A META-ANALYSIS OF GENDER DIFFERENCES IN LEADERSHIP EFFECTIVENESS

by

Samantha C. Paustian-Underdahl

A dissertation submitted to the faculty of
The University of North Carolina at Charlotte
in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in
Organizational Science

Charlotte

2012

Approved by:

_____
Dr. Lisa Slattery Walker

_____
Dr. David J. Woehr

_____
Dr. Steven G. Rogelberg

_____
Dr. Linda Rhoades Shanock

ABSTRACT

SAMANTHA C. PAUSTIAN-UNDERDAHL. A meta-analysis of gender differences in leadership effectiveness. (Under the direction of DR. LISA SLATTERY WALKER).

Despite years of evidence showing that men are typically considered to be more appropriate and effective in leadership positions than women, a recent debate has emerged over the potential existence of a female leadership advantage in certain contexts (Eagly & Carli, 2003a; Vecchio, 2003; Eagly & Carli, 2003b). This meta-analysis aims to contribute to this debate in the literature by quantitatively summarizing gender differences in leadership effectiveness using 99 independent samples from 95 studies. Results show that when all leadership contexts are considered together there is a non-significant gender difference in leadership effectiveness. Additionally, this study examines the influence of contextual moderators developed from role congruity theory (Eagly & Karau, 2002) and some competing theoretical frameworks. Overall, the findings support the core tenets of role congruity theory—that prejudice against female leaders can vary depending on a variety of features of the leadership context and characteristics of the perceivers of the leader's effectiveness. However, some of the hypotheses proposed by role congruity theory were only partially supported. The results of this meta-analysis point to ways in which the theory can be updated and expanded by taking into account findings supported by other theories presented in the literature on gender and leadership. Finally, this meta-analysis provides important practical implications for reducing the barriers women may face throughout the leadership labyrinth as they pursue the most elite leadership positions.

ACKNOWLEDGEMENTS

My four year adventure here at UNC Charlotte has been more fulfilling than I could have imagined. Faculty, graduate students, family members, and friends contributed to my success while I have been in graduate school, and I would like to take this moment to express my gratitude to these individuals. First off, the faculty members of the Organizational Science program have provided me with an incredible education—giving me the conceptual, research, and writing skills I need to be successful in academia. They have also shown me what a collaborative, kind, organized, and hard-working faculty looks like, which I will pursue in my search for a new academic home.

Lisa Walker has been a strong and supportive advisor as I have developed my research stream. She has provided me with the right balance of guidance and oversight when needed, along with the freedom to pursue independent work. Her sometimes strong opinions and willingness to express them in a warm and open manner are characteristics I strive for as I make the transition from graduate student to professor. In addition, her generosity in sharing resources with me has been incredibly kind and helpful. More importantly, she has demonstrated her faith in my abilities to complete a high quality dissertation in a somewhat abbreviated period of time, and she has given me timely feedback to help me meet critical deadlines. This project would not have been possible without her support.

Steven Rogelberg has provided me with exceptional career advice and guidance throughout my graduate career, which was particularly useful when I decided to pursue academia instead of an applied career. I want to thank him for letting me learn from his ability to teach and inspire learning in others. I hope to be as good of a leader, researcher,

and teacher as him one day. Steven has also provided me with critical feedback on several of my research projects, including this dissertation. He is a tremendous thinker and giver, always devising appropriate tests of hypotheses, presenting scientific findings in a stimulating manner, and knowing how to communicate these skills to students.

Linda Shanock kindly took me under her wing when I felt a bit lost in my first year of graduate school. She was the first professor to guide me through a research project from start to finish, and showed me what it takes to get published. She has always been generous with her time and insights, and I am very thankful for her continued support throughout the years.

In the short time I have known Dave Woehr, he has had a remarkable impact on the present research and a lasting impact on how I think about meta-analyses. His meta-analytic expertise and willingness to provide quick and thorough advice have been extremely helpful throughout the course of this project. I strive to be as kind, thoughtful, and calm with my future students, as he has been with me.

Though not members of my dissertation committee, several other professors have played a major role in the success of my dissertation. Jonathon Halbesleben has served as a tremendous mentor to me. I want to thank him for providing me with amazing research opportunities, having faith that I would finish this dissertation in a timely fashion, letting me borrow his copy of CMA, and giving me an equal voice in our work together. I have learned more about what it means to be a high quality management scholar from him than anyone else. I am also quite thankful to Alice Eagly for being a pioneer in developing a highly impactful stream of research and theory related to women and leadership, and for responding to my questions about this project in a quick and conscientious manner. I

would also like to express my gratitude to Anita Blanchard for teaching me not only about the writing process, but also about life in academia, during her seminar on writing and publishing. She has taught me a lot about managing work and life conflicts, while always being light-hearted and fun both in and outside of the classroom. Finally, I want to thank Loril Gossett for asking the question, 'What drives you?' in her qualitative methods class, and making me realize that I can build an academic career on understanding and resolving inequalities that occur in and outside of the workplace.

This dissertation would also not have been possible without the help of 3 highly skilled and dedicated research assistants. Lara Eller, Jessie Olien, and Christine Curcio were instrumental to the success of this dissertation. Their attentive and capable work enabled me to complete a comprehensive search for relevant primary studies, and to ensure coding accuracy for the primary studies in this meta-analysis. The methodological quality and the timeliness of this project were definitely enhanced because of their efforts.

My fellow graduate students have also made a huge impact on the scholar I am today. Not only have I learned technical skills from many of them, but they have also provided me with a sense of camaraderie and shared ambition that has kept me going. I especially want to thank Kate Frear who has given me daily insights, inspiration, and encouragement throughout the past four years. In our first year here she began using the phrase, "It's almost over", which we have told each other before any major paper, test, or otherwise significant and perhaps scary event. This phrase, along with other examples of her good grace and humor, have greatly contributed to my positive graduate school experience.

Last, but not least, I want to thank my family and friends for their continued support and love. I am grateful to my husband and partner in life Brian Underdahl, who has always encouraged me to follow my dreams, despite his own sacrifices. I am also thankful for his sense of humor and his ability to make me laugh and enjoy life outside of academia. My parents have also been unwavering in their support of my dreams, always encouraging me to do what I love, and to love what I do. Finally, my friends have enabled me to see the light at the end of the tunnel, and to push through to the end. I want to thank them for their continued efforts to be my friend, and for their understanding when I have had to say, 'Sorry, I can't see you….I am working on my dissertation". It's over now, and I will see you soon.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION AND STATEMENT OF PURPOSE

> We still think of a powerful man as a born leader and a powerful woman as an anomaly.
>
> —Margaret Atwood

> A snail could crawl the entire length of the Great Wall of China in 212 years, just slightly longer than the 200 years it will take for women to be equally represented in Parliament.
>
> —2008 UK Sex and Power Report (U.K. Equality and Human Rights Commission, 2008)

> After years of analyzing what makes leaders most effective and figuring out who's got the Right Stuff, management gurus now know how to boost the odds of getting a great executive: Hire a female.
>
> —R. Sharpe, 2000, Business Week

Introduction

While the number of women in the workplace has increased drastically within the past few decades, women remain significantly under-represented at the highest organizational levels (U.S. Bureau of Labor Statistics, 2010). Despite this under-representation, women's educational and work attainments are remarkable: in the United States (U.S.) women earn almost 60% of bachelor's and master's degrees, they make up nearly half of the labor force, and they occupy more than half of all professional and managerial positions (Catalyst, 2012; U.S. Bureau of Labor Statistics, 2011). Women have undeniably been making strides in attaining leadership positions. However, a closer examination of the data reveals the very real gap that still exists in the highest organizational ranks. Women occupy a mere 2.6% of Fortune 500 chief executive officer (CEO) seats and represent only 4% of the CEOs and heads of boards in the European Union (Eagly & Carli, 2007). The numbers are not any better in the political arena.

Currently, women hold only 90 of the 535 seats (16.8%) in the U.S. Congress (Center for American Women and Politics, 2012a), and 19.3% of parliamentary seats globally (Inter-Parliamentary Union, 2011). Why are women so under-represented in the most elite leadership positions? For decades researchers have attempted to answer this question (e.g., Berger, Fisek, Norman, & Zelditch, 1977; Eagly 1987; Eagly & Karau, 2002)—resulting in several potential theories and explanations.

A primary explanation for women's underrepresentation in elite leadership roles has focused on the idea that there are not enough women who enter into occupations that would eventually lead to leadership opportunities—known as the "pipeline problem." This pipeline problem has been attributed to a number of causes, including women's family responsibilities (Greenhaus & Parasuraman, 1999) and the social and dispositional tendencies for women to display fewer of the behaviors and interests that are needed to reach and succeed in leadership positions (e.g., Browne, 1999; Goldberg, 1993). However, given women's increasing educational attainments and their close to equal participation in professional and managerial roles as men, researchers have begun to question the sufficiency of such explanations (Eagly & Karau, 2002).

An alternative explanation for women's underrepresentation in elite leadership positions points to the undervaluation of women's effectiveness as leaders—an explanation supported by several theoretical perspectives including the influence of status based performance expectations (i.e., Berger et al, 1977; Ridgeway 1997; 2001), stereotypes surrounding gender roles and leadership (i.e., Schein, 1973; 1975; Heilman, 2001), and perceptions of role incongruity (Eagly & Karau, 2002). If women are generally seen as less effective leaders than men, then this may help to explain why there

are so few women who make it to the top of organizational hierarchies. Additionally, if measures of leadership effectiveness are used within organizations as part of performance management assessments, gender differences in these assessments can have important implications for employees and leaders.

Gender differences in leadership effectiveness within organizations could influence both pay and promotion opportunities for men and women (Robertson, 1986). Leaders who believe they have received an unfair or biased evaluation of their effectiveness could lose trust in their organization (Schoorman, Mayer, & Davis, 2007) or become less satisfied (e.g. Colquitt, Conlon, Wesson, Porter, & Ng, 2001; see also King, Hebl, George, & Matusik, 2010). Such attitudes have been shown to lead to withdrawal behaviors such as lateness, absenteeism, and turnover (Harrison, Newman, & Roth, 2006) which could contribute to women's underrepresentation in these roles. Understanding how measures of different effectiveness criteria and contextual variables such as industry, level of leadership, rater identity, and other aspects of the leadership setting influence the magnitude of any gender differences in the measurement of leadership effectiveness is also important to inform human resource practices and policies. If for example, there is a large gender difference in leadership effectiveness at a particular level of leadership or within a certain industry, it would be helpful for human resource professionals to be made aware of this in order to create better training, development, and performance assessment programs to improve the effectiveness of male and female leaders.

In addition to the organizational and practical implications discussed above, by taking a closer look at the relationship between gender and leadership effectiveness, the usefulness of theories that sometimes propose competing hypotheses regarding gender

and leadership can be examined. An examination of the literature shows that there are discrepant theoretical perspectives and empirical evidence regarding when and why there may or may not be a gender advantage in leadership effectiveness, pointing to the potential influence of contextual moderators (i.e., Berger, et al., 1977; Schein, 1973; Foschi, 2000; Kanter, 1977a). Thus, research is needed to clarify the relationship between gender and leadership effectiveness in order to help resolve discrepancies in the literature, and to examine contributions of different theoretical perspectives.

Debate in the Literature

In the past few decades popular press publications have paid significant attention to women leaders and their possible 'gender advantage' in leadership effectiveness (e.g., Helgesen, 1990; Rosener, 1995). Articles in newspapers and business magazines have exposed a new cultural trend in the United States for women to be considered particularly effective leaders in modern organizations. For example, Business Week announced that women have the ''Right Stuff'' (Sharpe, 2000), and Fast Company agreed that "The future of business depends on women" (Hefferman, 2002, p. 9). In a subsequent publication, Business Week included a cover story on the ''New Gender Gap,'' proposing that ''Men could become losers in a global economy that values mental power over might'' (Conlin, 2003, p. 78). Despite these articles potentially helping to reduce some of the negative stereotypes and status beliefs regarding women and leadership, their conclusive tone regarding a female advantage brought about a debate amongst academic leadership and gender researchers (see Vecchio, 2002; Eagly & Carli, 2003a; Vecchio, 2003; Eagly & Carli, 2003b; Yukl, 2002).

A Male Advantage

The traditional argument regarding gender differences in leadership outcomes has taken a male advantage perspective—proposing that men are seen as better suited for and thus more effective in positions of leadership than women. According to status characteristics theory (Berger, et al., 1977), men in general have higher status in society than women do. This higher status position means that they are typically seen as more suited for and more competent in powerful, leadership positions than women. Expectation states theory explains that beliefs regarding status characteristics create an implicit sense of the kinds of people (e.g. men) that are likely to be and should be competent as leaders (Berger et al., 1998; Ridgeway & Berger, 1986). When individuals assess a situation, relying on this implicit sense of expected competence based on status, they may refuse to accept or follow people from lower status groups (e.g. women) who attempt to lead others. Indeed, research has shown that women's efforts to assert power can result in resistance and dislike, which reduces their ability to be effective leaders (Heilman, Block, & Martell, 1995; Ridgeway, 2001).

This research is consistent with role congruity theory (RCT) which proposes that, in general, prejudice toward female leaders follows from the incongruity that many people perceive between the characteristics of women and the requirements of leader roles (Eagly and Karau, 2002). People tend to have dissimilar beliefs about the characteristics of leaders and women, and similar beliefs about the characteristics of leaders and men. According to the theory, by simply occupying leadership positions, women will likely encounter disapproval and negativity due to perceived gender role violation (Eagly & Karau, 2002).

Another theoretical framework which also provides support for a male advantage in leadership effectiveness is based on Schein's research related to gender and managerial stereotypes (1973; 1975). Schein's work shows that there is a persistent stereotype that associates management with being male. Research on the think manager-think male phenomenon shows that the attributes ascribed to managers are more closely related to attributes ascribed to men than women (e.g., Schein, 1973; 1975; Brenner, Tomkiewitz, & Schein, 1989; Heilman, Block, & Martell, 1995; Powell & Butterfield, 1979; Powell, Butterfield, & Parent, 2002; Willemsen, 2002). Such stereotypes can lead to more favorable evaluations of the leadership behaviors of men than women (Eagly & Karau, 2002).

Several studies have been conducted which provide support for these theories. A meta-analysis of experimental research included primary studies that presented written descriptions of managerial behavior to participants. These descriptions differed only in the sex of the leader. The authors found that there is an overall tendency for participants to evaluate female leaders less favorably than male leaders (Eagly, Makhijani, & Klonsky, 1992). Another study conducted in an organizational setting shows the same effect. Blanton (1996) examined principals of public middle schools and found that male principals were seen as significantly more effective than female principals. In a different study examining U.S. state legislators, followers were significantly more satisfied with male than female legislators, and males were seen as more effective overall than female legislators (Tucker, McCarthy, & Jones, 1999). Despite theoretical and empirical support for a male advantage in leadership effectiveness, other theoretical frameworks provide support for a female leadership advantage in which women may be more likely to lead in

ways that are seen as particularly effective in modern organizations (Foschi, 2000; Feldman, 1981; Weber & Cracker, 1983; Eagly & Carli 2003a & 2003b).

A Female Advantage

Eagly and Carli (2003a; 2003b) have proposed that women may be more likely than men to adopt transformational leadership behaviors, which are related to overall leadership effectiveness, especially in modern organizations (Judge & Piccolo, 2004). Over the course of the past several decades, there has been a shift in the ways organizations operate. Many modern organizations function globally, with diverse customers, shareholders, and employees, and more complex missions and contexts than in the past. Definitions of effective managerial practices have evolved in response to features of modern organizational environments (e.g., Avolio, 1999; Lipman-Blumen, 2000). Based on these shifts in organizational contexts, the effectiveness of transactional, command-and-control styles of leadership may be waning (e.g., Gergen, 2005; Lipman-Blumen, 2000; McCauley, 2004). There seems to be an increased demand for transformational managers and leaders who can inspire and engage their followers to work effectively in a complex environment.

Empirical research supports this view. A recent meta-analysis examined 87 studies testing the relationships between transformational, transactional, and laissez-faire leadership, with the outcome variable of leadership effectiveness. The authors found that transformational leadership has the strongest relationship with leadership effectiveness (Judge & Piccolo, 2004). There is also evidence that women may be more likely than men to engage in such effective leader behaviors. Eagly and colleagues (2003) found meta-analytic support that women are likely to behave in a more transformational style

than male leaders. They argue this is due to shared expectations that apply to persons who are members of a particular social category (Biddle, 1979; Sarbin & Allen, 1968). Gender role expectations create differential interpretations of male and female leadership behaviors, meaning that men and women could enact the same behavior, but it may be perceived differently by followers because of gender lenses which color interpretations of everyday behavior. The "feminine advantage" perspective of leadership (Yukl, 2002), contends that women are more skilled at inclusiveness, interpersonal relations, democratic decision making, and the nurturing of followers; and as a consequence, women should be more effective leaders.

Lauterbach and Weiner (1996) found support for a female advantage in leadership in their study of Fortune 100 leaders in middle management positions. Women managers were seen as more likely than males to act out of the organization's best interest, consider others' viewpoints, involve others in planning, and focus on both task and interpersonal aspects of their leadership. They were also seen as significantly more successful as leaders than the male leaders in the sample. Douglas and Ammeter (2004) found similar effects. They examined perceptions of leader performance for male and female unit supervisors of a large public school administration. Female supervisors were seen as having better interpersonal and networking skills, as well as being more effective leaders than men.

Additionally, research on double standards of competence (Foschi, 2000), and the boomerang effect (Feldman, 1981; Weber & Cracker, 1983), suggests that in certain situations, female leaders may receive more favorable evaluations of their leadership abilities than male leaders. These perspectives suggest that if women can make it to

leadership positions in highly masculine fields, or to the highest organizational levels, they may be seen as particularly effective due to perceptions of their ability to overcome significant barriers along the way (Crocker & Major, 1989; Rosette & Tost, 2010) and due to the reversing of stereotypes that reflect women as less effective leaders than men (Feldman, 1981; Weber & Cracker, 1983).

A recent study tested these ideas by presenting participants with job descriptions of either a middle or upper level leadership position, and then having the participants review and evaluate a performance assessment of someone currently holding that position (Rosette & Tost, 2010). The authors found that when a woman is employed in an upper leadership role, her successful performance conveys information to participants that increases their evaluation of the woman leader. Their findings suggest that these overvaluations occurred because women were perceived to face higher standards than men and were expected to engage in highly effective leadership behaviors. Feldman (1981) and Weber and Crocker (1983) have described a similar phenomenon; when observed behavior contrasts sharply with a stereotypically derived expectation, the incongruence between the observed behavior and the expected behavior can cause a boomerang effect. Consistent with this argument, several studies have found support for the overvaluation of women when success on a male sex-typed task is both undeniable and believed to be due to the woman's ability (Heilman, Martell, & Simon, 1988; Kryger & Shikiar, 1978).

