdeveloped area of research: the perceptual, evaluative, and judgmental processes of transformational bosses.

Main Findings

Overall, the results suggest that transformational bosses' ratings are characterized by predictable, stable trends. First, transformational bosses' ratings of follower interpersonal performance were elevated relative to the mean ratings provided across the entire sample as well as ratings provided by raters of the same target. These findings reveal an interesting paradox with theoretical conceptions of transformational leadership. Specifically, individually considerate bosses focus substantial attention on the development of their employees (Bass & Avolio, 1995) by "evaluat[ing] followers' potential both to perform their present job and to hold future positions of greater responsibility" (Bass, 1985, p.10). Thus, in contrast to the expectation that transformational bosses will provide more developmental ratings, these bosses actually evaluate

followers' interpersonal performance more favorably- suggesting some degree of inconsistency between these core components of transformational leadership theory and the actual behaviors displayed by transformational bosses.

On the other hand, it is noteworthy that this pattern of results is quite consistent with the observed relationships between transformational leadership and transactional leadership. Specifically, prior research suggests that transformational leadership ratings are negatively related with ratings of management by exception (Bycio et al., 1995), a dimension associated with giving negative feedback. Although prior research has questioned this finding as due to the biasing impact of rater liking on transformational leadership (Brown & Keeping, 2005), the results of the current study provide the first behavioral evidence that transformational bosses are actually less likely to give negative performance feedback, especially as it related to interpersonal performance domains.

Second, my results suggest that transformational bosses elevate their subordinates' performance ratings for interpersonal dimensions, but not task-related ones. Although not hypothesized, the study's current finding that transformational bosses evaluate their followers favorably on interpersonal dimensions is consistent with conceptualizations of transformational leadership that focus on their optimism, extraversion, and positive emotions (Bass, 1985; Berson, Shamir, Avolio, & Popper, 2001). Indeed, previous research on Leader-Member Exchange theory emphasizes the close interpersonal relationships formed between transformational bosses and their subordinates (Howell & Hall-Merenda, 1999). It seems as though these optimistic and extraverted bosses might allow the close relationships with their followers to color their judgments of followers' interpersonal skills, but this impact does not extend to the evaluation of non-interpersonal aspects of performance. Next results of this study suggest that transformational

leadership and halo effect do not co-vary consistently. It is likely that any observed differences in halo effect amongst transformational bosses in this study were an artifact of other unmeasured variables, such as job or ratee knowledge (Kozlowski, Kirsch, & Chao, 1986).

Lastly, results revealed a positive relationship between transformational leadership and within group agreement for ratings of interpersonal performance. After controlling for the effects of rating elevation, transformational leadership maintained a weak relationship with within-group agreement that approached conventional levels of statistical significance. From the perspective of Classical Test Theory, these results suggest that transformational bosses' provide ratings that are less impacted by random measurement error. These findings support prior theory characterizing transformational bosses as in tune with their followers and with the external environment (Barbuto & Burbach, 2006; Hoffman & Frost, 2006), particularly with regard to interpersonal dimensions. These results suggest that transformational bosses' ratings are less impacted by random measurement error on interpersonal dimensions of performance, relative to task-performance dimensions. Given the close relationships formed between transformational bosses and their followers, it is likely that they are privy to a comparatively large amount of interpersonal performance information and, thus, rate performance with a higher degree of agreement on interpersonal dimensions.

Implications and Avenues for Future Research

This study provides preliminary insight as to how transformational bosses evaluate and judge others' performance and by extension, perceive their environment. In this first perceptual examination, transformational bosses' ratings differed based on the performance category being rated. Although transformational bosses tended to rate with more elevation and in more agreement on interpersonal dimensions, their ratings are neither related to agreement nor

elevation when rating more administrative performance dimensions. The findings have implications for how transformational bosses perceive their environment, the measurement of transformational bosses' behaviors, the importance of investigating transformational leaders in the broader context of their role as managers, and the importance of continuing the investigation of leadership style as a predictor of rating characteristics.

By examining transformational bosses' evaluations of their followers' performance, this study indirectly sheds light on how transformational bosses perceive their environment. For example, it is possible that the current study's finding that transformational bosses provide elevated ratings on interpersonal dimensions could be explained by their optimism. Perhaps transformational bosses' optimistic perspective leads them to form more favorable evaluations of ambiguous stimuli, and it is this increased propensity to form favorable impressions that ultimately results in their expressing confidence in the future and engaging in risk-taking behaviors. On the other hand, transformational bosses' increased propensity to form positive impressions may be associated with a decreased likelihood of providing developmental feedback to followers. To this end, further research on how transformational bosses perceive a variety of environmental factors (e.g. perceptions of organizational climate or evaluations of organizational strength, weaknesses, opportunities, and threats), may provide insight into the perceptions and cognitions in other evaluative situations.

Next, this study is one of few to investigate the actual behaviors of transformational bosses. The vast preponderance of leadership research uses questionnaire-based measures that to some extent, measure follower attributions of leadership rather than actual leader behaviors (Lord, Foti, & DeVader, 1984). By using leader ratings of followers as a criterion variable, I was able to investigate central tenets of transformational theory using a behavioral criterion variable,

allowing for a deeper understanding of how transformational bosses interact with their followers, one unfettered by follower implicit leadership theories. As a practical, yet counterintuitive implication of these findings, transformational bosses may actually need training in giving negative feedback on interpersonal dimensions. Stated another way, these results suggest that transformational bosses might need help in differentiating amongst high and low performers on important organizational performance dimensions, such as team-building or confrontation skills. Similarly, organizations may need to consider the extent to which transformational bosses identify the best possible candidates for developmental assignments and advancement. Although this study sheds light on the main effect of transformational leadership on performance evaluations, the reasons for these effects are unclear. One potentially fruitful avenue for research is to investigate the degree to which transformational bosses have different goals when evaluating their followers' performance (Murphy, Cleveland, Skattebo, & Kinney, 2004; Tziner, Murphy, & Cleveland, 2001).

Next, this study represents one of few attempts to view transformational bosses in the broader context of their role as managers operating in organizational systems. Although the evaluation of follower performance is often included as a primary responsibility among the tasks associated with managerial work (see the Occupational Information Network or taxonomies of managerial performance, e.g. Borman & Brush, 1993), existing leadership research rarely investigates this important aspect of managerial work. Additional research linking transformational leadership with other aspects of managerial work, such as interacting with internal and external customers, developing strategic plans, staffing their workgroup, evaluating business strategies, making decisions, or performing technical job requirements of their specific job could further elucidate the nomological network of transformational leadership.

Finally, numerous primary studies and performance appraisal reviews have suggested the important role of rater characteristics in performance ratings (e.g. Cleveland & Murphy, 1992; Murphy & Cleveland, 1995; Tziner, Murphy, & Cleveland, 2005). Despite significant progress in this area, existing research has almost exclusively focused on broad personality constructs. The current study takes an important first step forward in the performance appraisal literature by providing the first integration of the performance appraisal rater effects and leadership literatures. Future research investigating other rater characteristics associated with the leadership role, such as exchange quality, staffing skills, and confrontation skills may help shed additional light on which rater characteristics are likely to explain variance in performance ratings. Such information could be valuable to organizations by indicating manager individual differences related to rating quality and the subsequent identification of high and low potential employees.

Limitations

Despite the contributions of this study, a few limitations warrant discussion. First, because the study's question necessitated a field setting, I was unable to calculate performance true scores to use as a referent when investigating the psychometric properties of ratings. Thus, these results do not allow for a determination of the degree to which transformational bosses' increased elevation and agreement is indicative of inflated ratings or reality (i.e., true score). Nevertheless, an understanding of the trends in performance ratings provide important information about the ways that transformational bosses evaluate follower performance, deliver feedback, and indirectly, cognitively represent aspects of their environment. In addition, the incorporation of other raters of the same target reduces this concern to some degree. That is, although the design did not allow for the computation of true scores, the relationship between elevation and agreement with transformational leadership gives an accurate portrayal of the way

that transformational bosses evaluate their followers relative to other raters of the same person. Moreover, the incorporation of psychometric characteristics served a unique purpose in this study, allowing a direct behavioral outcome of leadership, an area sorely lacking in leadership research. Indeed, previous research has shown that the evaluation of psychometric rating behaviors is meaningful and valuable, even in the absence of a true score comparison (Kane et al., 1995; Tziner et al., 2002).