No Gender Advantage

A third perspective concludes that there are likely to be no gender differences in leadership effectiveness assessments based on the structural influence of the leadership

environment in restricting any potential differences that may be brought about by leader gender (e.g. Kanter, 1977a). According to this argument, leadership positions demand particular behaviors and attitudes of leaders, essentially confining men and women in the same role to behave in the same ways (e.g., Kanter, 1977a; Nieva & Gutek, 1981; van Engen, van der Leeden, & Willemsen, 2001). These demands on leaders' behavior come in part from expectations organizational members form of how leaders should and should not behave (Phillips & Lord, 1982).

Kanter (1977a) conducted a study examining the behavior of men and women employees in a large corporation. Based on the data she gathered regarding the hiring of leaders, she observed, "the women who could get in were the ones who could demonstrate the ability to think like a man" (Kanter, 1977a, p. 25). Thus, people who are selected into leadership roles are likely to be similar already. Once they become members of the organization, the expectations associated with the leader role should create further standardization of leader behaviors (Eagly, 2007), which may lead to similar ratings of leader effectiveness for men and women. Bolman and Deal (1991) found support for this argument in a study examining the effectiveness of higher education department chairs and college presidents. They found non-significant differences in effectiveness between male and female leaders in their sample. Additionally, Vilkinas (2000) compared effectiveness assessments of male and female middle managers in Australian organizations, and found no differences between the groups.

Contextual Moderators of Gender Differences

Despite the theoretical perspectives presented above each providing support for a male advantage, female advantage, or a lack of gender advantage in leadership

effectiveness, respectively, they also provide some insights into how the effectiveness of leader behaviors may depend on contextual variables, such as the nature of the task and the characteristics of the followers (see Vecchio, 2002; Eagly & Carli, 2003a; Vecchio, 2003; Eagly & Carli, 2003b; Chemers, 1997). Status characteristics theory and expectations states theory each provide support for a general male advantage in performance assessments. Yet, a distinctive tenet of expectation states theory is that the biasing effect of gender status beliefs on performance expectations is not consistent across different situations. The biasing effect of gender status beliefs can vary depending on how salient and task relevant gender is in the situation compared to other more specific status characteristics (i.e., leader experience, success, behaviors, etc.) (Wagner & Berger, 1997; Ridgeway, 1997). Specifically, expectation states theory argues that gender status beliefs become salient enough to measurably affect evaluations of effectiveness when gender either distinguishes between the actors in a situation (Cota & Dion, 1986) or is linked by cultural beliefs to the task or goal being completed.

Similarly, RCT proposes unique contextual variables as moderators of the strength with which stereotypes regarding gender roles and leader roles, and their congruity play a role in effectiveness assessments (Eagly & Karau, 2002). In general, the theory argues that prejudice towards female leaders results from perceptions that the characteristics needed to be successful in leadership positions are incongruent with the characteristics associated with the female gender role. However, the theory also proposes that prejudice against female leaders can vary depending on a variety of features of the leadership context and characteristics of the perceiver(s) of the leader's effectiveness (Eagly & Karau, 2002). Eagly and Karau (2002) suggest that there may be a gender

advantage in leadership effectiveness favoring either men or women, depending on the extent to which the leadership context is congruent with male and female gender roles. Similarly, the extent to which the think manager-think male phenomenon holds depends on the specific time period in which leadership effectiveness is being evaluated. Research shows that stereotypes associating men with leadership seem to be waning over time (Powell, Butterfield, & Parent, 2002). Thus, the time period in which leadership effectiveness is being assessed may influence whether and the extent to which any gender differences may exist.

In addition, the perspective presented by Eagly and Carli (2003a; 2003b), proposing that women may be more likely to lead in a transformational manner that is seen as particularly effective in modern organizations, also supports the existence of contextual moderators. As Vecchio (2003) discusses briefly in his rebuttal of Eagly and Carli (2003b) and as supported in Carson's recent dissertation (2011), transformational leadership may only be seen as a particularly effective leadership style in certain leadership contexts in which relations-oriented, developmental leadership is needed by followers. Carson (2011) argues that within traditional, task-oriented, and hierarchically structured organizations, "it may not be as necessary for leaders to exhibit the behaviors associated with transformational leadership in order to be perceived as effective by their direct reports. Instead transactional leadership behaviors such as clearly outlining job expectations, rewarding/incentivizing desired behaviors, and enforcing rules in an effort to avoid mistakes (Avolio, Bass, & Jung, 1999; Yukl, 2006) may be perceived as more effective" (pg. 60).

Thus, it seems that in some settings (i.e., collaborative, flatter organizations), transformational leadership behaviors may be more important for leadership effectiveness. However, in other settings (i.e., independent, task-focused, hierarchical organizations) a transactional or a hands-off leadership style may be more relevant to leaders' success. Thus, while women may be more likely than men to adopt transformational leadership behaviors, those behaviors may only be seen as particularly effective in certain leadership contexts. Additionally, gender stereotypes play a very important role in how leader behaviors are perceived in the workplace. Even when male and female leaders behave similarly, they are not necessarily evaluated similarly (Jago & Vroom, 1982; Nieva & Gutek, 1982; Powell, 1988). Thus, even if women adopt transactional behaviors needed in leadership contexts requiring more of a task-focused kind of leadership, they may be seen as less effective due to the incongruity between these behaviors and the female gender role.

Research has shown that both women and men are evaluated less favorably when they behave in ways that are incongruent with their gender roles than when they behave in a gender-congruent manner (Haccoun, Haccoun, & Sallay, 1978; Petty & Lee, 1975; Schein, 1973; Watson, 1988; Wiley & Eskilson, 1982). Eagly and others have proposed that women may overcome perceptions of their incongruity with leadership roles by behaving in a transformational manner (Eagly, 2007; Eagly & Carli, 2003a). This transformational style of leadership may be considered by raters to be appropriate and acceptable for women leaders because of its balance between the communion-related qualities (i.e., individualized consideration, inspirational motivation) associated with women, and the agentic qualities (i.e., idealized influence, intellectual stimulation)

associated with leaders. Thus, when leaders are in settings that require transformational leadership behaviors (and they adopt such behaviors), women may be seen as more effective than men, due to the congruity between their behaviors and their gender role. However, when leaders are in settings that require transactional or laissez-faire leader behaviors (and they adopt such behaviors), men may be considered to be more effective than women due to the congruity between these behaviors and their gender role.

Additionally, the double standards of competence perspective (Foschi, 2000), as well as the boomerang effect (Feldman, 1981; Weber & Cracker, 1983), also suggest that there may be contextual moderators to gender differences in leadership effectiveness. These theoretical paradigms suggest that only in certain situations—in leadership contexts deemed to be particularly difficult for women—female leaders may receive more favorable evaluations of their leadership abilities than male leaders. In other contexts, these biases that help women are likely to disappear, in favor of the more dominant biases that tend to favor male leaders.

Summary

Taken all together, these perspectives and discrepant findings regarding whether or not there is a gender advantage in leadership effectiveness point to the potential influence of moderators in affecting the direction and strength of effects. This idea is supported by RCT which provides what is likely to be the most comprehensive theory of prejudice against female leaders that has been developed to date. As mentioned above, RCT proposes that prejudice against female leaders can vary depending on a variety of features of the leadership context as well as characteristics of the evaluators of the leader (Eagly & Karau, 2002). Yet, there are other theoretical frameworks proposed in the

literature surrounding gender and leadership which provide hypotheses that compete with some of those presented by RCT (the boomerang effect: Feldman, 1981; Weber & Cracker, 1983; double standards of competence: Foschi, 2000; self categorization theory: Turner, 1985; Turner, Hogg, Oakes, Reicher, & Wetherall, 1987; social identity theory: Tajfel & Turner, 1979; Turner, 1985; expectation states theory: Ridgeway, 2001; research on individuating information: Reskin, 2000).

Thus, a meta-analysis (i.e., quantitative summary) of the literature examining gender and leadership effectiveness is needed to: 1) summarize the overall relationship between gender and leadership effectiveness, and 2) clarify the effects of contextual moderators on gender differences in leadership effectiveness. The ability of meta-analytic methods to detect and estimate moderator effects enables them to provide boundary conditions for theories and relationships within a field of research (Cortina, 2003). As Aguinis and colleagues (2011) argue in their recent review of meta-analytic choices and judgment calls, "meta-analyses (i.e., quantitative literature reviews) have revolutionized the field of management. Conclusions based on meta-analytic findings set the standard for what is considered state of the science, for what we know and do not know, and for which theory is considered valid and which one is not (Schmidt, 1992, 1996)" (p. 2). The current meta-analysis will contribute to the debate over gender advantages in leadership, as well as discrepancies between hypotheses proposed by RCT, and other theoretical perspectives. Further, this research will provide practical implications for understanding women's underrepresentation in elite leadership positions.

Purpose of the Present Study

In order to clarify the relationship between gender and leadership effectiveness—and to examine potential moderators of this relationship—a quantitative synthesis of laboratory and organizational studies which have examined gender differences in leadership effectiveness across a variety of time periods and leadership contexts is needed. A similar meta-analysis was published by Eagly, Karau, and Makhijani in 1995, which integrated the results of 76 effect sizes from studies conducted in the U.S. up until 1989. This meta-analysis found that male leaders were seen as more effective than female leaders in roles that were male dominated or masculine in other ways (i.e., numerically male-dominated roles; military roles). Additionally, female leaders were seen as more effective than male leaders in less male-dominated or less masculine roles (i.e., educational and social service organizations). Since this meta-analysis was published, there have been additional theoretical perspectives proposed in the literature as well as more rigorous meta-analytical methods developed, which can be used to complement and compete with the findings and theories presented by Eagly and colleagues (1995). These theoretical and methodological factors provide many reasons to conduct a new meta-analysis of research on gender and leadership effectiveness.

First, the 1995 meta-analysis included studies completed up until 1989; approximately 22 years of new data and theories have accumulated since that time. The current meta-analysis updates and expands upon the 1995 meta-analysis by examining competing hypotheses and research questions originating from primary studies conducted up until 2011. Second, cultural shifts have occurred over time, making it important to assess how the passage of time may have affected gender differences in leadership effectiveness. Third, there has not been a comprehensive examination of gender

differences in leadership effectiveness across different countries. The current meta-analysis will include data from primary studies conducted in multiple countries. Fourth, statistical methods of meta-analysis have advanced. At the time of the 1995 meta-analysis, only fixed-effects models were available.

Fixed effects models have since been criticized and random-effects models have been developed and shown to be better suited for meta-analyses of different populations (Hedges & Vevea, 1998; Lipsey & Wilson, 2001). The current meta-analysis utilizes random-effects methodology, as well as the Hunter-Schmidt approach to meta-analysis, which allows for corrections of attenuation due to unreliability of the effectiveness measure (Hunter & Schmidt, 2004). Thus, the current study updates and expands upon the Eagly et al., (1995) meta-analysis to clarify how the relationship between gender and leadership effectiveness may vary depending on the influence of a variety of contextual moderators proposed by RCT and competing theories, including the passage of time, the nature of the effectiveness criterion, the leadership industry, the hierarchical level of the leader position, the sex of the raters, national differences, and the percentage of male raters. Practically relevant moderators including the type of effectiveness criteria, rater type, and publication type are also explored.

In the next section, I review the conceptualization of leadership effectiveness used in this study and present theoretical support for hypotheses and research questions developed from RCT and other theoretical perspectives surrounding gender and leadership. I review in greater depth previous research related to gender and leadership, including descriptions of each construct I will include in the meta-analysis. In the methods section, I describe the methodology used to test the proposed hypotheses and

research questions, including the study design, measures, and analytical techniques. In the final sections of this dissertation, I provide the results of the study and conclude with a discussion of the study findings, implications, limitations, and directions for future research.

LITERATURE REVIEW AND STATEMENT OF HYPOTHESES

Defining Leadership Effectiveness

Before presenting hypotheses and research questions developed from RCT, and some competing theories, I first discuss my conceptualization of the primary dependent variable in this meta-analysis—leadership effectiveness. The matter of how leaders' effectiveness should be conceptualized and measured has been addressed repeatedly by organizational scientists (e.g., Bass, 1990; Hunt, 1991; Tsui, 1984; Yukl & Van Fleet, 1992; Avolio, Sosik, Jung, & Berson, 2003; Yukl, 2006). From the 1950's to the 1980's, a common way to describe leadership effectiveness was through two-factor models of leadership behavior. Examples include task-oriented versus relations-oriented leadership, initiating structure and consideration, autocratic versus participative leadership, transformational versus transactional leadership, and charismatic versus non-charismatic leadership (Yukl, 1999; Avolio et al., 2003). More recently, however, scholars have been moving away from describing effectiveness using two-factor models of leadership behavior. Instead, effective leadership has been described as a leader's capacity to, "facilitate the ability of a group or organization to meet its goals and maintain itself over time" (Eagly, et al., 1995, p. 128; Hunt, 1991; Mumford, 1986). Therefore, effectiveness is currently regarded as an outcome of leaders' behavior rather than a particular type of behavior (e.g., Hunt, 1991; Mumford, 1986; Hogan, Curphy, & Hogan, 1994; Yukl, 2006).

Leadership effectiveness has also been defined as, "judgments about a leader's impact on an organization's bottom line (i.e., the profitability of a business unit, the quality of services rendered, market share gained, or the win— loss record of a team)"

(Hogan et al., 1994, p. 4). The leadership effectiveness domain is a fairly large and complex construct which has been measured in many ways, including through the use of objective and subjective indicators. In a recent review of the leadership literature, DeRue, Nahrgang, Wellman, and Humphrey (2011) proposed that leadership effectiveness criteria can be conceptualized along three dimensions: (a) content, (b) level of analysis, and (c) target of evaluation. Examples of the first dimension—content—include objective measures such as unit or organizational task performance, and subjective content such as affective and relational criteria (e.g., satisfaction with the leader). An additional kind of effectiveness content may consist of overall judgments of effectiveness that encompass both objective and subjective elements. Levels of analysis define whether the effectiveness measure consists of an individual-level indicator, or a dyadic, unit, or organizational type of indicator. Finally, the third dimension—target of evaluation—refers to exactly whose performance is being assessed. The target could be the leader him or herself (i.e., satisfaction with the leader or leader performance) or the target could be a more distal source of effectiveness for the leader (i.e., group or subordinate performance) (DeRue et al., 2011).

Yukl (2006, p. 11) argues that, "it is usually best to include a variety of criteria in research on leadership effectiveness". Thus, based on the dimensions developed by DeRue and colleagues (2011) as well as the coding scheme used in a previous meta-analysis of gender differences in leadership effectiveness (i.e., Eagly, et al., 1995), I incorporate a variety of leadership effectiveness criteria in this meta-analysis. The following measures of leadership effectiveness are included in the current study: (a) measures of leaders' effectiveness, performance, or leadership ability that were based on

self-ratings or ratings by supervisors, peers, subordinates, or judges; (b) ratings (by self or others) of satisfaction with leaders or satisfaction with leaders' performance; (c) coding or counting of effective leadership behaviors by trained judges or observers; and (d) measures of organizational productivity or group performance. I will code these measures for 1) different content, 2) different targets, and 3) different sources.

Magnitude of Gender Differences

The theoretical perspectives discussed in the introduction indicate that there are likely to be contextual moderators which will lead to a male advantage in some situations, and a female advantage in others, cancelling each other out when the effects are combined and summarized. Further, in leadership contexts in which gender is not salient to employees and managers, or to the tasks at hand, there may not be any gender differences in measures of leadership effectiveness (Ridgeway, 2001). Thus, when the results of several studies from a variety of settings and time periods are quantitatively summarized, I hypothesize:

Hypothesis 1: Overall, gender differences in leadership effectiveness will be small or non-existent ($d < .2$).

Hypotheses and Research Questions Developed from Role Congruity Theory

RCT provides a comprehensive and valuable theory of prejudice against female leaders, which proposes that prejudice against female leaders can vary depending on a variety of features of the leadership context as well as characteristics of the evaluators of the leader (Eagly & Karau, 2002). The theory is developed in part from social role theory, which argues that individuals develop descriptive and prescriptive gender role expectations of others' behavior based on the social roles filled by men and women

(Eagly, 1987). Men have historically filled higher status, breadwinner roles which tend to require agentic characteristics like being assertive, aggressive, and independent, while women have traditionally filled lower status, caregiving roles which require communal characteristics of being other-oriented and sympathetic. Based on this sex-based division of social roles, women are typically described and expected to be more communal than men, while men are believed and expected to be more agentic than women. RCT builds and expands upon social role theory by considering the congruity between gender roles and leadership roles, as well as by proposing moderators that influence congruity perceptions and subsequent prejudicial behaviors (Eagly & Karau, 2002).

The theory argues that variations in the perceived lack of fit of women in leadership positions, either due to the extent stereotypes are applied to an individual woman or due to the extent a given job is male gender-typed, should correspondingly influence the degree and frequency of gender bias that results. Several contextual moderators of gender differences in leadership effectiveness are presented by Eagly and Karau (2002) in their review of RCT. Yet, the vast body of literature examining gender and leadership provides a number of hypotheses that compete with some of RCT's hypotheses regarding the nature with which moderators play a role in gender differences of leadership effectiveness. In the following sections, I describe in more detail hypotheses derived from RCT (as presented in Eagly and Karau, 2002) and some competing hypotheses proposed by other theoretical paradigms. Table 1 summarizes these hypotheses.

TABLE 1: Role Congruity Theory and Some Competing Hypotheses

Moderator	RCT Hypotheses	Competing Hypotheses
Time of study	Time of study will moderate gender differences in leadership effectiveness such that there will be greater gender differences (favoring men) seen amongst older studies, and smaller gender differences seen amongst newer studies (Spence & Buckner, 2000 as cited in Eagly & Karau, 2002).	None
Industry	As the masculinity of the leadership industry increases, the incongruity between the female gender role and the leadership role increases, leading to reduced perceptions of women's' effectiveness (Eagly & Karau, 2002).	The boomerang effect proposes that the extent to which observed behavior contrasts sharply with a stereotypically derived expectation, the incongruence between the observed behavior and the expected behavior can cause a boomerang effect (Feldman, 1981; Weber & Cracker, 1983). As the masculinity of the industry increases, women's success in obtaining these roles will create a boomerang effect in which their perceived effectiveness is exaggerated.