Next, the rating system utilized in the current study was for developmental purposes and may not generalize to performance appraisals taken for administrative purposes. Alternatively, each boss providing ratings in this study was aware that their ratings would be presented to their followers for feedback purposes, so this issue does not threaten this study's ability to interpret the feedback provided by transformational bosses. In addition, comparisons of administrative and developmental ratings find that although the level of ratings differs between the two purposes, the pattern or rank order of assessees is relatively consistent (Viswesvaran, Ones, & Schmidt, 1996); thus, the correlation results should be relatively consistent for administrative ratings. Still, future research replicating these effects with administrative performance evaluations is needed.

Finally, although the current study provided a first look at the characteristics of transformational bosses' evaluations of their followers, it does not allow for inferences regarding how these bosses administer feedback to their subordinates. It is possible that transformational bosses' ratings are not representative of how they provide feedback to followers. For example, a boss might provide elevated performance ratings, but may temper these positive ratings with more direct, developmental feedback in face-to-face feedback settings. Unfortunately, the data collected in the current study were not sufficient to examine the feedback associated with these

performance ratings. One final limitation is that the rather modest sample size may have precluded the detection of statistically significant effects. For instance, both comparisons between task and interpersonal dimensions approached conventional levels of statistical significance, but were arguably evidence of practically meaningful trends.

Conclusion

Although transformational leadership has been one of the most researched leadership paradigms of the past twenty years, existing research has rarely examined transformational bosses in the context of their day-to-day administrative responsibilities. By examining the rating behaviors of transformational bosses, not only did this study uncover preliminary behavioral evidence for the characteristics of transformational bosses' evaluations, it also facilitated inferences with respect to the cognitions of transformational bosses. Future research investigating the ways that transformational bosses cognitively represent their external environment may be particularly fruitful in understanding the nature of transformational leadership.

Footnotes

¹ Since rating elevation and depression are inverse concepts (i.e., one represents higher than average and the other represents lower than average scores), the remainder of the study will only reference rating elevation in order to avoid redundancy.

² Although ambiguity exists in the appropriate interpretation of LPC scores, high LPC scores are generally viewed as akin to relation oriented leaders and low LPC scores are akin to task oriented leaders.

³ I conducted all analyses with peer and subordinate raters used as a referent and there were no differences across sources in any of the primary relationships under examination in this study.

References

- Antonioni, D., & Park, H. (2001). The relationship between rater affect and three sources of 360-degree feedback ratings. *Journal of Management*, *27*, 479-495.
- Austin, J. T., & Villanova, P. (1992). The criterion problem: 1917-1992. *Journal of Applied Psychology*, *77*, 836-874.
- Bartels, L. K., & Doverspike, D. (1997). Assessing the assessor: The relationship of assessor personality to leniency in assessment center ratings. *Journal of Social Behavior & Personality*, *12*, 179-190.
- Bass, B. M. (1985). Leadership: Good, better, best. *Organizational Dynamics*, *13*, 26-40.
- Bass, B. M., & Avolio, B. (1995). *Multifactor leadership questionnaire (form 5x-short)*. Menlo Park, CA: Mind Garden.
- Bass, B. M., Avolio, B. J., & Goodheim, L. (1987). Biography and the assessment of transformational leadership at the world-class level. *Journal of Management*, *13*, 7-19.
- Bernardin, H. J., Alvares, K. M., & Cranny, C. J. (1976). A recomparison of behavioral expectation scales to summated scales. *Journal of Applied Psychology*, *61*, 564-570.
- Bernardin, H. J., Cooke, D. K., & Villanova, P. (2000). Conscientiousness and agreeableness as predictors of rating leniency. *Journal of Applied Psychology*, *85*, 232-236.
- Bernardin, H. J., & Orban, J. A. (1990). Leniency effect as a function of rating format, purpose for appraisal, and rater individual differences. *Journal of Business and Psychology*, *5*, 197-211.
- Bernardin, H. J., & Walter, C. S. (1977). Effects of rater training and diary-keeping on psychometric error in ratings. *Journal of Applied Psychology*, *62*, 64-69.

- Berson, Y., Shamir, B., Avolio, B. J., & Popper, M. (2001). The relationship between vision strength, leadership style, and context. *Leadership Quarterly*, *12*(1), 53-73.
- Bono, J. E., & Judge, T. A. (2004). Personality and Transformational and Transactional Leadership: A Meta-Analysis. *Journal of Applied Psychology*, *89*, 901-910.
- Borman, W. C. (1977). Consistency of rating accuracy and rating errors in the judgment of human performance. *Organizational Behavior & Human Performance*, *20*, 238-252.
- Borman, W. C. & Brush, D. H. (1993). More progress toward a taxonomy of managerial performance requirements. *Human Performance*, *6*, 1-21.
- Borman, W. C., & Motowidlo, S. J. (1997). Task performance and contextual performance: The meaning for personnel selection research. *Human Performance*, *10*, 99-109.
- Brown, D. J., & Keeping, L. M. (2005). Elaborating the construct of transformational leadership: The role of affect. *Leadership Quarterly*, *16*, 245-272.
- Burns, J. M. (1978). *Leadership*. Oxford, England: Harper & Row.
- Bycio, P., Hackett, R. D., & Allen, J. S. (1995). Further assessments of Bass's (1985) conceptualization of transactional and transformational leadership. *Journal of Applied Psychology*, *80*, 468-478.
- Cardy, R. L., & Dobbins, G. H. (1994). *Performance appraisal: Alternative perspectives*. Cincinnati, OH: South-Western Publishing.
- Cleveland, J. N. & Murphy, K. R. (1992). Analyzing performance appraisal as goal-directed behavior. In G. Gerris and K. Rowland (eds.) *Research in Personnel and Human Resources Management*, (vol. 10, pp. 121-185). Greenwich, CT: JAI Press.
- Conger, J. A., & Kanungo, R. N. (1998). *Charismatic leadership in organizations*. Thousand Oaks, CA, US: Sage Publications, Inc.

- Cooper, W. H. (1981). Ubiquitous halo. *Psychological Bulletin*, 90, 218-244.
- Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. Fort Worth, TX : Harcourt, Brace, Jovanovich.
- Dvir, T., Eden, D., Avolio, B. J., & Shamir, B. (2002). Impact of transformational leadership on follower development and performance: A field experiment. *Academy of Management Journal*, 45, 735-744.
- Edison, J. D., & Adams, H. E. (1992). Depression, self-focus, and social interaction. *Journal of Psychopathology and Behavioral Assessment*, 14, 1-19.
- Feldman, J. M. (1981). Beyond attribution theory: Cognitive processes in performance appraisal. *Journal of Applied Psychology*, 66, 127-148.
- Fiedler, F. E. (1971). Validation and extension of the contingency model of leadership effectiveness: A review of empirical findings. *Psychological Bulletin*, 76, 128-148.
- Fisicaro, S. A. (1988). A reexamination of the relation between halo error and accuracy. *Journal of Applied Psychology*, 73, 239-244.
- Fleishman, E. A. (1957). The Leader Opinion Questionnaire. In R. M. Stogdill & A.E. Coons (Eds.), *Leader behavior: Its description and measurement* (pp.120-133). Columbus, OH: Bureau of Business Research.
- Forret, M. L., & Dougherty, T. W. (2001). Correlates of networking behavior for managerial and professional employees. *Group & Organization Management*, 26, 283.
- Fuller, J. B., Patterson, C. E. P., Hester, K., & Stringer, D. Y. (1996). A quantitative review of research on charismatic leadership. *Psychological Reports*, 78, 271-287.
- Geyer, A. L. J., & Steyrer, J. M. (1998). Transformational leadership and objective performance in banks. *Applied Psychology: An International Review*, 47, 397-420.