TABLE 1 (Continued)

Moderator	RCT Hypotheses	Competing Hypotheses
Leadership level	Eagly and Karau (2002) draw from literature on abilities associated with leadership levels (e.g. Pavett & Lau, 1983; Martell et al, 1998) and gender differences in leadership behaviors (Eagly, et al., 2003), proposing that at the lowest leadership levels, more task-oriented behaviors are needed to be successful, and since men are more likely to engage in such behaviors, they are likely to be seen as more effective. At middle leadership levels, more relational and transformational leadership behaviors are needed, and women are more likely than men to engage in such behaviors, thus women may be seen as more effective in middle management than men. At the highest levels, agentic behaviors are needed for success, and thus men will be seen as more effective in these upper positions.	According to the double standards of competence model, women may be seen as more effective than men at the highest leadership levels due to the idea that they have overcome significant barriers to make it to the top, and thus must be particularly effective as leaders (Foschi, 2000).
Sex of raters	Men have a more masculine construal of leadership than women do. This perception will lead to male raters being more inclined to favor male leaders in rating their effectiveness. Women have a more androgynous view of leadership, and will rate men and women leaders similarly (Schein, 2001 as cited in Eagly & Karau, 2002).	According to self categorization theory (Turner, 1985; Turner, et al., 1987) and social identity theory (Tajfel & Turner, 1979), raters will see same-gender leaders as members of their in-group, whereas opposite-gender leaders will be seen as out-group members and will be subject to stereotyping and discrimination in favor of the in-group. Male leaders will be rated as more effective by male raters. Female leaders will be rated as more effective by female raters.
Nationality	There are likely to be national cultural variations (i.e., gender egalitarianism) of the congruity between gender roles and leader roles (Eagly & Karau, 2002). Yet, research shows that organizational cultural variables may play a greater role than national culture on employee perceptions (Pothukuchi et al., 2002). Thus, it is unclear how national culture may affect gender differences in leadership effectiveness.	

TABLE 1 (Continued)

Moderator	RCT Hypotheses	Competing Hypotheses
Setting (lab or organizational study)	As cognitive resources become limited, raters are more likely to rely on stereotypes in making judgments of leadership effectiveness. Organizational settings are often busy and chaotic places, in which evaluators of leader effectiveness are likely to have greater cognitive load than evaluators in a laboratory setting. Thus, gender differences in leadership effectiveness will be greater (favoring men) in organizational settings than in laboratory settings (e.g. Macrae, Hewstone, & Griffiths, 1993 as cited in Eagly & Karau, 2002).	Due to the effects of individuating information and knowledge of leaders' specific status characteristics available to evaluators in organizational settings, gender differences in leadership effectiveness (favoring men) should be smaller in organizational settings than in laboratory settings (Reskin, 2000; see also Eagly et al., 1992, 1995).
Percent of male raters	As the percentage of male raters increases, women leader's female-stereotypical qualities become more salient, and they will be seen as less effective. Fewer male raters will lead to differences in effectiveness favoring female leaders (Kanter, 1977 as cited in Eagly & Karau, 2002).	According to expectation states theory, as the percentage of male raters reaches either low or high extremes, men will be seen as more effective due to the salience of gender to the context (Ridgeway, 2001). There will be a curvilinear relationship such that as the percentage of male raters is closer to 50%, gender will be less salient, and gender differences in perceived effectiveness will be small, while at the extremes, men will be seen as more effective.

Time of Study as a Moderator

RCT proposes that it is important to consider how time may moderate gender differences in leadership effectiveness (Eagly & Karau, 2002). Over the past several decades, more and more women have become active participants in the world of work generally, and in leadership positions specifically. Women's participation in the U.S. labor force has increased from 33% in 1950 to 61% in 2010 (U.S. Department of Labor, 2010). Women now constitute 25% of chief executive officers when all organizations in the U.S. are considered and more than half of all professional and managerial positions (U.S. Department of Labor, 2010; Catalyst, 2012). Women have also become more active in political leadership positions, making up 16.8% of the U.S. Congress (vs. 2% in 1950; Center for Women and Politics, 2012a) and 12% of governor seats (vs. 0% in 1950; Center for American Women and Politics, 2012b). This same pattern is being observed in countries other than the U.S. as well (see European Commission, 2010; Inter-Parliamentary Union, 2010).

This increase in female participation in leadership roles may be associated with a weakening of the perceived incongruity between women and leadership. Powell and colleagues (2002) propose that stereotypes may change over time in the presence of disconfirming information. According to the bookkeeping model of stereotype change, stereotypes are open to revisions, and may change gradually if there is a steady stream of disconfirming information (Rothbart, 1981). Thus, as more women have entered into and succeeded in leadership positions, it is likely that people's stereotypes which associate leadership with masculinity have been dissolving slowly over time.

Koenig, Eagly, Mitchel, and Ristikari (2011) conducted a meta-analysis examining the extent to which stereotypes of leadership are culturally masculine. They examined the influence of study publication date on this association, and found that leadership was seen as more gender-neutral over time. Additionally, when examining shifts in the think manager–think male paradigm (Schein, 1973), Koenig et al. (2011) found that the change over time took the form of increasing similarity between leaders and women without change in the similarity between leaders and men. It seems that people's perceptions of leadership are shifting to become more associated with feminine characteristics than in the past. The authors' conclusion was, "that leadership now, more than in the past, appears to incorporate more feminine relational qualities, such as sensitivity, warmth, and understanding, thus adding them to the masculine dominance and strength qualities traditionally associated with leadership." (Koenig et al., 2011, p. 634). Given the decrease over time in the perceived incongruity between women and leadership, RCT proposes that assessments of men's and women's leadership effectiveness will become more similar, or favor women over time.

> Hypothesis 2: Time of study will moderate gender differences in leadership effectiveness such that there will be greater gender differences (favoring men) seen amongst older studies, and smaller gender differences, or differences that favor women seen amongst newer studies.

Industry as a Moderator

RCT highlights the importance of the fit between gender roles and the requirements of leader roles, and proposes that the relative success of male and female leaders should depend on the particular demands of these roles (Eagly & Karau, 2002).

According to the theory, industries that are highly male dominated or culturally masculine in their demands present particular challenges to women because of their incompatibility with people's expectations about women. This incompatibility not only restricts women's access to such industries but also can compromise perceptions of their effectiveness. When leader roles are particularly masculine, people may suspect that women are not qualified for them, and they may resist women's authority (Eagly & Karau, 2002; Heilman, 2001). Although leadership positions are traditionally considered to be masculine, and are often dominated by males, they can vary widely in these aspects. Some industries are less masculine and in recent years have been occupied by more women than men (e.g., social service organizations, educational fields; U.S. GAO, 2010). The extent to which industries are defined in a masculine manner is one moderator proposed by RCT.

Many studies have been carried out with a focus on the congruity between the type of job role or task and the gender of the person who performs it. Davison and Burke (2000) conducted a meta-analysis on primary studies that examined the influence of role congruity on gender biased selection decisions in simulated settings. Their findings were in line with the RCT of prejudice toward female leaders, such that men and women were evaluated less favorably when they were selected to perform tasks that were incongruent with their gender. Additionally, a meta-analysis integrated the results of 76 effect sizes assessing how well male and female leaders performed as leaders (Eagly, et al., 1995). This meta-analysis found that male leaders were seen as more effective than female leaders in roles that were male dominated or masculine in other ways (i.e., numerically male-dominated industries; military roles; Eagly et al., 1995). Additionally, female

leaders were seen as more effective than male leaders in less male-dominated or less masculine industries (i.e., educational, governmental, and social service organizations).

> Based on RCT, I could hypothesize that in industries that are male-dominated, and culturally masculine, women will be considered to be less effective leaders than men, and in organizations that are female dominated, and culturally feminine, men will be considered to be less effective than women (Eagly & Karau, 2002).

However, the boomerang effect proposes a different hypothesis. This effect suggests that the extent to which observed behavior contrasts sharply with a stereotypically derived expectation, the incongruence between the observed behavior and the expected behavior can cause a boomerang effect—or a reversal of the stereotypical bias that exists (Feldman, 1981; Weber & Crocker, 1983). Feldman (1981) and Weber and Crocker (1983) have found that when an observed behavior contrasts sharply with a stereotypically derived expectation, the incongruence between the observed behavior and the expected behavior can cause a boomerang effect. For instance, if a female succeeds in becoming a leader in a highly masculine industry, rather than treating the leader as a member of the general male or female population to which he or she belongs, a "subtype" of that population is evoked, and the leader will be characterized very differently from those in the general population (Ashmore, 1981).

This perspective proposes that the characterization of those in the newly evoked subtype may be more counter-stereotypic than that of a typical member of a different group altogether. Thus, a woman's divergence from stereotypic expectations (e.g. succeeding in becoming a leader in a male-dominated industry) can create a paradox,

with the woman being viewed as having few stereotypically feminine qualities, perhaps even fewer than a typical man, and her competence and likely future success is consequently evaluated more favorably. Studies have found support for this effect—indicating an overvaluation of female as compared to male job applicants of particularly masculine types of jobs (Bigoness, 1976; Hamner, Kim, Baird, & Bigoness, 1974; Kryger & Shikiar, 1978) as well as for leadership positions (Jacobson & Effertz, 1974).

> Based on this literature, I could hypothesize that as the masculinity of the industry increases, women's success in obtaining these roles will create a boomerang effect in which their perceived effectiveness is exaggerated.

Given the competing nature of these hypotheses, I propose the following research question.

> Research Question 1: How does the masculinity of the leadership industry influence gender differences in leadership effectiveness?

Leadership Level as a Moderator

RCT proposes that perceptions of leadership are likely to be more masculine for higher status leadership positions, thereby increasing role incongruity for women in these positions (Eagly & Karau, 2002). The higher the level of leadership is, the more masculine and agentic are the expected behaviors for the leader (Hunt, Boal, & Sorenson, 1990; Lord & Maher, 1993; Martell, Parkers, Emrich, & Crawford, 1998; Eagly & Karau, 2002). As the incongruity between the female gender role and the leadership roles increases, so should the discrepancy in gender differences of leadership effectiveness. Recently, a meta-analysis was conducted on how leadership and gender stereotypes may depend on the level of the leadership position. Koenig, et al. (2011) found that greater

masculinity was associated with high-status leadership positions (e.g., upper-level managers) than mid-level positions (e.g., middle-level managers or managers in general). The authors suggest the effect appears primarily because higher level leaders are stereotypically more similar to men than women. Consistent with RCT, Eagly and Karau (2002) argued that "the incongruity between the female gender role and the leader roles is likely to be the most extreme at the highest levels of leadership" (p. 577).

Eagly and Karau (2002) also discuss literature on leader behaviors associated with other leadership levels (e.g. Pavett & Lau, 1983; Martell et al, 1998) and gender differences in leadership behaviors (Eagly, et al., 2003). This literature proposes that at the lowest leadership levels, more task-oriented, transactional behaviors are needed to be successful, and since men are more likely to engage in such behaviors, they are likely to be seen as more effective. Also, at middle leadership levels, more relational and transformational leadership behaviors are needed, and women are more likely than men to engage in such behaviors, thus women may be seen as more effective in middle management than men. Research that asked managers at different levels to rate the importance of various abilities or activities for their jobs (e.g., Alexander, 1979; Gomez-Mejia et al., 1985; Paolillo, 1981; Pavett & Lau, 1983) supports these ideas. Lower level managers reported relying on abilities involving direct supervision of employees' task involvement such as monitoring potential problems, and managing conflict. Middle level managers believed that their roles required greater relational skills like fostering cooperative effort and motivating and developing subordinates.

In a study of the attributes that people ascribe to political officeholders (Huddy & Terkildsen, 1993), agentic characteristics were perceived to be the most important for

higher level offices. Thus, Eagly and Karau (2002) concluded that, "In short, the incongruity between the female gender role and leader roles is likely to be most extreme at the highest levels of leadership. Yet, such incongruity might be somewhat lower for middle managers than for first level, or "line" managers, given the more socially complex elements of middle management activity" (pg 577). In an early meta-analysis on gender differences in leadership effectiveness, Eagly and colleagues (1995) found support for these ideas. Effectiveness assessments favored men for first-level or line leadership positions, and favored women for second or middle-level leadership. Upper level leadership positions were not included in the analysis, and thus, it is uncertain whether men or women would have been seen as more effective in those positions.

> Based on RCT, I could hypothesize that: At lower leadership levels, men will be seen as more effective; at middle leadership levels, women will be seen as more effective; and at the highest leadership levels, men will be seen as more effective.

Research and theory on double standards of competence (Foschi, 1996, 2000) proposes a different hypothesis. While double standards can produce barriers to women's career advancement (Lyness & Thompson, 2000), there is also reason to believe that these double standards can provide a basis for an advantage for women leaders who reach the highest positions. Foschi (1996, 2000) argues that a woman's presence in a top leadership position provides information about her abilities to others in the organization—that she must be exceptionally competent to have made it in such a high status and agentic leadership position. When positive evaluations regarding a leader's skills or abilities can be viewed as occurring in spite of some shortcoming, the individual is likely to be perceived as possessing a particularly high level of competence (Crocker &

Major, 1989; Rosette & Tost, 2010). A recent laboratory study using fictitious leader vignette's showed support for this unique effect (Rosette & Tost, 2010). The authors found that top female leaders were given more positive evaluations than mid-level managers because women at the top were perceived to face higher standards than men.

Based on this information, I could propose a hypothesis that is somewhat counter to RCT: that women who reach and succeed at the very top of organizations may be evaluated favorably relative to men because they have demonstrated that they have overcome double standards both to arrive in their top position and further to excel in that top position that is dominated by men and perceived to be particularly masculine.

Due to the competing nature of these hypotheses, I aim to explore the effects of leadership level on gender differences in leadership effectiveness rather than test a specific prediction.

Research Question 2: How does the level of leadership moderate gender differences in leadership effectiveness?

Sex of Raters as a Moderator

RCT proposes that the individual differences in the evaluators of leadership effectiveness should moderate gender differences (Eagly & Karau, 2002). Specifically, they discuss raters' sex as one example of an individual difference characteristic that may act as a contextual moderator based on men and women's views of the leadership role. Eagly and Karau (2002) argue that men have a more masculine construal of leadership than women do. Research by Schein (2001) supports this argument. According to Schein's (2001) review of research, studies have shown that women, compared with men,

generally have a more androgynous view of managerial roles as requiring communal qualities as well as agentic ones.

One explanation for this more androgynous view of female leaders is that women tend to have considerably greater experience with female managers than men do (Reskin & Ross, 1995). Because of their lesser experience with female managers, men may be more likely to rely upon traditional stereotypes of women and leadership when forming impressions of each social category. As a result of their more masculine perception of leadership, male raters are likely to have a stronger tendency than female raters to view women as less qualified and less effective as leaders than men.

> Based on RCT, I could hypothesize that male raters are more inclined to favor male leaders in rating their effectiveness. Women have a more androgynous view of leadership, and will rate men and women leaders similarly.

However, the social psychological literature on stereotyping supports a different hypothesis. People mentally separate others into members of another group (the out-group) or members of their own group (the in-group) to maintain a positive social identity (Brewer & Kramer, 1985). According to self categorization theory (Turner, 1985; Turner, et al., 1987) and social identity theory (Tajfel & Turner, 1979; Turner, 1985), people identify themselves as members of distinct social groups and may show an innate favoritism and bias towards other members of those groups. This sense of group membership typically fosters in-group/self-categorization in a manner that favors the in-group at the expense of the out-group.

Gender similarity is one way in which individuals may self-categorize themselves into the in-group or out-group within an organization (Tsui & O'Reilly, 1989). The

similarity-attraction paradigm argues that individuals who are more similar to each other demographically are likely to be more attracted to, and thus like each other more than those who are different from each other (Byrne, 1971). Other research suggests that demographically similar employees like and prefer to work with one another to a greater degree than with employees who are demographically different (Glaman, Jones, & Rozelle, 1996). Research on relational demography among supervisor-subordinate dyads in Taiwan found that demographic similarity in the dyads positively affected supervisors' perceptions of their subordinates' performance (Farh, Tsui, and Cheng, 1998).

Previous research has examined the effects of in-group favoritism on various aspects of the supervisor-subordinate relationship. Tsui and O'Reilly (1989) found that demographic similarity among 272 supervisor-subordinate dyads was associated with higher effectiveness as perceived by subordinates and more attraction on the part of supervisors towards subordinates. Generally, people feel more familiar with members of their in-group than with members of the out-group, and that members of the in-group are more similar than members of the out-group (in-group homogeneity). These perceptions make out-group members more at risk to stereotyping, because less is known about the out-group and differentiation of out-group members is more difficult than that of in-group members.

In performance assessments, the ratings of leader performance often come from the manager's peers, subordinates, and/ or boss. Thus, the raters may see same-sex leaders as members of their in-group, while opposite-sex leaders may be seen as out-group members. In this case, out-group members will be subject to stereotyping because a stereotype is informative. Further, when both the in-group and out-group are assessed

on the same criteria, discrimination that favors the in-group is likely (Brewer & Kramer, 1985). Thus, raters of a leader's performance may assume that the characteristics of their same sex boss are more favorable than those of an opposite sex boss, leading to discrimination against out-group bosses. In assessing a leader's performance, then, it is proposed that raters will see same-sex leaders as members of their in-group, whereas opposite-sex leaders will be seen as out-group members and will be subject to stereotyping and discrimination in favor of the in-group.

Based on this information, I would propose a hypothesis that is counter to RCT:

Men will be rated as more effective by male raters. Women will be rated as more effective by female raters.

Based on these competing hypotheses, I propose the following research question:

Research Question 3: How does the sex of raters influence gender differences in leadership effectiveness?

Nationality as a Moderator

Given the limited amount of research examining cross-cultural moderators of gender differences in leadership effectiveness, a specific hypothesis was not proposed by Eagly and Karau (2002) in their review of RCT. However, the authors did discuss the importance of examining national cultural moderators of the congruity between gender and leader roles, and their effect on prejudicial behavior. With increasing globalization and interdependence of the world's economies, national culture is becoming more, rather than less, important, especially in regards to organizational leadership. Gender role perceptions are created through an interaction between culture and social expectations, and thus perceptions of leadership along gender lines can be affected by cultural values.

This idea is consistent with social role theory (Carless, 1998), which describes individual behavior as driven by societal expectations.