- Ghorpade, J., & Chen, M. M. (1995). Creating quality-driven performance appraisal systems. *Academy of Management Executive, 9*, 32-39.
- Glynn, M. A., & Dowd, T. J. (2008). Charisma (un)bound: Emotive leadership in Martha Stewart Living magazine, 1990-2004. *Journal of Applied Behavioral Science, 44*, 71-93.
- Hater, J. J., & Bass, B. M. (1988). Superiors' evaluations and subordinates' perceptions of transformational and transactional leadership. *Journal of Applied Psychology, 73*, 695-702.
- Hedge, J. W., & Kavanagh, M. J. (1988). Improving the accuracy of performance evaluations: Comparison of three methods of performance appraiser training. *Journal of Applied Psychology, 73*, 68-73.
- Hendel, T., Fish, M., & Galon, V. (2005). Leadership style and choice of strategy in conflict management among Israeli nurse managers in general hospitals. *Journal of Nursing Management, 13*, 137-146.
- Hjelle, L. A. (1969). Personality characteristics associated with interpersonal perception accuracy. *Journal of Counseling Psychology, 16*, 579-581.
- Hoffman, B. J., Blair, C. A., Meriac, J. P., & Woehr, D. J. (2007). Expanding the criterion domain? A quantitative review of the OCB literature. *Journal of Applied Psychology, 92*, 555-566.
- Hoffman, B.J., Bynum, B., Piccolo, R., & Williams, A. (2008). Elucidating the transformational effect: Person-organization and person-supervisor fit. Paper to be presented at the 32nd annual meeting of the Academy of Management in Anaheim, CA.
- Hoffman, B.J. & Frost, B.C. (2006). Multiple intelligences of transformational leaders: An empirical examination. *International Journal of Manpower, 27*, 37-51.

- Hoffman, B.J., Lance, C., Bynum, B., & Gentry, B (in press). Rater source effects are alive and well after all. *Personnel Psychology*.
- Hoffman, B. J. & Woehr, D. J. (in press). Disentangling the meaning of multisource feedback: An examination of the nomological network surrounding source and dimension factors. *Personnel Psychology*.
- House, R. J., Spangler, W. D., & Woycke, J. (1991). Personality and charisma in the U.S. presidency: A psychological theory of leader effectiveness. *Administrative Science Quarterly*, 36, 364-396.
- Howell, J. M., & Hall-Merenda, K. E. (1999). The ties that bind: The impact of leader-member exchange, transformational and transactional leadership, and distance on predicting follower performance. *Journal of Applied Psychology*, 84, 680-694.
- Isen, A. M., Shalcker, T. E., Clark, M., & Karp, L. (1978). Affect, accessibility of material in memory, and behavior: A cognitive loop? *Journal of Personality and Social Psychology*, 36, 1-12.
- James, L. R., Demaree, R. G., & Wolf, G. (1984). Estimating within-group interrater reliability with and without response bias. *Journal of Applied Psychology*, 69, 85-98.
- Jöreskog, K., & Sörbom, D. (2004). LISREL 8.70. Chicago: Scientific Software International.
- Judge, T. A., & Piccolo, R. F. (2004). Transformational and Transactional Leadership: A Meta-Analytic Test of Their Relative Validity. *Journal of Applied Psychology*, 89, 755-768.
- Kane, J. S., Bernardin, H. J., Villanova, P., & Peyrefitte, J. (1995). Stability of rater leniency: Three studies. *Academy of Management Journal*, 38, 1036-1051.
- Kingsbury, F. A. (1922). Analyzing ratings and training raters. *Journal of Personnel Research*, 1, 377-383.