Cultural values are principles that countries and regions endorse to direct people regarding appropriate behaviors in different situations, and have been characterized through what is practiced (how things are) and valued in that area (how things should be; House, Hanges, Javidan, Dorfman, & Gupta, 2004). One potentially important cultural dimension that can vary across cultures is gender egalitarianism (GE) from the Global Leadership and Organizational Behavior research (GLOBE; House, et al., 2004). The GLOBE framework represents the latest and most comprehensive attempt to conceptualize and measure culture at the national level, and it has been shown to overlap considerably with other conceptualizations of culture (i.e., Hofstede, 2001).

GE is described as, "the extent to which an organization or a society minimizes gender role differences and gender discrimination" (House, Javidan, Hanges, & Dorfman, 2002, p 5). The lower the value of GE, the more traditional are gender roles (i.e., men as providers and women as homemakers and caregivers) in that region. Additionally, in contexts with low GE, there is typically more gender discrimination in organizations, with positions that are high in the organizational hierarchy being held by more men than women (Emrich, Denmark, & Den Hartog, 2004). The higher the value of GE in a region, the less traditional are gender roles, and the less gender discrimination is present in organizations (Emrich et al., 2004). Thus, national cultural norms surrounding gender egalitarianism may influence leader's behaviors in the workplace, as well as others' perceptions of leadership behavior and effectiveness. However, there may also be cultures and practices nested within organizations of the same national culture that could

supersede the influence of national cultural values on the evaluation of leaders (Eagly & Karau, 2002).

Hofstede, Neuijen, Ohayv, & Sanders (1990) found that while organizations from different countries differ in cultural values, organizations from the same country differ in organizational practices. Weber, Shenkar, and Raveh, (1996) also found that in international and domestic mergers and acquisitions, national and organizational cultures are separate constructs that have different relationships with attitudinal and behavioral antecedents and outcomes. A recent study examined the effect of two levels of cultural distance, national and organizational, on joint venture performance (Pothukuchi, Damanpour, Choi, Chen, & Park, 2002). The authors found that aggregate indices of national and organizational culture distance influence organizational outcomes differently. National culture distance more significantly affects the macro, organizational outcomes of efficiency and competitiveness, while organizational culture distance is a better predictor of the micro level joint partners' reported satisfaction. Similarly, different levels of gender egalitarianism between the national culture and the organizational culture could lead to different outcomes regarding gender and leadership effectiveness in organizations. In sum, RCT supports the exploration of national culture as a moderator of gender differences in leadership effectiveness. Research suggests that more proximal organizational cultural variables may also play a role in these relationships.

Based on these arguments, I propose the following research question:

Research Question 4: How do nationality and national gender egalitarianism influence gender differences in leadership effectiveness?

Study Setting as a Moderator

RCT proposes that as cognitive resources become limited, raters are likely to rely on stereotypes in making judgments of leadership effectiveness (Eagly & Karau, 2002). Stereotypes are a strategy used to preserve valuable cognitive resources (e.g., Allport, 1954; Fiske & Neuberg, 1990); therefore, it is likely that biases against female leaders are more likely to emerge when raters are distracted or are under a high level of cognitive load. Research in social cognition has demonstrated that when individuals experience a cognitive load, feel tired, or are under time pressure, they tend to rely more on stereotypes to form impressions of others (e.g., Devine, 1989; Kruglanski & Freund, 1983; Macrae, Bodenhausen, Milne, & Ford, 1997; Pendry & Macrae, 1994). Many studies of assessments of leaders' performance are conducted in organizational settings. Such settings are often busy and noisy, with organizational members frequently distracted by multiple tasks, responsibilities, and interruptions (Banks & Murphy, 1985).

Yet, other studies of gender differences in leadership effectiveness are conducted in laboratory settings. These types of studies often involve a group of undergraduate students working together on a task, while being 'led' by their group's student leader (e.g. Jacobson & Effertz, 1974; Eskilson, 1975; York, 2005). Often times, the group members are solely focused on the task at hand, and thus may have few distractions when they assess the leaders' effectiveness. Given the reduced cognitive load of participants in laboratory settings compared to those in organizational settings, it seems that these individuals may be less likely to rely upon gender and leadership-related stereotypes and biases in making judgments of leadership effectiveness.

Thus, I could hypothesize that gender differences in leadership effectiveness will be greater (favoring men) in organizational settings than in laboratory settings.

However, expectation states theory and status characteristics theory (see Berger, Cohen, & Zelditch, 1966; Berger, Rosenholtz, & Zelditch, 1980; Berger, Wagner, & Zelditch, 1985; Correll & Ridgeway, 2003) provide a different hypothesis regarding how the research study setting may moderate gender differences in leadership effectiveness. According to these theories, gender is considered to be one of several diffuse status characteristics that influence expectations about the knowledge, ability, or influence of a given person. Generally, people assign greater status and task competence to men than they do women and these beliefs produce expectations that affect social interactions and the process through which leadership is enacted, which can result in women emerging less as leaders than men (Berger, Rosenholtz, & Zelditch, 1980).

Specific status characteristics are more directly related to performance expectations than diffuse status characteristics are. Specific status characteristics usually have a closer link to performance criteria than diffuse status characteristics do. For example, if given the choice between a man and a woman being more competent at repairing a car, people might expect that a man would be more competent based on his higher status position in society (diffuse status characteristic); whereas if people are given the choice between a man and woman, and they are told the woman is a certified mechanic and the man is not, a person would likely use the specific status characteristic (mechanical training) to form an expectation of competence in the female. Thus, when raters know the specific status characteristics a leader has, this should enable them to make a more accurate judgment of the leader's effectiveness than if raters only know the leader's gender (a diffuse characteristic).

Similarly, the construct of individuating information has been used to understand perceptions of gender differences in the workplace (e.g., Reskin, 2000; see also Eagly et al., 1992, 1995). Research on the influence of individuating information on perceptions of competence in men and women targets, suggests that even a small amount of individuating information can overcome the effects of stereotypes and diffuse status characteristics on judgments of others (e.g. Locksley, Hepburn, & Ortiz, 1982; Beckett & Park, 1995). Being able to observe a leader's behaviors and abilities on the job could be an essential form of individuating information for someone assessing the leader's effectiveness. This use of individuating information is quite similar to the role of specific status characteristics in that both kinds of information should minimize the role of dynamics such as stereotypes and diffuse status characteristics.

These perspectives have implications for the moderating role of research study settings on gender differences in leadership effectiveness. In field studies of leadership effectiveness, these theories propose there will be smaller gender differences in measures of leadership effectiveness because the raters presumably have extensive knowledge of their leader's specific status characteristics (i.e., their leadership experience, knowledge, abilities, etc.). In contrast, in laboratory studies on leader effectiveness, typically the raters and leaders have not known each other for very long, and thus the raters may be more likely to rely upon stereotypes of gender and leadership, and to use diffuse status characteristics (e.g. gender) to form expectations of effectiveness.

Based on this information, I could propose a hypothesis that is counter to RCT:

Gender differences in leadership effectiveness will be greater (favoring men) in laboratory settings than in organizational settings.

Given the competing nature of the hypothesis proposed by RCT, with that proposed by expectation states theory, and research on individuating information, I propose the following research question.

Research Question 5: How does the study setting influence gender differences in leadership effectiveness?

Percent of Male Raters as a Moderator

RCT proposes that sex ratios in work groups should moderate gender differences in leadership effectiveness based on the concept of tokenism (Kanter, 1977a as cited in Eagly & Karau, 2002). Tokenism proposes that as the percentage of male raters increases, women leaders' female-stereotypical qualities become more salient, and thus, their perceived 'lack of fit' in leadership positions should become stronger. Kanter's (1977a) research and theoretical development of the construct of tokenism included a case study of 20 saleswomen in a 300-person sales force at a multinational, Fortune 500 corporation. She found that tokenism has several consequences for minority group members within the workplace: higher visibility (and increased scrutiny), exaggeration of differences from majority group members, exclusion from informal workplace interactions, and assimilation. Assimilation implies that tokens are forced into stereotypical categories defined by the dominants. In other words, tokens may not be seen as they really are. Kanter (1977a) labels this as role encapsulation, which forces tokens into limited and caricatured roles (for example expectations as to what is "suitable behavior" for a female).

These findings have been replicated across a variety of settings. The first women to enter the U.S. Military Academy at West Point reported feeling highly visible, socially

isolated, and gender stereotyped (Yoder, Adams, & Prince, 1983). Similar patterns can be seen in a study of enlisted military women (Rustad, 1982), by the first women to serve as corrections officers in male prisons (Jurik, 1985; Zimmer, 1986), and by the first policewomen on patrol (Martin, 1980). Early studies of women physicians (Floge and Merrill, 1985) and academics (Young, Mackenzie, & Sherif, 1980) show similar effects of their token status. These findings are explained with the concept of numeric gender imbalance. Kanter (1977b) proposes that the damaging effects of tokenism, or numerical underrepresentation, will take place for both men and women tokens. Because tokens are numerically few, they stand out (creating performance pressures), and are isolated by the majority (who exaggerate their differentness). Thus, based on RCT and tokenism, it would seem that as the percentage of male raters increases, women will be seen as less effective than men, and as the percentage of male raters decreases (more female raters), women will be seen as more effective than men.

In addition, if men have a more masculine construal of leadership than women do (as proposed above), then it follows that the more male raters there are for a female leader, it is likely that her effectiveness will be rated as lower than if there were an equal amount of male and female raters, or a majority of female raters. Because women have a more androgynous view of leadership, it follows that the fewer the percentage of male raters (more female raters), there may be small or no differences in effectiveness between male and female leaders.

Thus, I could hypothesize that as the percentage of male raters increases, gender differences in leadership effectiveness will favor men, and when the percent of

male raters decreases, differences in effectiveness will favor women or be very small.

Yet, expectation states theory proposes a competing hypothesis (Wagner & Berger, 1997; Ridgeway, 1997). According to the theory, the biasing effect of gender status on self–other performance expectations can vary, depending on how salient and task relevant gender is in the situation compared to other social roles and status-valued social distinctions that are also salient for the participants (Ridgeway, 2001). In mixed-sex work settings, expectation states theory predicts that, the effects of gender status beliefs on performance expectations should be smaller than in settings that are particularly male or female dominated. Because leadership is generally considered to be a masculine task, when there is a majority of either male or female members in a leader's work group, this is likely to make gender more salient to the group members, and increase performance expectations in favor of male leaders. Thus, as the percentage of male raters reaches either low or high extremes, men may be seen as more effective due to the salience of gender to the context. Yet, when the percentage of male and female raters is close to equal, there should be small or nonexistent gender differences in leadership effectiveness.

These ideas compete somewhat with those proposed by Kanter (1977a; as cited by Eagly & Karau, 2002) who maintained that being a token would result in similar consequences for women and men. Research suggests that for men, being a token either has no negative effects (Budig, 2002) or actually results in more positive outcomes (Fairhurst & Snavely, 1981; Williams, 1992). Therefore, negative consequences of being a manager in a work group occupied primarily with members of the opposite sex, seems

to affect only women. Reskin (1988) explained women's more negative tokenism experiences by discussing the greater gender discrimination often directed towards women. Status characteristics theory (Berger et al., 1977) and expectations states theory (Ridgeway, 2001) also support these findings. These theories refer to the wide differential in social status between women and men, such that women are ascribed much lower status by society (Fairhurst & Snavely, 1983; Yoder, 1991), and there are higher performance expectations for men than women.

Based on this information, I would propose a hypothesis that is counter to RCT:

> There will be a curvilinear relationship between percentage of male raters and gender differences in leadership effectiveness such that as the percentage of male raters is close to 50%, gender will be less salient, and gender differences in perceived effectiveness will be small, while at the extremes, men will be seen as more effective.

Based on these competing hypotheses, I propose the following research question:

> Research Question 6: How does the percentage of male raters influence gender differences in leadership effectiveness?

Exploratory Moderators

In addition to the hypotheses and research questions presented above that are derived from RCT, I also examine three exploratory moderator variables that are important for practical reasons: type of leadership effectiveness criteria, type of rater, and type of publication. One potential moderator of gender differences in effectiveness may be the type of leadership criteria being measured. Researchers have argued that subjective performance evaluations can be particularly biased because they may be easily tainted by prejudice against women (Eagly, 2007; Frear & Paustian-Underdahl, 2011). However, a

leader cannot truly affect the behaviors of others in order to reach organizational goals unless followers accept his or her leadership. Thus, subjective performance evaluations, even if prejudiced, provide one relevant measure of leadership effectiveness. Despite the relevance of perceptual effectiveness criteria, these measures may result in larger gender differences in effectiveness outcomes than more objective measures of effectiveness (i.e., group or organizational task performance).

Heilman (2001) suggests that gender stereotypes and the expectations they produce have a tendency to affect how performance information is attended to and interpreted by organizational members. For people to reject a well-established stereotype (i.e., that men are better leaders than women), they must accept disconfirming information (i.e., that women are effective leaders too). Unfortunately, the easiest response to disconfirming information is to reject it (Heilman, 2001). Thus, evaluators of leader performance typically engage in a cognitive distortion process that enables them to pay more attention to and remember precisely what they expected to see to begin with. Research has shown that unless the quality of the work outcome is incontrovertible (as it may be using objective indicators of effectiveness), women's performance is undervalued compared to men's (see Heilman, 1983, 1995; and Nieva & Gutek, 1980).

> For these reasons, I could hypothesize that gender differences in leadership effectiveness will be greater (favoring men) when the leadership criterion is based on subjective rather than objective assessments of leaders' performance.

However, when considering the effect size of gender differences in objective effectiveness measures, it is also important to consider the nature of the time period in which the study was conducted as well as the type of task being conducted by the leader

and his or her group of followers. If the study was conducted in a time where there were very few women in leadership positions, when women were considered to be less competent and to have less legitimacy than male leaders, than even objective indicators of the unit's task performance may be biased in favor of groups led by men. Additionally, if the group task involves a particularly feminine (caregiving) or masculine (changing tires) assignment, this is likely to moderate any gender differences, even when examining objective rather than subjective effectiveness indicators. If studies examining gender and leadership effectiveness choose to utilize masculine tasks in their measures of leadership effectiveness, than objective measures may not be less biased than subjective measures of effectiveness.

For example, a recent lab study conducted by Riggio, Riggio, Salinas, and Cole (2003), instructed the leaders that, "he or she was 'in charge' of the group and that he or she had responsibility for overseeing the assembly of a metal contraption, an assembled model of which was on the table in front of them. The metal contraption was made out of different-sized steel braces held together with nuts and bolts." (p. 93). In this example, one of the indicators of leadership effectiveness was how well the leader's group completed this male-typed task—assembling a metal contraption. This study included undergraduate students who have not been trained to complete this type of task. Thus, objective measures of leadership effectiveness may not be less biased than subjective measures of effectiveness if studies examining objective effectiveness include untrained participants conducting a gender-typed task.

Based on this logic, I could hypothesize that objective and subjective measures are both susceptible to the effects of gender status beliefs and biases, and thus

there will not be differences in effect sizes based on the type of effectiveness criterion.

These discrepant perspectives lead to the following exploratory research question.

Research Question 7: How does the type of effectiveness criteria moderate gender differences in leadership effectiveness?

It is important from a practical human resource and performance management perspective to examine how the type of rater may moderate gender differences in ratings of leadership effectiveness. If for example, a specific group of raters (i.e., peers or bosses) provides a particularly strong influence on gender differences in leadership effectiveness, it is important for human resource practitioners to be aware of this potential source of bias in performance assessments. Additionally, the "file drawer" problem proposes that many studies are never published due to non-significant findings or findings that oppose the hypothesized direction proposed in the study (Rosenthal, 1979). Thus, it is important to understand how the publication type (e.g. published journal article vs. book vs. unpublished thesis/ dissertation vs. unpublished data/ white paper) may influence the direction and strength of effects.

Research Question 8: How does the type of rater moderate gender differences in leadership effectiveness?

Research Question 9: How does the type of publication moderate gender differences in leadership effectiveness?

METHOD

Literature Search and Inclusion Criteria

To gather primary studies to include in this meta-analysis, an extensive literature search was conducted to select studies published through 2011, initially using the keywords 'leadership performance' and 'leadership effectiveness'. Studies found from these keywords were manually searched for data on gender differences in leadership outcomes. Additionally, the keywords of 'leader', 'leadership', 'manager', and 'supervisor' were used, and were paired with terms such as 'gender', 'sex', 'sex differences', and 'women'. The first step in the search involved the utilization of seven major computer databases that abstract sources with relevance to gender and leadership effectiveness (Google Scholar, PsycInfo, Business Source Complete, PsycArticles, Sociological Abstracts, Academic Search Complete, and Communication and Mass Media Complete).

Additionally, I searched through numerous review articles, books, and recent Academy of Management and Society for Industrial-Organizational Psychology conference proceedings (2010-2012), as well as the reference lists of other related meta-analyses (Eagly et al., 1992; Eagly et al., 1995). In addition, I completed a manual search through journals that might have had relevant articles including the Journal of Applied Psychology, Academy of Management Journal, Personnel Psychology, Journal of Management, and the Psychological Bulletin. Finally, emails were sent to relevant listservs and research groups (OB listserv, GDO listserv, SPSP listserv, Organizations, Occupations, and Work ASA listserv, LDRNET listserv) to request in press or unpublished manuscripts and datasets. I reviewed the abstracts of the articles that were found during this search process for indications that their content would be appropriate

for this meta-analysis (e.g., they included results of an empirical study where gender and leadership effectiveness were measured in some quantitative form). The search yielded 270 potential articles and dissertations, which I reviewed for their ability to meet the specific inclusion criteria discussed below.

Extensive criteria for including studies in the meta-analysis were specified, which consisted of the following (a) the study compared male and female leaders, managers, supervisors, officers, principals, department heads, or coaches; (b) participants were at least 18 years old, and not sampled from abnormal populations; (c) the study assessed the effectiveness of at least five leaders of each sex; (d) measures of leaders' effectiveness included one of the following: performance, or leadership ability that were based on self-ratings or ratings by supervisors, peers, subordinates, or judges; ratings (by self or others) of satisfaction with leaders or satisfaction with leaders' performance; coding or counting of effective leadership behaviors by trained judges or observers; or measures of organizational productivity or group performance, and (e) the reported results had to have been sufficient either to calculate a sex-of-participant effect size or to determine the statistical significance or direction of the sex difference. This last criterion eliminated studies that provided only a multiple regression equation in which gender appeared as one of the predictors as well as studies that provided only a multivariate analysis of variance on leadership effectiveness combined with other measures. Authors were contacted for more information if the appropriate data appeared to have been collected but not reported in the paper.

If a document included data from different countries or different types of organizations, the samples of leaders were treated as separate studies, if the data were

reported separately. I carefully tracked authors of multiple studies in order to determine if the same sample and data may have been reported in multiple studies. If I determined that the same sample was used in more than one study, I only used and coded data from one of the studies (e.g. Bolman, 1992; Bolman & Deal, 1991). Primary studies that reported only measures of salary were not included because these outcomes are determined by a variety of factors (e.g., industry, supervisor preferences, job tenure) that do not necessarily reflect the leader's effectiveness. Studies that report only measures of likability or interpersonal attractiveness were also excluded. Ratings of effectiveness had to pertain to the general quality of the leaders' performance, rather than to specific stylistic features (e.g., consideration, initiation of structure). Thus, studies were excluded if leadership effectiveness was merely inferred on the basis of a measure of leadership style. Moreover, studies were omitted if they reported only subordinates' general job satisfaction (rather than their satisfaction with the leader) or if they examined only stereotypes of leaders or peoples' attitudes toward male and female leaders in general.

Lab studies which involved leaderless group discussions in which no one was designated to fill a leadership role were excluded. Also, studies of non-supervisory employees performing "in-basket" exercises, or any other kind of management simulation not involving group interaction were excluded because the participants in these studies did not assume an actual leadership role. Finally, studies reporting teachers' performance were excluded because teachers do not occupy a clear leadership position (since they are responsible for teaching and managing the well-being of children). After each study was reviewed for these criteria, the final sample consisted of 95 studies, and 99 independent effect sizes.

Coding the Studies

A thorough coding manual including instructions for coding articles and abstracting appropriate effect sizes was developed in order to aid in the coding process (see Appendix A). Lisa Walker and I independently tested this coding manual on four diverse primary studies, and then met to discuss coding inconsistencies and technical questions. Based on this discussion, we tweaked the coding manual to ease the coding process for myself and two research assistants who we trained to help me complete the coding of all subsequent primary studies. Two coders (one of them always being me) independently coded each primary study.

Coders agreed on 89.5% of the initial codes and disagreements were resolved via discussion. Appendix B provides more detailed information on rater agreement statistics for each pair of coders across three coding time periods. General information about each study was coded (see Appendix A). Additionally, the reliability coefficient (alpha) of the effectiveness measure from each study was recorded to be used to correct for measurement unreliability. Finally, all relevant information was coded to aid in the calculation of the standardized mean difference effect size (means and standard deviations, Fs, ts, or Zs, and correlations).

Effect Size Computation

Cohen's d (Cohen, 1988) was the effect size used in this study. It is the effect size for the standardized mean difference between two groups on a continuous variable (e.g., the mean difference between males and females on a continuous measure of leadership effectiveness). Thus, for each independent sample within an article, d was computed, with $d = (M_m - M_f)/s_w$; M_m = the mean for males, M_f = the mean for females, and s_w = the

pooled standard deviation. The numerator of the d statistic was the mean male rating minus the mean female rating. A positive sign indicates that men were more effective than women, and a negative sign indicates that women were more effective than men. The denominator is the pooled or averaged value of the male and female group standard deviations. Thus, a d of –.25 would indicate that the mean male score is one quarter of an averaged standard deviation lower than the mean female score. If means and standard deviations were not available, the effect size was computed from other statistics such as t, F, or r, using formulas provided by Lipsey and Wilson (2001).

To reduce computational error, the effect sizes were calculated with the aid of a computer program (Borenstein, Hedges, Higgins, & Rothstein, 2009). Researchers recommend that the best way to average a set of independent standardized mean difference effect sizes is by weighting each effect by its inverse variance (Hedges & Olkin, 1985; Sanchez-Meca &Marín-Martínez, 1998). Thus, in the current study, each effect size was weighted by its inverse variance such that effects with greater precision received greater weight (Borenstein et al, 2009). The random effects model assumes that the studies in the meta-analysis are a random sample from a population of possible studies and that the population of true effect sizes is normally distributed. In the random effects model, the variance in the distribution of observed effects is attributed to: within-study variance (mainly due to measurement artifacts), and between-study variance (due to differences of true effect sizes in the population). On the other hand, the common effects model assumes that there is one common true effect underlying the distribution of observed effects. The random effects model was used in this case, allowing effects to be weighted by their precision, and by the random effects variance component.

The Hunter–Schmidt approach was used for this meta-analysis (Hunter & Schmidt, 2004). This approach has been labeled the psychometric meta-analysis. Researchers who favor the psychometric approach to meta-analysis site the importance of measurement theory, which suggests that a substantial portion of the variability observed in the relationship between two variables across primary-level studies is mostly the result of methodological and statistical artifacts (Aguinis, 2001; Aguinis & Whitehead, 1997). If these artifacts are not controlled for upfront in the research design or after data collection, then researchers may conclude that there is an interaction, when in actuality, the source of variance across primary studies may not be substantive. This can lead to an increased likelihood of a Type I error (finding an effect that does not actually exist) in meta-analyses.

Thus, I used the recommended transformation calculations to help resolve artifacts which can be due to unreliability in the dependent variable (Hunter & Schmidt, 2004). Because not all of the individual studies provided alpha coefficients for reliability information, the effect sizes were corrected following the optimal two-stage procedure recommended by Hunter and Schmidt (2004, pp. 173–175). In the first step, individually known artifacts were corrected. The distributions of the artifacts available from the first step were then used to correct for the remaining artifacts (Hunter & Schmidt, 2004, pp. 174–175). Table 2 reports the alpha coefficients for each of the 42 studies in which it was reported, as well as the effect sizes (before and after being distribution-corrected) from studies that did not report an alpha.

Some primary studies included more than one indicator of effectiveness (e.g. satisfaction with leader and leader objective performance), ratings from multiple sources

(boss, self, peers, etc), and/ or reported effectiveness separately by the sex of the raters. In these instances, I calculated a general effectiveness construct using the formula proposed by Rosenthal and Rubin (1986), thereby removing any non-independence in the data. I calculated separate effect sizes for 11 primary studies that reported effectiveness separately by rater sex, and used these in a separate meta-analysis in order to assess the effect of this moderator. However, for all other analyses, I removed the non-independence of the effect sizes by representing each study with only one effect size. Specifically, the effect sizes were combined across sex of raters to create study level effect sizes for the overall meta-analysis.

Moderator Analyses

Multiple methods were used to assess for the need to test for moderators—the chi-square based Q statistic, the 75% rule, and the 80% credibility interval. The Q statistic tests homogeneity in the true effect sizes across studies (Hunter & Schmidt, 1990). The Q statistic is distributed as an approximate chi-squared distribution and gives an indication that moderators are present when significant. A significant Q indicates that study-level ds do not estimate a common population effect size, and thus a subsequent search for moderators is warranted (Aguinis & Pierce, 1998). If a significant Q statistic becomes non-significant when looking at separate moderator levels, or drops substantially, it suggests that the moderator explains variance in the effect. In addition, we include the between-class effect (Q_b). If the categorical moderator fully explains variance in the data, the between-class effect is significant.

Hunter and Schmidt (2004) propose a meta-analytic approach based on psychometric principles, in which artifactual sources of variance must be corrected for

prior to estimating any moderating effects. This approach—known as the 75% rule—indicates that if less than 75% of the variance in effect-size estimates is due to artifacts (i.e., unreliability), it is likely that there is substantive variance which may be explained by moderators. The Hunter and Schmidt approach assumes that if correctable factors account for at least 75% of the across-study variance, then the remaining variance is also likely to be accounted for by other uncorrectable artifacts (Aguinis, Gottfredson, & Wright, 2010). However, this method of detecting moderators has been criticized based on its high Type I error rate (Sackett, Harris, & Orr, 1987; Cortina, 2003).

Another method for testing the influence of moderators involves reporting a credibility interval around the mean across-study effect (which is obtained by weighting each primary-level study by its inverse variance). A credibility interval that includes zero or is very large suggests that moderators may be impacting the data (Arthur, Bennett, & Huffcutt, 2001). While this technique has a lower risk of committing a type I error than the 75% rule, it has very low power to detect moderators (Cortina, 2003). In a review of methods to detect meta-analytic moderators, Cortina (2003) concludes that tests that use the chi-square statistic, including the Q statistic, fall somewhere in between the 75% rule and the credibility interval on both their risk of committing a type I error, and in their power to detect moderators. Thus, in this meta-analysis all three methods are reported, while the results are discussed based on the Q statistic.

TABLE 2: Effect Sizes and Corrections for the Studies Included in the Meta-Analysis

Author (Year)	obs d	cor d	dist cor d	alpha
Hemphill (1962)	-0.150	-0.177	-0.177	.72
Hoyle & Randall (1967)	-0.200	-0.200	-0.215	
Day & Stodgill (1972)	0.170	0.170	0.183	
Bartol (1973)	0.220	0.220	0.237	
Jacobson & Effertz (1974)	-0.670	-0.670	-0.721	
Eskilson (1975)	0.090	0.090	0.097	
Francke (1975)	-0.010	-0.010	-0.011	
Petty & Lee (1975)	0.260	0.260	0.280	
Gross & Trask (1976)	-0.270	-0.270	-0.290	
Osborn & Vicars (1976)				
Sample 1	-0.300	-0.300	-0.320	
Sample 2	-0.260	0.260	-0.280	
Quinn (1976)	0.410	0.410	0.441	
Mohr & Downey (1977)	1.410	1.410	1.517	
Renwick (1977)	0.130	0.130	0.140	
King (1978)	-0.090	-0.094	-0.094	.91
Patrick (1978)	0.320	0.341	0.341	.88
Deaux (1979)	0.840	0.840	0.904	
Munson (1979)	-0.490	-0.490	-0.527	
Couch (1980)	-0.080	-0.086	-0.086	.87
Ezell et al. (1980)	-0.070	-0.086	-0.086	.67
Friedman (1980)	-0.180	-0.180	-0.194	
Rice et al. (1980)	0.200	0.200	0.215	
Schneier & Bartol (1980)	-0.040	-0.040	-0.043	
Barker (1982)	0.140	0.140	0.151	
Wexley & Pulakos (1982)	-0.200	-0.211	-0.211	.90
Gupta et al. (1983)	-0.260	-0.260	-0.280	
Caccese & Mayerberg (1984)	0.490	0.490	0.527	
Peters et al. (1984)	-0.170	-0.175	-0.175	.94
Rice et al. (1984)	0.390	0.421	0.421	.86
Yoder & Adams (1984)	0.490	0.490	0.527	
Millard & Smith (1985)	0.160	0.160	0.172	
Simpson (1985)	-0.220	-0.220	-0.237	
Morrison et al (1987)	-0.200	-0.200	-0.215	
Goktepe & Schneier (1988)	-0.250	-0.268	-0.268	.87
Horgan & Simeon (1988)	0.250	0.250	0.269	
Moehlman (1988)	-0.240	-0.240	-0.258	
Ragins (1989)	-0.260	-0.263	-0.263	.98
Schulman (1989)	-0.230	-0.234	-0.234	.97
Bolman and Deal (1991)	0.020	0.020	0.022	
Komives (1991)	0.150	0.150	0.161	

TABLE 2 (continued)

Ragins (1991)	-0.180	-0.183	-0.183	.97
Bolman and Deal (1992)				
US sample	-0.240	-0.240	-0.260	
Singapore sample	-0.170	-0.170	-0.180	
Atwater & Roush (1994)	0.050	0.050	0.054	
Davidson (1995)	-0.100	-0.100	-0.108	
Moss & Jensrud (1995)				
Sample 1	-0.230	-0.230	-0.250	
Sample 2	-0.230	-0.230	-0.250	
Nogay (1995)	-0.380	-0.422	-0.422	.81
Bass et al. (1996)				
Sample 1	-0.350	-0.370	-0.368	.90
Sample 2	-0.100	-0.110	-0.106	.89
Blanton (1996)	1.340	1.340	1.441	
Lauterbach & Weiner (1996)	-1.030	-1.144	-1.144	.81
Pratch & Jacobowitz (1996)	-0.140	-0.140	-0.151	
Daughtry & Finch (1997)	-0.250	-0.250	-0.269	
Knott & Natalle (1997)	-0.080	-0.085	-0.085	.88
Leithwood & Jantzi (1997)	-0.150	-0.162	-0.162	.86
Reyes (1997)	0.420	0.488	0.488	.74
Carless (1998)	-0.180	-0.199	-0.199	.82
Kabacoff (1998)	-0.090	-0.090	-0.097	
Pasteris (1998)	-0.220	-0.220	-0.237	
Randall (1998)	-0.390	-0.416	-0.416	.88
Sutton (1998)	0.440	0.440	0.473	
Tucker et al.(1999)	0.630	0.630	0.678	
Adams & Keim (2000)	-0.260	-0.260	-0.280	
Gore (2000)	-0.160	-0.160	-0.172	
Thompson (2000)	0.070	0.076	0.076	.84
Vilkinas (2000)	-0.010	-0.011	-0.011	.83
Bess (2001)	0.060	0.060	0.065	
Burke & Collins (2001)	-0.100	-0.100	-0.108	
Bartone et al. (2002)	-0.300	-0.344	-0.344	.76
Riggio, et al. (2003)	-0.437	-0.437	-0.470	
Rosser (2003)	-0.270	-0.285	-0.285	.90
Rivero (2003)	0.030	0.030	0.032	.90
Davy & Shipper (2004)	-0.630	-0.630	-0.678	
Douglas & Ammeter (2004)	-0.870	-0.973	-0.973	.80
Stephens (2004)	0.050	0.053	0.053	.89
Hua (2005)	0.610	0.610	0.656	
Hura (2005)	-0.190	-0.198	-0.198	.92
York (2005)	-0.160	-0.169	-0.169	.90
Duehr (2006)	-0.100	-0.100	-0.108	
Love (2007)	0.320	0.320	0.344	

TABLE 2 (continued)

Singh (2007)	0.160	0.186	0.186	.74
White (2007)	-0.320	-0.320	-0.344	
Johnson, et al. (2008)	-0.040	-0.043	-0.043	.87
Mohr & Wolfram (2008)	0.300	0.300	0.323	
Skottheim (2008)	-0.710	-0.710	-0.764	
Vilkinas et al. (2009)	-0.260	-0.265	-0.265	.96
Ayman, et al. (2009)	-0.060	-0.061	-0.061	.98
Eckert, et al. (2010)	-0.140	-0.148	-0.148	.90
Rohmann & Rowald (2009)	-0.390	-0.390	-0.420	
Vecchio & Anderson (2009)	-0.140	-0.144	-0.144	.94
Antonaros (2010)	-0.100	-0.108	-0.108	.86
Burns & Martin (2010)	-0.080	-0.092	-0.092	.76
Agezo & Hope (2011)	-0.070	-0.073	-0.073	.93
Carson (2011)	0.260	0.273	0.273	.91
Douglas (2011)	-0.120	-0.130	-0.130	.85
Elsesser & Lever (2011)	0.040	0.046	0.046	.76
Folkman & Zenger (2011)	-0.015	-0.015	-0.016	
Sparks & Kuhnert (2011)	-0.050	-0.050	-0.054	
Weighted Population Means	**-0.047**	**-0.048**	**-0.050**	
Weighted Variances	**0.001**	**0.001**	**0.001**	

Note. Obs d = the uncorrected observed effect size. Cor d = only the effect sizes from studies that reported alpha were corrected for unreliability. Dist cor d = all effect sizes were corrected for unreliability with a distribution correction. Alpha = reported alpha reliability. Alpha reported for 42 out of the 99 samples.

Categorical moderator variables were examined using sub-groups analysis, while continuous moderators (percentage of male raters and time of study) were examined using meta-analytic regression analysis. I used weighted least squares (WLS; Neter, Wasserman, & Kutner, 1989) in SPSS/ PASW 18 to estimate the continuous variables following Steel and Kammeyer-Mueller (2002) and Geyskens, Krishnan, Steenkamp, & Cunha (2009), who found it the most accurate method. The weights were the same as used in the sub-group meta-analyses, the inverse variance of each effect size. This technique requires that the reported standard error is corrected by dividing it by the square root of the mean square residual, since the weighting procedure assumes

incorrectly that the weights refer to numbers of subjects (Lipsey & Wilson, 2001). All PASW output related to statistical significance testing is thus incorrect because is it based on sample size assumptions not accurate for meta-analysis. To test the significance of each moderator, Z was calculated, which is the unstandardized regression coefficient divided by the corrected standard error, in order to determine whether it exceeded the critical value (Lipsey & Wilson, 2001).

To test Hypothesis 2, regarding the effect of time on gender differences in leadership effectiveness, both meta-analytic regression techniques as well as sub-group analyses were adopted. Theoretically meaningful sub-groups were created using data on the percent of women in management in the U.S. from 1960 to present day (Catalyst, 2012b). Primary studies included in the meta-analysis were conducted between 1962 and 2011. The sub-group categories were developed based roughly on token status percentage groups developed by Kanter (1977a). Kanter (1977b) referred to percentages of 15% or less as skewed, percentages between 15% and 40% as tilted, and percentages between 40% and 50% as balanced. Too few studies were published when women made up 15% or less of management positions, so this grouping was extended to include studies published when women made up close to 25% or less of management positions.

To test Research Question 1, regarding the extent to which the masculinity of the industry influences gender differences in leadership effectiveness, I used secondary data from the U.S. Bureau of Labor Statistics (BLS) (2011) as well as Statistics on Women in the Military (2011). Heilman (1983) suggests that one way to conceptualize the masculinity or femininity of a job is by the percentage of men and women occupying that job**.** The BLS database reports the average percent of men and women in different

industries, and the Statistics on Women in the Military report includes the percent of women in the U.S. military. Additionally, in order to compare the findings of this meta-analysis with previous studies, the same categories of industries that were used in the Eagly, et al., (1995) meta-analysis of gender differences in leadership effectiveness were used in the present study. To gather secondary data to test Research Question 4, regarding the moderating effects of gender egalitarianism present within the country the leaders reside in, I gathered scores from the Global Leadership and Organizational Behavior research (GLOBE) (House, et al., 2004). This data was added to the coding database in order for gender egalitarianism at the national level to be included for sub-group comparisons.

Associations among Moderators

Many researchers overlook the importance of detecting potentially spurious relationships among meta-analytic moderators (Lipsey, 2003). Lipsey (2003) noted that confounding can occur if moderator variables are related to each other and to effect size estimates in meta-analytic research. Such confounding can make the interpretation of the results of any one moderator misleading. For example, it may appear that objective indicators of leadership effectiveness show that men are more effective leaders than women, but if most or all of the studies using objective measures of effectiveness are also older studies (which show the same effect on d), the conclusions drawn regarding objective indicators of effectiveness may be spurious.