- Kirkpatrick, S. A., & Locke, E. A. (1996). Direct and indirect effects of three core charismatic leadership components on performance and attitudes. *Journal of Applied Psychology, 81*, 36-51.
- Kozlowski, S. W., Kirsch, M. P., & Chao, G. T. (1986). Job knowledge, rater familiarity, conceptual similarity and halo error: An exploration. *Journal of Applied Psychology, 71*(1), 45-49.
- Kram, K. E. (1985). *Mentoring at work*. Glenview, IL: Scott, Foresman and Company.
- Lahey, M. A., Downey, R. G., & Saal, F. E. (1983). Intraclass correlations: There's more there than meets the eye. *Psychological Bulletin, 93*, 586-595.
- Lance, C.E., Foster, M.R., Gentry, W.A., Thoresen, J.D. (2004). Assessor cognitive processes in an operational assessment center. *Journal of Applied Psychology, 89*, 22-35.
- Lance, C. E., Hoffman, B. J., Gentry, W.A., & Baranik, L. E. (2008). Rater source factors represent important subcomponents of the criterion construct space, not rater bias. *Human Resource Management Review*.
- Lance, C. E., LaPointe, J. A., & Stewart, A. M. (1994). A test of the context dependency of three causal models of halo rater error. *Journal of Applied Psychology, 79*, 332-340.
- Lindell, M. K., Brandt, C. J., & Whitney, D. J. (1999). A revised index of interrater agreement for multi-item ratings of a single target. *Applied Psychological Measurement, 23*, 127-135.
- Longenecker, C. O., Sims, H.P., & Gioia, D. (1987). Behind the mask: The politics of employee appraisal. *Academy of Management Executive, 1*, 183-191.

- Lord, R. G., Foti, R. J., & DeVader, C. L. (1984). A test of leadership categorization theory: internal structure, information processing, and leadership perspectives. *Organizational Behavior and Human Performance*, 34, 343-378.
- Lucas, R. E., Le, K., & Dyrenforth, P. S. (2008). Explaining the extraversion/positive affect relation: Sociability cannot account for extraverts' greater happiness. *Journal of Personality*, 76, 385-414.
- Meindl, J. R., & Ehrlich, S. B. (1987). The romance of leadership and the evaluation of organizational performance. *Academy of Management Journal*, 30, 91-109.
- Meng, X.-L., Rosenthal, R., & Rubin, D. B. (1992). Comparing correlated correlation coefficients. *Psychological Bulletin*, 111, 172-175.
- Mintzberg, H., & Vecchio, R. P. (2007). The manager's job: Folklore and fact. In *Leadership: Understanding the dynamics of power and influence in organizations (2nd ed.)*. (pp. 33-50). Notre Dame, IN US: University of Notre Dame Press.
- Mount, M. K., Judge, T. A., & Scullen, S. E. (1998) Trait, rater and level effects in 360-degree performance ratings. *Personnel Psychology*, 51, 557-576
- Murphy, K.R. (2008). Explaining the weak relationship between job performance and ratings of job performance. *Industrial and Organizational Psychology, Perspectives on Research and Practice*, 1, 161-166.
- Murphy, K. R., & Anhalt, R. L. (1992). Is halo error a property of the rater, ratees, or the specific behaviors observed? *Journal of Applied Psychology*, 77, 494-500.
- Murphy, K. R. & Cleveland, J. N. (1995). *Understanding Performance Appraisal: Social, Organizational, and Goal-based Perspectives*. Thousand Oaks, CA: Sage Publications.

- Murphy, K. R., Cleveland, J. N., & Skattebo, A. L., & Kinney, T. B. (2004). Raters who pursue different goals give different ratings. *Journal of Applied Psychology, 89*, 158-164.
- Noe, R. A. (1988). An investigation of the determinants of successful assigned mentoring relationships. *Personnel Psychology, 41*, 457-479.
- Piccolo, R. F., & Colquitt, J. A. (2006). Transformational Leadership and Job Behaviors: The Mediating Role of Core Job Characteristics. *Academy of Management Journal, 49*, 327-340.
- Podsakoff, P. M., MacKenzie, S. B., Moorman, R. H., & Fetter, R. (1990). Transformational leader behaviors and their effects on followers' trust in leader, satisfaction, and organizational citizenship behaviors. *Leadership Quarterly, 1*, 107-142.
- Rice, R. W. (1978). Construct validity of the Least Preferred Co-Worker score. *Psychological Bulletin, 85*, 1199-1237.
- Rowold, J. (2008). Effects of transactional and transformational leadership of pastors. *Pastoral Psychology, 56*, 403-411.
- Saal, F. E., Downey, R. G., & Lahey, M. A. (1980). Rating the ratings: Assessing the psychometric quality of rating data. *Psychological Bulletin, 88*, 413-428.
- Saal, F. E., & Landy, F. J. (1977). The mixed standard rating scale: An evaluation. *Organizational Behavior & Human Performance, 18*, 19-35.
- Scullen, S. E., Mount, M. K., & Goff, M. (2000). Understanding the latent structure of job performance ratings. *Journal of Applied Psychology, 85*, 956-970.
- Shamir, B., House, R. J., & Arthur, M. B. (1993). The motivational effects of charismatic leadership: A self-concept based theory. *Organization Science, 4*, 577-594.