In order to address this possibility, I examined the interrelationships between all of the moderators in this meta-analysis. A series of chi-square tests of independence were used instead of correlations because the majority of moderators in this study are

categorical. The two continuous moderators (e.g. time and percent of male raters) were broken into sub-groups so they could be included in these tests of association. Due to the large number of tests conducted, I used a p-value of .01 to reduce the likelihood of Type I error.

Assessing Publication Bias

Publication bias occurs when the primary studies that appear in a meta-analysis are unrepresentative of the population of studies on those hypotheses, because they only include studies that have been published in the literature (Aguinis et al, 2010). Rosenthal (1979) proposed a method of examining meta-analytic publication bias, or what he called the "file drawer problem". Rosenthal's solution involves conducting a file drawer analysis in which the meta-analyst calculates the number of unpublished studies that would be needed to make the meta-analytically derived mean effect size non-significant. This number is called the failsafe N. When the failsafe N is relatively small compared to the number of articles included in the meta-analysis, this suggests that the meta-analytic results may be affected by publication bias (Becker, 2005). If the failsafe N is relatively large, this indicates that the meta-analytic results are valid, and practically important and meaningful (Rosenthal, 1979).

Despite the popularity of this method, it has been criticized for its assumption that excluded studies show a null result whereas many may show a result in the opposite direction (Aguinis et al, 2010). Additionally, the failsafe N ignores primary-level study sample size information and variance information (e.g., it assumes that the effect of adding N studies showing a null effect would be the same even if each of those studies had different sample sizes and variances) (Becker, 2005). A relatively recent

advancement in the detection of publication bias in meta-analyses is called the trim-and-fill technique (McDaniel, Rothstein, & Whetzel, 2006 as cited in Aguinis et al, 2010). The main goal of the trim-and-fill technique is to assess how much the value of the estimated summary effect size might change if there are missing studies.

The trim-and-fill technique was conducted in this meta-analysis by creating and inspecting a funnel plot with study-level effect-size estimates on the horizontal axis against a measure of the standard error on the vertical axis to determine whether there is (asymmetry) or is not (symmetry) possible publication bias. If the funnel plot is asymmetrical, then the trim-and-fill method imputes the missing study-level effect-size estimates needed to make the funnel plot symmetrical, adds them to the meta-analysis, and calculates a trim-and-fill adjusted mean effect-size estimate. This allows the researcher to compare their observed effect size with the adjusted effect size in order to see the potential effect publication bias may have on the findings of the study.

Despite the advantages provided by this new technique of detecting publication bias, there are some limitations. The biggest concern is that publication bias is not the only source of asymmetry in funnel plots. Alternative sources of asymmetry include true heterogeneity (i.e., substantive factors related to effect size), artifacts (e.g., heterogeneity due to poor choice of effect measure), data irregularities (e.g., poor methodological design of small studies, inadequate analysis, fraud), and chance (Egger, Smith, Schneider, & Minder, 1997; Terrin, Schmid, Lau, & Olkin, 2003). Both the failsafe N and the trim-and-fill methods of detecting publication bias were conducted and reported in this study.

RESULTS

A total of 99 effect sizes from 58 journal publications, 30 unpublished dissertations or theses, 5 books, and 6 other sources (e.g. white papers, unpublished data) were examined in this meta-analysis. The sample sizes ranged from 10 to 60,470 leaders and the mean sample size across all the samples was 1,011 leaders (SD = 6,151). The majority of samples reported data from studies conducted within the United States or Canada (86%). The mean age of leaders across the 40 samples in which age was reported was 39.04 years (SD = 9.72). Table 3 includes study characteristics for the primary studies examined in the meta-analysis, including the observed, uncorrected mean effect sizes (d), the number of participants in each analysis (N), as well as the codes for the source, industry, setting, level, rater, content, and subjective type of content for each study. The reference list for all of the primary studies included in this meta-analysis is included in Appendix C.

Magnitude of Gender Differences

Table 4 includes a summary of the effect sizes for gender differences in overall effectiveness. Figure 1 shows the distribution of effect sizes, which is approximately normal and centered around 0. The overall analysis of effectiveness measures resulted in a mean corrected d of -.05 (K = 99, N = 101,676). I examined the data for any extreme outliers (+/- 3 SDs), and found two effect sizes that met this criteria (d = 1.44, N =30 and d= 1.52, N= 40). Hunter and Schmidt (2004) argue that when sample sizes of outliers are small to moderate, extreme outliers can occur due to sampling error. They note that such outliers should not be removed from the data in order to avoid an overcorrection of sampling error. Additionally, the purpose of a meta-analysis is to summarize an entire

TABLE 3: Study Characteristics for the Studies Included in the Meta-Analysis

Author (Year)	d	N	Source	Industry	Setting	Level	Rater	Content	Subjective Type
Hemphill (1962)	-0.150	232	3	2	3	3	7	2	5
Hoyle & Randall (1967)	-0.200	30	1	2	1	3	2	2	2
Day & Stodgill (1972)	0.170	73	1	4	1	1	2	2	1
Bartol (1973)	0.220	24	3	6	2	1	6	1	
Jacobson & Effertz (1974)	-0.670	20	1	6	2	1	7	3	1
Eskilson (1975)	0.090	48	2	6	2	1	6	1	
Francke (1975)	-0.010	463	2	1	1	1	7	2	2
Petty & Lee (1975)	0.260	164	1	2	1	1	2	2	4
Gross & Trask (1976)	-0.270	189	3	2	1	3	7	2	2
Osborn & Vicars (1976)									
Sample 1	-0.300	34	1	3	1	1	2	2	4
Sample 2	-0.260	39	1	3	1	1	2	2	4
Quinn (1976)	0.410	161	2	2	1	3	1	2	2
Mohr & Downey (1977)	1.410	40	1	4	1	4	1	2	1
Renwick (1977)	0.130	95	1	1	1	1	3	2	2
King (1978)	-0.090	48	2	2	1	3	2	2	2
Patrick (1978)	0.320	85	2	2	1	1	2	2	4
Deaux (1979)	0.840	82	1	1	1	1	1	2	1
Munson (1979)	-0.490	65	1	5	1	1	2	2	4
Couch (1980)	-0.080	275	2	3	1	2	1	2	2
Ezell et al. (1980)	-0.070	304	1	5	1	3	1	2	2
Friedman (1980)	-0.180	256	2	1	1	1	5	2	1
Rice et al. (1980)	0.200	72	1	6	2	1	3	3	2
Schneier & Bartol (1980)	-0.040	52	1	6	2	1	5	2	1
Barker (1982)	0.140	70	2	2	1	1	2	2	1

TABLE 3 (Continued)

Wexley & Pulakos (1982)	-0.200	286	1	6	1	1	2	2	2
Gupta et al. (1983)	-0.260	158	1	1	1	4	2	2	2
Caccese & Mayerberg (1984)	0.490	231	1	2	1	1	1	2	1
Peters et al. (1984)	-0.170	637	1	1	1	2	3	2	1
Rice et al. (1984)	0.390	1110	1	4	1	1	7	2	2
Yoder & Adams (1984)	0.490	148	1	4	1	1	1	2	1
Millard & Smith (1985)	0.160	38	1	3	1	4	2	2	4
Simpson (1985)	-0.220	116	2	2	1	3	2	2	2
Morrison et al (1987)	-0.200	357	3	1	2	3	5	2	2
Goktepe & Schneier (1988)	-0.250	28	1	6	2	1	2	2	1
Horgan & Simeon (1988)	0.250	55	4	1	1	1	1	1	
Moehlman (1988)	-0.240	32	2	2	1	3	7	2	2
Ragins (1989)	-0.260	110	1	6	1	2	2	2	1
Schulman (1989)	-0.230	69	2	2	1	3	7	2	1
Bolman and Deal (1991)	0.020	187	1	2	1	1	7	2	1
Komives (1991)	0.150	74	1	2	1	1	2	2	5
Ragins (1991)	-0.180	110	1	6	1	2	2	2	2
Bolman and Deal (1992)									
US sample	-0.240	205	1	2	1	3	7	2	1
Singapore sample	-0.170	274	1	2	1	3	7	2	1
Atwater & Roush (1994)	0.050	550	1	6	2	1	1	2	5
Davidson (1995)	-0.100	51	2	2	1	1	1	2	2
Moss & Jensrud (1995)									
Sample 1	-0.230	551	1	2	1	4	7	2	1
Sample 2	-0.230	143	1	2	1	1	7	2	1
Nogay (1995)	-0.380	61	2	2	1	3	7	2	2
Bass et al. (1996)									
Sample 1	-0.350	229	1	1	1	4	2	2	5

TABLE 3 (Continued)

Sample 2	-0.100	285	1	6	1	4	2	2	5
Blanton (1996)	1.340	30	2	2	1	3	7	2	1
Lauterbach & Weiner (1996)	-1.030	21	1	1	1	2	1	2	1
Pratch & Jacobowitz (1996)	-0.140	48	1	6	2	1	4	2	1
Daughtry & Finch (1997)	-0.250	102	1	2	1	4	7	2	1
Knott & Natalle (1997)	-0.080	296	1	1	1	4	3	2	2
Leithwood & Jantzi (1997)	-0.150	147	1	2	1	3	2	2	2
Reyes (1997)	0.420	73	2	2	1	1	1	2	1
Carless (1998)	-0.180	304	1	1	1	3	7	2	2
Kabacoff (1998)	-0.090	1800	4	1	1	4	3	2	1
Pasteris (1998)	-0.220	124	2	2	1	3	7	2	2
Randall (1998)	-0.390	279	2	4	1	4	1	2	2
Sutton (1998)	0.440	80	2	6	2	1	7	2	2
Tucker et al.(1999)	0.630	727	1	3	1	3	1	2	5
Adams & Keim (2000)	-0.260	21	1	6	1	3	7	2	2
Gore (2000)	-0.160	14	2	2	1	3	2	2	1
Thompson (2000)	0.070	57	1	2	1	4	2	2	2
Vilkinas (2000)	-0.010	509	1	1	1	2	7	2	1
Bess (2001)	0.060	24	2	2	1	3	2	2	1
Burke & Collins (2001)	-0.100	1031	1	6	1	4	1	2	2
Bartone et al. (2002)	-0.300	855	1	4	1	1	3	3	
Riggio, et al. (2003)	-0.437	58	1	6	2	1	7	1	
Rosser (2003)	-0.270	28	4	6	3	4	7	2	2
Rivero (2003)	0.030	42	2	6	1	2	3	3	
Davy & Shipper (2004)	-0.630	1035	4	1	1	2	3	2	1
Douglas & Ammeter (2004)	-0.870	26	1	2	1	4	7	2	1
Stephens (2004)	0.050	155	2	6	1	2	1	2	2
Hua (2005)	0.610	10	2	2	1	3	6	1	
Hura (2005)	-0.190	172	2	1	1	4	5	2	1

TABLE 3 (Continued)

York (2005)	-0.160	46	2	6	2	1	5	2	5
Duehr (2006)	-0.100	186	2	6	3	4	7	2	2
Love (2007)	0.320	132	2	2	1	3	2	2	2
Singh (2007)	0.160	340	1	1	1	2	1	2	1
White (2007)	-0.320	47	2	2	1	3	7	2	2
Johnson, et al. (2008)	-0.040	109	1	6	1	4	2	2	1
Mohr & Wolfram (2008)	0.300	140	1	1	1	1	1	2	1
Skottheim (2008)	-0.710	36	3	6	3	1	2	2	5
Vilkinas et al. (2009)	-0.260	48	1	1	1	2	1	2	1
Ayman, et al. (2009)	-0.060	109	1	6	1	2	2	2	1
Eckert, et al. (2010)	-0.140	4019	1	1	1	4	7	2	2
Rohmann & Rowald (2009)	-0.390	45	1	1	1	4	2	2	5
Vecchio & Anderson (2009)	-0.140	1221	1	6	1	2	7	2	1
Antonaros (2010)	-0.100	199	2	2	1	3	2	2	2
Burns & Martin (2010)	-0.080	14	1	2	1	3	2	2	1
Agezo & Hope (2011)	-0.070	52	1	2	1	3	2	2	2
Carson (2011)	0.260	121	2	1	1	4	7	2	1
Douglas (2011)	-0.120	26	1	1	1	1	7	2	2
Elsesser & Lever (2011)	0.040	60470	1	6	1	4	2	2	2
Folkman & Zenger (2011)	-0.015	7280	4	1	1	4	7	2	2
Sparks & Kuhnert (2011)	-0.050	8441	4	1	1	3	3	2	1

Note. d = uncorrected effect size. N = sample size of relevant effect size. Source of effect size is coded as 1= journal, 2= dissertation or thesis, 3= book, 4= other (white paper, etc). Industry is coded 1=business, 2=education, 3=government, 4= military, 5= social service/ social work, 6= mixed. Setting is coded as 1= organizational, 2= laboratory, 3= mixed. Level is coded as 1= first-line supervisor, 2= middle manager, 3= upper level leader/ executive, 4=mixed. Rater is coded as 1= self, 2= subordinate(s), 3= boss (es), 4= peers, 5= judges/ trained observers, 6= objective counting device, 7= mixed or unclear. Content of effectiveness is coded as 1= objective, 2= subjective, 3=mixed. Subjective Type is the type of subjective effectiveness and is coded as 1) overall effectiveness/ performance, 2) ability, 3) effort/ motivation to perform, 4) satisfaction with leader, and 5) other or mixed.

TABLE 4: Meta-Analysis of Overall Gender Differences and by Time of Study

	Mean d	Cor d	K	N	Var.	% Art.	95% CI	80% CR	Q	Q_b
Overall leader effectiveness	-.047	-.05	99	101676	.001	23.6	-.102, .001	-.084, -.017	415.3**	
Time of Study									410.88**	3.349
1962-1981 (15.6%-26.2 %)	.044	.046	23	4363	.004	41.748	-.075, .167	-.033, .088	52.697**	
1982-1995 (26.2%-38.7%)	-.053	-.055	25	5651	.003	39.404	-.159, .05	-.123, .102	60.908**	
1996-2003 (38.7%-50.6%)	-.058	-.064	25	7183	.003	19.764	-.170, .043	-.134, .107	121.431**	
2004-2011 (50.6%-51.4%)	-.091	-.097	26	84478	.003	14.218	-.197, .002	-.162, .110	175.833**	

Note: Mean d is the observed d across studies, cor d is corrected for measurement reliability in effectiveness, K is the number of studies, N is the number of participants, Var. is the variance of the corrected d, % Art. is the percentage of variance due to the artifacts of sampling error and measurement unreliability, 95% CI is a 95% confidence interval, 80% CR is an 80% credibility interval, Q is the Chi-square test for homogeneity of effect sizes, Q_b is the between group test of homogeneity. Percentages listed with the year of publication indicate the approximate percentage of women in management positions in the U.S. (Catalyst, 2012b). * = p< .05, ** = p<.01.

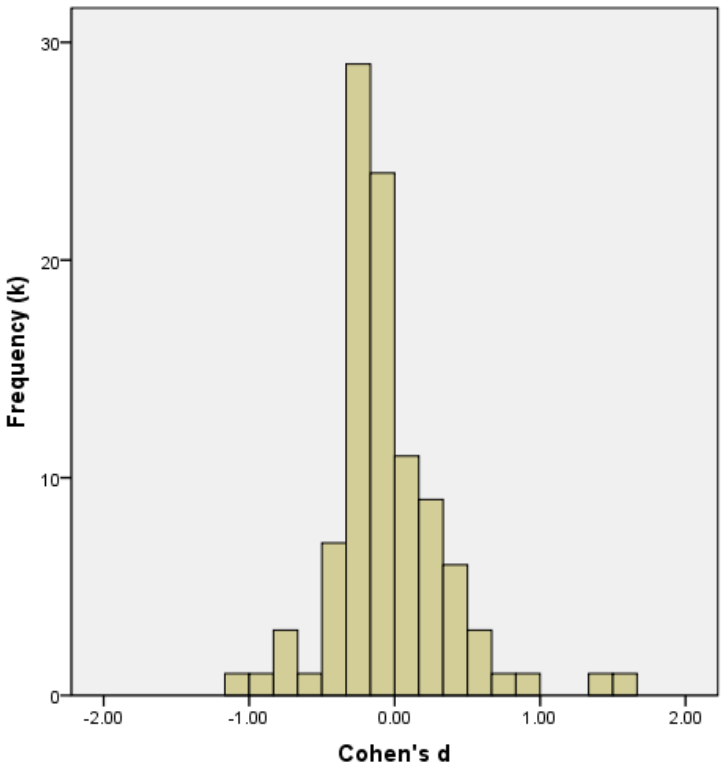

FIGURE 1. Histogram of Effect Sizes.

Note. Distribution (histogram) of the 99 ds of gender differences in leadership effectiveness included in the meta-analysis.

body of literature, thus only very extreme outliers should be removed from the data (Hunter & Schmidt, 2004). I re-ran the meta-analysis with these two effect sizes removed from the data, and the overall effect size changed slightly—by .01—becoming d = -.06. Due to the small sample sizes associated with these outliers, and to the small change in the summary effect size that resulted from removing the effects from the data, I did not eliminate them from the data.

The overall effect size of d = -.05 and is not significantly different from zero. Artifacts due to unreliability and sampling error explained only 23.6% of the variance, suggesting sufficient heterogeneity to search for moderators (Hunter & Schmidt, 2004).

The 80% credibility interval (Whitener, 1990) did not include zero suggesting that there is not enough heterogeneity to warrant a search for moderators. However, the Q test of homogeneity did indicate that moderation is likely (i.e., Q = 415.3 ($p < .01$)). This suggests that there is substantial variation in estimated population values and that in some cases males are more effective (positive values) and that in a somewhat greater proportion of cases females are more effective (negative values). Overall, the data support the hypothesis (Hypothesis 1) of relatively small differences in measures of leadership effectiveness.

Moderator Analyses

Time of Study as a Moderator

To test Hypothesis 2, that there will be greater gender differences (favoring men) seen amongst older studies, and smaller gender differences, or differences that favor women seen amongst newer studies, both sub-group analyses and meta-analytic regression techniques were utilized. The approximate percentages of women in management per each time-based sub-group can be seen in Table 4. The time categories exhibited a non-significant moderating effect on gender differences in leadership effectiveness ($Q_b = 3.349$, $p = .34$). Yet, the pattern of effects does support the hypothesis. Men were seen as more effective leaders in the oldest group of studies (1962-1981) with d= .05 (K= 23, N= 4,363). In all subsequent years, women were seen as more effective than men, with the effect growing in magnitude over time. The gender difference in leadership effectiveness between 1982 and 1995, is d= -.06 (K= 25, N= 5,651). The effect size held steady between 1996 and 2003 with a d= -.06 (K= 25, N=

7,183), and between 2004 and 2011 the effect increases slightly to a d= -.10 (K= 26, N= 84,478).

Meta-analytic weighted least squares regression techniques were also used to examine the effect of the continuous variable of time on overall gender differences in leadership effectiveness. The unstandardized beta term for time is not statistically significant (ß = .001, $p > .05$, R^2=.005), however the pattern of the effect is consistent with Hypothesis 2 (see Figure 2). A post-hoc power analysis was conducted to determine if a low sample size (99 studies) could explain the non-significance of the effect. Consistent with this idea, the observed power was only .35. Additionally, the analysis revealed that 431 studies would have been needed to detect an effect of this size. Overall, the sub-group analyses and the scatter plot of the continuous variable of time show some support for the hypothesis proposed by RCT (Eagly & Karau, 2002) that male leaders are seen as more effective in older studies, while female leaders are seen as more effective in newer studies. Yet, it should be noted that while in the expected direction, the effect is quite small and non-significant.

Industry as a Moderator

Research Question 1 asked how the industry being studied and the masculinity of each industry influences gender differences in leadership effectiveness. Industry exhibited a significant moderating effect on gender differences in leadership effectiveness ($Q_b = 11.72$, $p < .05$). Consistent with RCT (Eagly & Karau, 2002), industries that are more masculine in nature and male dominated numerically tended to show that male leaders are more effective (Table 5). Government organizations (37.3% female) have a d= .27 (K= 5, N= 1,113), and military organizations (14.5% female) have a d= .12 (K= 6,

N= 2,505). In addition, as expected based on RCT, industries that are more feminine and female dominated have effect sizes that show

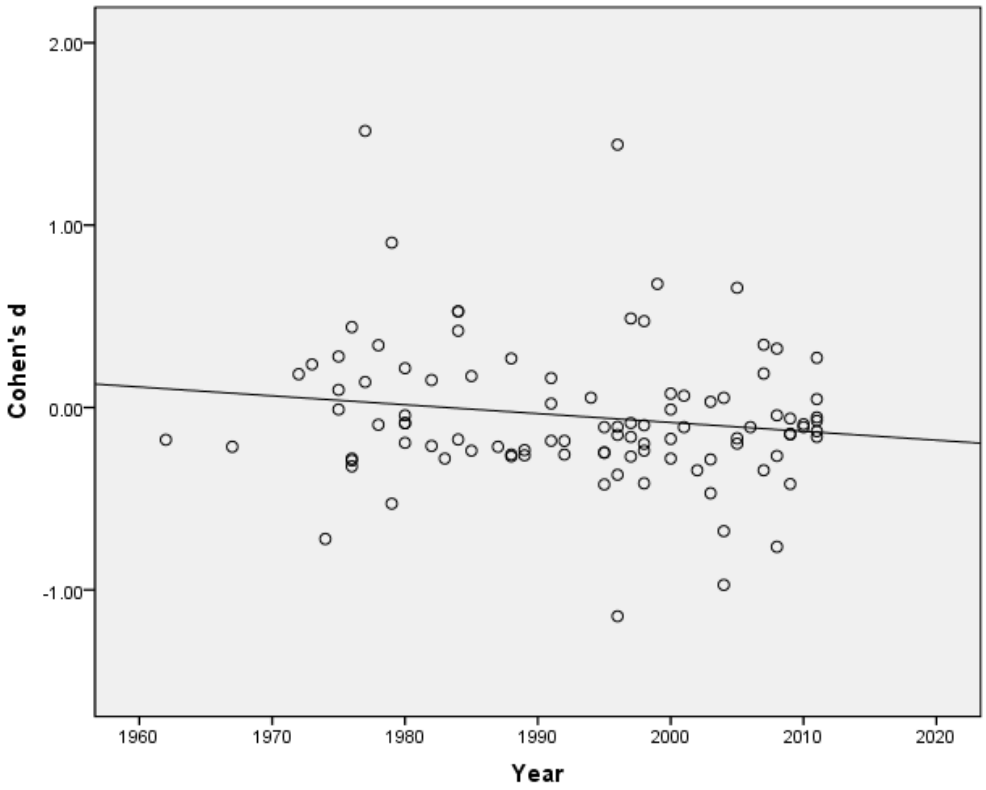

FIGURE 2. Scatter Plot of Time of Study and Effect Sizes.

women as more effective than men. Females were seen as more effective in social service organizations (85% female) with a d= -.23 (K= 2, N= 369) and as slightly more effective in education organizations (68.4% female) with a d= -.03 (K= 36, N= 4,051). The former effect should be interpreted cautiously given its small number of studies. Interestingly, in business settings which are made up of 42.5% females, women are also seen as more effective with a d= -.12 (K= 25, N= 28,440). Finally, in studies that included leaders from a variety of industries, women are slightly more effective than men with a d= -.08 (K=

25, N= 65,197). Overall, the results support the hypothesis proposed by RCT (Eagly & Karau, 2002), and do not support the hypothesis proposed by the boomerang effect (Feldman, 1981; Weber & Cracker, 1983).

Leadership Level as a Moderator

Research Question 2 asked how the level of leadership influences gender differences in leadership effectiveness. Leadership level exhibited a significant moderating effect on gender differences in leadership effectiveness ($Q_b = 10.71$, $p < .05$). The results of a sub-group analysis are partially consistent with the hypothesis proposed by RCT based on the literature on abilities associated with leadership levels (e.g. Pavett & Lau, 1983; Martell et al, 1998) and gender differences in leadership behaviors (Eagly, et al., 2003), and are consistent with the hypothesis developed by the double standards of competence model (Foschi, 2000) (Table 5). Men are seen as slightly more effective than women in lower leadership levels/ supervisor positions with a d= .07 (K= 37, N= 7,421). Women are more effective in middle management positions with a d= -.17 (K= 12, N= 4,570). However, inconsistent with the hypothesis developed by Eagly and Karau (2002) and consistent with the double standards of competence model (Foschi, 2002), women are slightly more effective than men in upper level leadership positions with a d= -.04 (K= 28, N= 12,364). In studies that included leaders from two or more of these leadership levels, women are also seen as more effective, with a d= -.12 (K= 22, N= 77,321).

Sex of Raters as a Moderator

In order to test Research Question 3, regarding the nature with which the sex of raters affects gender differences in leadership effectiveness, separate effects were calculated for the 11 primary studies that reported effectiveness separately by rater sex. In

all other analyses, the effects from these 11 studies were combined such that Cohen's d reflected ratings by men and women combined in a way that removed any interdependence in the data. In this meta-analysis, sex of raters exhibited a non-significant moderating effect on gender differences in leadership effectiveness ($Q = 1.15$, $p = .56$). Despite this non-significant effect, the pattern of effect sizes is consistent with the hypothesis developed from self categorization theory (Turner, 1985; Turner et al., 1987) and social identity theory (Tajfel & Turner, 1979; Turner, 1985). Raters generally perceived same-sex leaders as more effective than opposite sex leaders (Table 5).

Nationality as a Moderator

Research Question 4 asked how nationality and national gender egalitarianism influence gender differences in leadership effectiveness. Sub-groups were created by grouping countries within the same region together (see Table 5). Nationality exhibited a non-significant moderating effect on gender differences in leadership effectiveness ($Q_b = 4.29$, $p = .63$). Results show that there are small differences between each of the sub-groups, and all of the effects are in the same direction (slightly favoring women).

Setting as a Moderator

Research Question 5 asked how the study setting influences gender differences in leadership effectiveness. Study setting exhibited a non-significant moderating effect on gender differences in leadership effectiveness ($Q_b = 1.73$, $p = .42$). Additionally, the pattern of effects was not consistent with either of the competing hypotheses proposed by RCT or individuating information (Table 5). Studies conducted within laboratories and within organizational settings both show that women are seen as slightly more effective

TABLE 5: Meta-Analysis for Gender Differences by Industry, Level of Leadership, Rater Sex, Nationality, and Setting

	Mean d	Cor d	K	N	Var.	% Art.	95% CI	80% CR	Q	Q_b
Industry									283.15**	11.72**
Business (42.5% female)	-.098	-.106	25	28440	.002	23.657	-.194, -.017	-.163, -.048	101.450**	
Education (68.4% female)	-.030	-.033	36	4051	.002	47.997	-.128, .063	-.095, .029	72.921**	
Government (37.3% female)	.253	.265	5	1113	.016	16.733	.017, .512	.103, .426	23.905**	
Military (14.6% female)	.108	.117	6	2505	.011	11.428	-.085, .319	-.015, .249	43.754**	
Social Service (85% female)	-.192	-.225	2	369	.033	49.059	-.579, .129	-.456, .006	2.038	
Mixed	-.076	-.08	25	65197	.003	61.409	-.184, .024	-.148, -.012	39.082	
Level of leadership									377.72**	10.71**
Lower level supervisors	.066	.069	37	7421	.002	40.997	-.028, .166	.006, .132	87.811**	
Middle managers	-.163	-.172	12	4570	.005	21.127	-.310, -.034	-.262, -.082	52.066**	
Upper level leaders	-.038	-.042	28	12364	.003	24.319	-.150, .066	-.112, .029	111.022**	
Mixed	-.116	-.123	22	77321	.003	16.559	-.228, -.019	-.192, -.055	126.817**	
Rater Sex									433.24**	1.14
Female	-.089	-.093	11	846	.008	49.58	-.272, .085	-.210, .023	20.169*	
Male	.038	.042	11	775	.009	100	-.143, .227	-.079, .163	6.872	
Mixed	-.038	-.048	91	99139	.001	22.156	-.102, .005	-.084, -.013	406.198**	
Nationality									357.04**	4.29
Asian (GE= 2.98)	-.029	-.035	3	662	.02	46.117	-.313, .243	-.217, .147	4.337	
Australian (GE= 3.4)	-.09	-.100	2	813	.022	83.678	-.390, .190	-.290, .089	1.195	
European (GE= 3.08)	-.031	-.041	3	221	.039	25.314	-.428, .346	-.294, .211	7.901*	
US/ Canadian (GE= 3.52)	-.041	-.044	85	88260	.001	24.718	-.101, .013	-.215, -.141	339.830**	
Mixed	-.167	-.178	4	11595	.011	79.340	-.389, .032	.351, .625	3.781	
Setting									412.04**	1.734
Laboratory	-.021	-.023	13	1425	.008	100.000	-0.203, .158	-141, .095	11.485	
Organizational	-.043	-.046	82	99769	.001	20.358	-0.101, .009	-.082, -.010	397.884**	
Mixed	-.207	-.233	4	482	.021	100.000	-0.515, .049	-.417, -.049	2.672	

76

TABLE 5 (Continued)

Note: Mean d is the observed d across studies, cor d is corrected for measurement reliability in effectiveness, K is the number of studies, N is the number of participants, Var. is the variance of the estimate of the corrected d, % Art. is the percentage of variance due to the artifacts of sampling error and measurement unreliability, 95% CI id a 95% confidence interval, 80% CR is an 80% credibility interval, Q is the Chi-square test for homogeneity of effect sizes, Q_b is the between group test of homogeneity. * = p< .05, ** = p<.01.

leaders than men.

Percent of Male Raters as a Moderator

Research Question 6 asked how the percentage of male raters in a study influences gender differences in leadership effectiveness. Meta-analytic weighted least squares regression analyses were used to examine the centered continuous variable. Consistent with expectation states theory (Ridgeway, 2001), the squared term is statistically significant (ß =.0001, $p < .01$), which indicates that percentage of male raters is curvilinearly related to gender differences in leadership effectiveness and that its positive sign indicates a U-shape (see Table 6). The squared term explains significantly more variance in the effect size than the non-squared term, providing support for the hypothesis that when there are an extreme number of male or female raters, gender becomes salient to the leadership task, and men are seen as more effective leaders than women. While this is a small effect, the scatter plot also supports this proposed relationship (see Figure 3). Overall, the results support the hypothesis based on expectation states theory (Ridgeway, 2001) and partially support the hypothesis proposed by RCT (Eagly & Karau, 2002).

TABLE 6: Percent of Male Raters as a Moderator

Variable	Step 1		Step 2	
	β	Z	β	Z
Percent of male raters	-.003	3.00**	.01	2.43*
Percent of male raters X Percent of male raters			.0001	2.53**
R^2	.083*		.159*	

Note. *p<.05, **p<.01

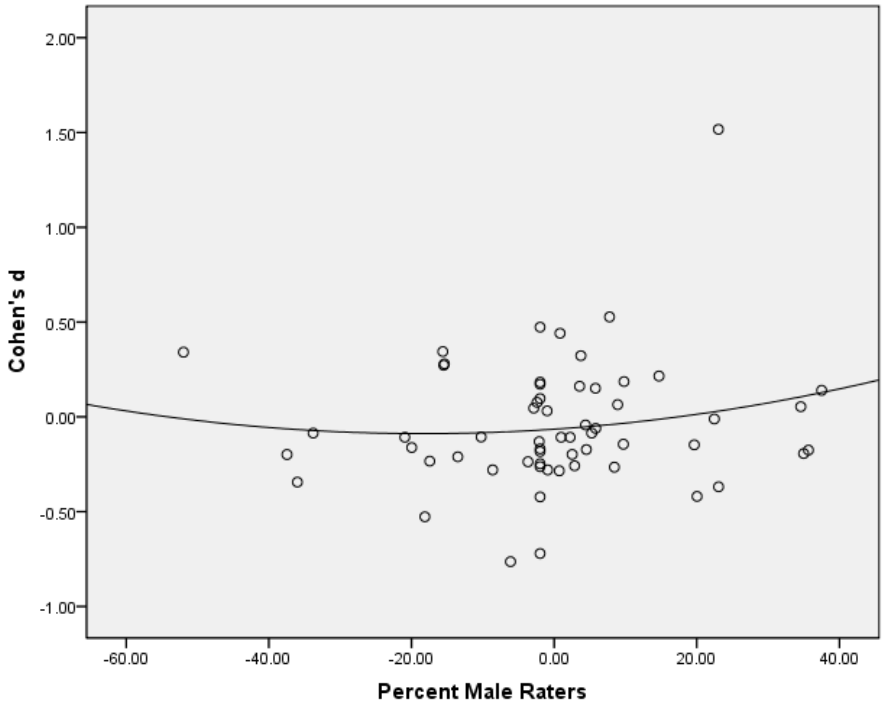

FIGURE 3. Scatter Plot of Percent of Male Raters and Effect Sizes

Exploratory Moderators

To examine Research Question 7 regarding the influence of effectiveness content, I compared the effects of objective versus subjective effectiveness measures on gender differences in leadership effectiveness (see Table 7). Type of effectiveness exhibits a non-significant moderating effect on gender differences in leadership effectiveness (Q_b = 1.77, p = .41). Results for objective versus subjective measures are reasonably consistent with overall effectiveness ratings in which gender differences were quite small (although females score slightly higher on average for subjective assessments, while males score slightly higher on average for objective assessments).

Additionally, given the broad range of types of subjective measures included across the primary studies and that the Q statistic associated with the subjective measures

indicated the presence of further moderators within the subgroup ($Q = 380.88$, $p < .01$), I examined gender differences in leadership effectiveness by type of subjective measure (see Table 4). Measures of perceptions of leadership ability showed that women leaders are seen as more able leaders than men with d= -.08 (K= 38, N= 81,102). Perceptions of leaders' performance also showed a tendency for women to be rated slightly higher than men with d= -.05 (K= 40, N=17,688). Assessments of satisfaction with the leader showed that raters were slightly more satisfied with male leaders than female leaders with d= .04 (K= 6, N= 425). Again, all of these effects are quite small.

The within subgroup Q statistics for each of these groups provided support that additional moderators are affecting these groups, and warranted a search for further moderators (Table 7). Thus, two hierarchical subgroups were created to further examine the interaction between passage of time on subjective versus objective measures of effectiveness. One group included studies conducted between 1962 and 1995, because during this time women were still tokens within companies, making up close to 40% or less of management positions. The second group consisted of studies conducted between 1996 and 2011, a time when women were considered to make up a 'balanced' proportion of management positions (40% or more).

Results show that the only shifts in the direction of gender differences in subjective and objective effectiveness ratings across these two time periods occurred for perceptions of leaders' performance, as well as for objective indicators of effectiveness. In the older group of studies, there is no gender difference in perceptions of leadership performance with d= .001 (K= 19, N= 3,292), while in the newer studies, performance evaluations show a tendency for women to be seen as more effective with a d= -.08 (K=

21, N= 14,396). When examining shifts in gender differences in objective effectiveness assessments, men are more effective in older studies with d= .197 (K= 3, N= 127), while women are more effective in newer studies with d= -.26 (K= 2, N= 68). However, these values should be interpreted with caution because the number of studies and sample sizes are small.

The effects of rater type (Research Question 8) and type of publication (Research Question 9) were also explored using sub-group analyses. Rater type exhibited a significant moderating effect on gender differences in leadership effectiveness (Q_b = 24.45, $p < .01$). As can be seen in Table 7, women are seen as more effective than men when rated by bosses (d= -.17, K= 9, N= 13,273), by subordinates (d= -.08, K= 32, N= 63,450), by judges/ trained observers (d= -.18, K= 5, N= 882), and by a mixed group of raters (d= -.18, K= 5, N= 882). Men are seen as more effective when rated by themselves with a d= .20 (K= 19, N= 4,711). This is the largest effect for this set of moderators. Men are also more effective than women when rated by an objective counting device with a d= .15 (K= 2, N= 72). Finally, there are no gender differences when leaders are rated by their peers (d= .02, K= 2, N= 58). However, the effects for both the objective counting device and peers need to be considered with caution as they both have a very small number of studies and sample sizes.

Publication type exhibited a non-significant moderating effect on gender differences in leadership effectiveness (Q_b = 6.48, p = .09). A summary of the hypothesized results is presented in Table 8.

Associations among Moderators

A series of chi-square analyses show that there are 10 significant associations among the moderators. The study setting (laboratory vs. organization) is significantly related to 5 of these moderators. First, study setting and type of effectiveness measure are significantly related, χ^2 (8, N = 99) = 26.50, p < .01. Among the studies conducted in an organizational setting (K= 83), 43% of them measured effectiveness using a measure of leader performance, and 41% measured effectiveness using a measure of leader ability. Among the lab studies (K= 13), measures of effectiveness were fairly evenly distributed with 31% using measured of leader performance, and 23% using measures of leader ability, objective indicators, or mixed indicators, respectively. Of the 5 studies that measured effectiveness using objective criteria, 60% of them were conducted in laboratory settings. Second, the study setting is related to industry, χ^2 (10, N = 99) = 43.15, p < .01. Among the studies conducted in an organizational setting (K= 83), 42% occurred in educational settings, and 29% occurred in business settings, with the remaining studies being evenly distributed across other industries. Among the lab studies (K= 13), 92% of the studies were conducted in mixed industries. Third, study setting is also related to type of rater, χ^2 (12, N = 99) = 32.26, p < .01. Among the studies conducted in an organizational setting (K= 83), 36% used subordinate raters of effectiveness, 30% used a mixed group of raters, and 22% used self-reports, with the remaining studies being evenly distributed across other rater categories. The rater groups were evenly distributed across the lab studies (K= 13).