- Shore, T. H., Thornton, G. C., & Shore, L. M. (1990). Construct validity of two categories of assessment center dimension ratings. *Personnel Psychology, 43*, 101-116.
- Solomonson, A. L., & Lance, C. E. (1997). Examination of the relationship between true halo and halo error in performance ratings. *Journal of Applied Psychology, 82*, 665-674.
- Tejeda, M. J., Scandura, T. A., & Pillai, R. (2001). The MLQ revisited: Psychometric properties and recommendations. *Leadership Quarterly, 12*, 31-52.
- Thorndike, E. L. (1920). A constant error in psychological ratings. *Journal of Applied Psychology, 4*, 25-29.
- Tziner, A., Murphy, K. R., & Cleveland, J. N. (2001). Relationships between attitudes toward organization and performance appraisal systems and rating behavior. *International Journal of Selection and Assessment, 9*, 226-239.
- Tziner, A., Murphy, K. R., & Cleveland, J. N. (2002). Does conscientiousness moderate the relationship between attitudes and beliefs regarding performance appraisal and rating behavior? *International Journal of Selection and Assessment, 10*, 218-224.
- Tziner, A., Murphy, K. R., & Cleveland, J.N. (2005). Contextual and Rater Factors Affective Rating Behavior. *Group & Organization Management, 30*, 89-98.
- Vandenberg, R. J. and C. E. Lance (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods, 3*, 4-69.
- Van Scotter, J. R., Moustafa, K., Burnett, J. R., & Michael, P. G. (2007). Influence of prior acquaintance with the ratee on rater accuracy and halo. *Journal of Management Development, 26*, 790-803.

- Vingoe, F. J., & Antonoff, S. R. (1968). Personality characteristics of good judges of others. *Journal of Counseling Psychology, 15*, 91-93.
- Viswesvaran, C., Ones, D. S., & Schmidt, F. L. (1996). Comparative analysis of the reliability of job performance ratings. *Journal of Applied Psychology, 81*, 551-574.
- Viswesvaran, C., Schmidt, F. L., & Ones, D. S. (2005). Is there a general factor in ratings of job performance? A meta-analytic framework for disentangling substantive and error influences. *Journal of Applied Psychology, 90*, 108-131.
- Waldman, D. A., Ramirez, G. G., House, R. J., & Puranam, P. (2001). Does leadership matter? CEO leadership attributes and profitability under conditions of perceived environmental uncertainty. *Academy of Management Journal, 44*, 134-143.
- Wells, F. L. (1907). A statistical study of literary merit. (Columbia Univ. Cont. to Phil. & Psych., 16, 3.). *Archives of Psychology, 7*.
- Woehr, D. J., & Lance, C. E. (1991). Paper people versus direct observation: An empirical examination of laboratory methodologies. *Journal of Organizational Behavior, 12*, 387-397.
- Woehr, D. J., Sheehan, M. K., & Bennett, W., Jr. (2005). Assessing measurement equivalence across rating sources: A multitrait-multirater approach. *Journal of Applied Psychology, 90*, 592-600.
- Yukl, G. A. (2006). *Leadership in organizations* (6th ed.). Englewood Cliffs, NJ: Prentice-Hall.