Fourth, study setting is related to publication type, χ^2 (6, N = 99) = 30.70, p < .01. Among the studies conducted in an organizational setting (K= 83), 63% were published in academic journals, and 30% were found in unpublished theses/ dissertations, with 1

study found in a book, and the remaining 3 in other sources (i.e., whitepapers). Among the lab studies (K= 13), 7 were published in journals, 4 were found in theses/ dissertations, and the remaining 2 were published in books. Finally, study setting is related to level of leadership, χ^2 (6, N = 99) = 20.09, p < .01. The studies conducted in an organizational setting (K= 83), were fairly evenly distributed across levels of leadership with 33% including senior leaders, 29% including lower level supervisors, 24% including leaders from multiple levels, and the remaining 14% including middle level managers. However, among the lab studies (K= 13), 92% included lower level supervisors, while 1 included senior level leaders.

In addition to being related to study setting, type of leadership effectiveness is significantly related to two other moderators. Type of rater is related to type of leadership effectiveness, χ^2 (24, N = 99) = 67.58, p < .01. Among studies conducted using overall leadership performance as the criterion for effectiveness (K= 40), 33% used a group of mixed raters, 25% used subordinate raters, 23% used self ratings, and remaining 8 studies were evenly distributed between boss ratings, peer ratings, and judges/ trained observers. The same pattern occurred for studies measuring effectiveness as leader ability (K= 38). All 6 of the studies that measured effectiveness as satisfaction with leaders used ratings from subordinates. Thus, it is difficult to tease apart the influence of subordinate perceptions from the influence of perceptions of satisfaction.

The industry being examined is significantly related to type of effectiveness measure, χ^2 (20, N = 99) = 44.82, p < .01. Among studies conducted using overall leadership performance as the criterion for effectiveness (K= 40), 40% were conducted in education settings, 33% were conducted in business settings, 20% occurred in multiple

industries, and the remaining 3 studies took place in military settings. A very similar pattern occurred for studies measuring effectiveness as leader ability (K= 38). The other types of effectiveness were spread out fairly evenly across the types of industries. Type of industry is also related to level of leadership, χ^2 (15, N = 99) = 45.93, p < .01.

TABLE 7: Meta-Analysis for Exploratory Moderators

	Mean d	Cor d	K	N	Var.	% Art.	95% CI	80% CR	Q	Q_b
Objective vs. Subjective									409.21**	1.77
Objective	.040	.044	5	195	.032	88.636	-.304, .393	-.183, .272	4.513	
Subjective	-.043	-.046	88	100417	.001	21.802	-.099, .007	-.081, -.011	399.038**	
Mixed	-.192	-.209	6	1064	.017	88.363	-.467, .048	-.378, -.041	5.658	
Type of effectiveness									380.88**	1.518
Subjective- Ability	-.072	-.078	38	81102	.002	26.510	-.160, .005	-.131, -.024	139.57**	
Subjective- Performance	-.045	-.047	40	17688	.002	24.631	-.132, .037	-.103, .008	158.34**	
Subjective- Satisfaction	.039	.038	6	425	.019	55.730	-.235, .310	-.141, .216	8.97	
Objective	.042	.046	5	195	.033	88.64	-.309, .401	-.186, .278	4.51	
Mixed	.005	.005	10	2266	.007	12.953	-.164, .173	-.105, .114	69.48**	
Type X Date										
1962-1995									112.53**	1.057
Subjective- Ability	-.044	-.048	17	5314	.006	47.674	-.198, .102	-.146, .050	33.561*	
Subjective- Performance	.001	.001	19	3292	.005	26.458	-.142, .144	-.092, .095	68.032**	
Subjective- Satisfaction	.023	.022	6	425	.024	55.730	-.282, .325	-.176, .220	8.972	
Objective	.183	.197	3	127	.057	100.000	-.272, .666	-.109, .503	0.170	
Mixed	-.002	-.004	3	856	.029	100.000	-.338, .331	-.222, .215	1.798	
1996-2011									261.01**	1.295
Subjective- Ability	-.095	-.102	21	75788	.003	18.924	-.208, .004	-.171, -.033	105.688**	
Subjective- Performance	-.075	-.081	21	14396	.003	22.192	-.197, .034	-.157, -.006	90.122**	
Objective	-.242	-.259	2	68	.098	45.782	-.873, .355	-.660, .142	2.184	
Mixed	.016	.014	7	1410	.011	9.521	-.192, .220	-.121, .148	63.016**	

TABLE 7 (Continued)

Rater									273.67**	24.455**
Boss	-.159	-.172	9	13273	.005	13.529	-.315, -.028	-.265, -.078	59.133**	
Peers	.015	.017	2	58	.108	91.758	-.629, .662	-.405, .438	1.090	
Subordinates	-.077	-.082	32	63450	.003	70.799	-.185, .021	-.150, -.015	43.786*	
Self	.185	.201	19	4711	.004	16.353	.084, .318	.124, .277	110.074**	
Judges/ trained observers	-.167	-.177	5	882	.015	100.000	-.418, .063	-.334, -.020	0.257	
Objective counting device	.137	.147	2	72	.082	100.000	-.412, .707	-.218, .513	0.067	
Mixed/ unclear	-.109	-.119	30	19229	.002	48.931	-.212, -.025	-.179, -.058	59.267**	
Publication Type									347.39**	6.483
Journal	-.027	-.033	58	77218	.001	23.960	-.101, .035	-.077, .012	237.893**	
Thesis/ Dissertation	.001	.001	30	4980	.003	58.507	-.104, .107	-.068, .070	49.567*	
Book	-.212	-.247	5	838	.017	100.000	-.502, .009	-.414, -.080	3.404	
Other	-.198	-.201	6	18639	.008	8.846	-.371, -.031	-.312, -.090	56.524**	

Note: Mean d is the observed d across studies, cor d is corrected for measurement reliability in effectiveness, K is the number of studies, N is the number of participants, Var. is the variance of the estimate of the corrected d, % Art. is the percentage of variance due to the artifacts of sampling error and measurement unreliability, 95% CI id a 95% confidence interval, 80% CR is an 80% credibility interval, Q is the Chi-square test for homogeneity of effect sizes, Q_b is the between group test of homogeneity. * = p< .05, ** = p<.01.

For the most part, the level of leadership being examined in each study was fairly evenly distributed across industries, except for senior leadership positions. Among studies examining senior leaders (K= 29), 79% occurred in education settings. Thus, it is difficult to know the relative influences of senior leadership positions and that of educational settings on gender differences in leadership effectiveness.

The chi-square analyses also revealed that the type of rater is related to type of publication, χ^2 (18, N = 99) = 35.54, p < .01. The majority of studies measuring effectiveness using ratings from subordinates (K= 32) and self-reports (K= 19), were published in academic journals (72% and 63%, respectively). Other rater types were more evenly distributed among the types of publications. The final significant relationship among the moderators occurred between level of leadership and the year of publication, χ^2 (9, N = 99) = 22.34, p < .01. Most of the levels of leadership being examined in each study were fairly evenly distributed across publication year categories, except for lower level leadership positions. Studies that examined lower level supervisors (K= 37), were primarily conducted between 1962 and 1981 (41%) or between 1982 and 1995 (32%).

Publication Bias

Our analyses revealed that the fail-safe N (Rosenthal, 1979) for the overall relationship was 194. This relatively large number indicates that publication bias was unlikely to be present in the current study. Because this method of detecting publication bias assumes that excluded studies show a null result whereas many may show a result in the opposite direction (Aguinis et al, 2010), I also utilized a funnel plot and trim-and-fill method to test for publication bias. In the funnel plot (see Figure 4), the effect size for each study is plotted as a function of its standard error. Visual inspection of the funnel

plot indicates that the studies included in the present meta-analysis were more or less evenly distributed around the overall effect size; this pattern suggests that no or very little publication bias was present (Egger et al., 1997).

The trim-and-fill procedure (see Duval & Tweedie, 2000) builds on the information obtained in the funnel plot by predicting where potentially missing studies are likely to fall. The overall effect size and 95% confidence interval observed for the combined studies was -0.05 (-0.102, 0.001). Using the trim-and-fill technique, which imputes effect sizes where there may be missing studies, the imputed point estimate was calculated to be -0.003 (-0.056, 0.049). Thus, it appears that this meta-analysis may have a slight publication bias in that it has a greater number of studies with effect sizes favoring females than ones favoring males. However, there could be alternative sources of asymmetry which could be due to true heterogeneity (i.e., substantive factors related to effect size), artifacts (e.g., heterogeneity due to poor choice of effect measure), data irregularities (e.g., poor methodological design of small studies, inadequate analysis, fraud), and chance (Egger et al., 1997; Terrin, et al, 2003).

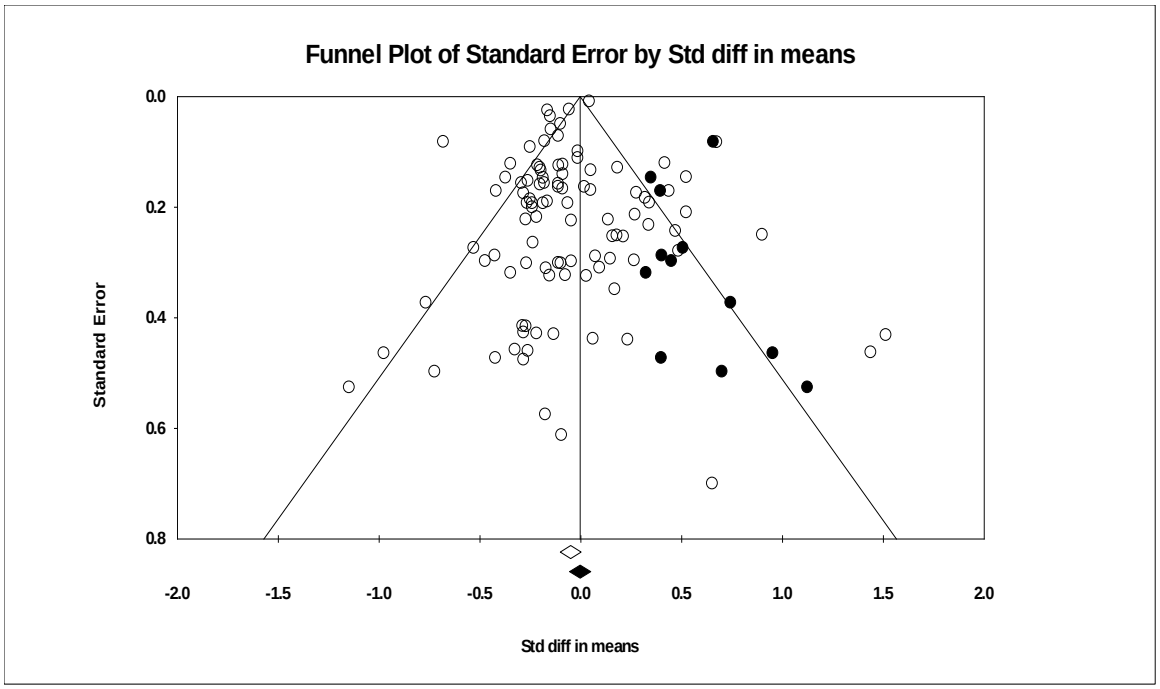

FIGURE 4. Funnel Plot of Standard Error by Effect Sizes

Note. The observed effect size is indicated with the white diamond, while the trim-and-fill imputed effect size is indicated by the black diamond. Potentially missing studies are indicated by the black dots.

TABLE 8: Summary of Abbreviated Hypotheses and Meta-Analytic Findings

Moderator	RCT Hypotheses	Support	Competing Hypotheses	Support
Time of study	There will be greater gender differences (favoring men) seen amongst older studies, and smaller gender differences seen amongst newer studies (Spence & Buckner, 2000 as cited in Eagly & Karau, 2002).	n.s.	None	N/A
Industry	As the masculinity of leadership industry increases, the incongruity between the female gender role and the leadership role increases, leading to reduced perceptions of women's' effectiveness (Eagly & Karau, 2002).	Yes	As the masculinity of the industry increases, women's success in obtaining these roles will create a boomerang effect in which their perceived effectiveness is exaggerated (Feldman, 1981; Weber & Cracker, 1983).	No
Leadership level	At the lowest leadership levels, men are likely to be seen as more effective. At middle leadership levels, women may be seen as more effective than men. At the highest levels, agentic behaviors are needed for success, and thus men will be seen as more effective in these upper positions (e.g. Pavett & Lau, 1983; Martell et al, 1998; Eagly, et al., 2003).	Partial	According to the double standards of competence model, women may be seen as more effective than men at the highest leadership levels (Foschi, 2000).	Yes
Sex of raters	Male raters will favor male leaders in rating their effectiveness. Women raters will rate men and women leaders similarly (Schein, 2001 as cited in Eagly & Karau, 2002).	n.s.	Male leaders will be rated as more effective by male raters. Female leaders will be rated as more effective by female raters. (Turner, 1985; Turner, et al., 1987; Tajfel & Turner, 1979).	n.s.
Nationality	It is unclear how national culture may affect gender differences in leadership effectiveness.			N/A

TABLE 8 (Continued)

Setting	Gender differences in leadership effectiveness will be greater (favoring men) in organizational settings than in laboratory settings (e.g. Macrae, Hewstone, & Griffiths, 1993 as cited in Eagly & Karau, 2002).	n.s.	Gender differences in leadership effectiveness (favoring men) should be smaller in organizational settings than in laboratory settings (Reskin, 2000; see also Eagly et al., 1992, 1995).	n.s.
Percent of male raters	As the percent of male raters increases, women leaders will be seen as less effective. Fewer male raters will lead to differences in effectiveness favoring female leaders (Kanter, 1977a as cited in Eagly & Karau, 2002).	Partial	There will be a curvilinear relationship such that as the percentage of male raters is closer to 50%, gender will be less salient, and gender differences in perceived effectiveness will be small, while at the extremes, men will be seen as more effective (Ridgeway, 2001).	Yes

DISCUSSION

This meta-analysis aimed to offer practical explanations and implications for understanding women's underrepresentation in elite leadership positions, as well as to contribute to the academic debate over gender advantages in leadership, and to better understand the value of RCT and other theoretical perspectives regarding gender and leadership. Specifically, I quantitatively summarized a body of research that examined gender differences in leadership effectiveness between 1962 and 2011, including effect sizes from 99 independent samples of leaders. In response to the debate in the literature regarding whether or not there is a gender advantage in leadership (Vecchio, 2002; Eagly & Carli, 2003a; Vecchio, 2003; Eagly & Carli, 2003b; Yukl, 2002), I found that the overall summary effect was non-significantly different from zero.

Thus, it seems that when all leadership contexts are considered together, differences in leadership effectiveness ratings between males and females are not to blame for women's underrepresentation in leadership positions. This idea is supported by the diverse perspectives and discrepant findings in the literature that propose that there are several contextual moderators that influence whether there is a male or female advantage, or no gender difference in leadership effectiveness. Hypotheses regarding the effects of a variety of contextual moderators developed from RCT (Eagly & Karau, 2002) and other theoretical frameworks proposed in the literature surrounding gender and leadership were examined in this meta-analysis to better understand the circumstances in which gender may make a difference in leadership effectiveness assessments.

Support for Role Congruity Theory Hypotheses

Time of Study

Eagly and Karau (2002) argued that as definitions of leadership become more androgynous and less masculine over time, the tendency to perceive women as less effective leaders than men should also lessen due to an ease in women's role incongruity in relation to leader roles. Despite the sub-group analysis being non-significant, the pattern of sub-group comparisons did show support for this idea—men were considered slightly more effective leaders in older studies, while women were favored in studies conducted more recently.This fits with a recent meta-analysis that examined the extent to which stereotypes of leadership are culturally masculine; the authors found that leadership was seen as more gender-neutral over time (Koenig, et al, 2011).

Additionally, leadership may be defined in a more androgynous way as a result of the increased number of women entering into management positions (Koenig, et al (2011). Prior to 1981, the percentage of women in management positions in the U.S. was less than 26% (Catalyst, 2012b). Studies conducted during this time period tended to show that men were more effective leaders than women. However, after 1981, the percentage of females in management increased to close to 40% by 1995, and slightly above 50% by 2011 (Catalyst, 2012b). During these time periods, I found that gender differences in effectiveness were quite small, and in the most recent group of studies, women were considered to be more effective than men. Yet, it is important to consider the small size of this effect.

When weighted least squares regression was used to supplement the sub-group analyses for examining the effect of time of study, the probability of the effect being different from zero was greater than 5%. While the pattern of the effect (Figure 2) was consistent with the sub-group analysis findings, the effect was not statistically significant.

Thus, it appears that in older studies men were more effective than women, and in newer studies women were more effective than men, but the size of this effect is quite small. As discussed in the results section, it is likely that this effect was not statistically significant due to low power associated with the relatively small sample size (K=99). It is somewhat surprising that the effect is not greater in magnitude given the empirical evidence showing how shifts in the think manager–think male paradigm (Schein, 1973) have also changed over time such that there is now a greater perceived similarity between leaders and women than before (Koenig, et al, 2011). Perhaps it takes longer for outcomes of prejudice (effectiveness ratings) to change than it does for attitudes associated with prejudice to change.

The findings associated with time of study provide hope for future women leaders. It seems that over time the incongruity between women's gender role and leader roles has decreased, leading to small gender differences between men and women's leadership effectiveness. Thus, it is important for scholars to focus on other explanations for women's underrepresentation in leadership positions. While perceptions of women's leadership effectiveness are now equal to or slightly higher than men's, perceptions of their lack of fit in these roles may harm their chances of being hired into leadership positions, being developed or mentored to be better leaders, being paid as much as male leaders, being invited to important networking events, and/or being promoted into more senior level leadership positions. Future research should explore these alternative explanations for women's underrepresentation in leadership positions.

Industry

Also consistent with RCT, the masculinity of the industry being examined significantly moderated gender differences in effectiveness. Industries that are male dominated (i.e., military and government) showed a tendency for males to be more effective, and industries that are female dominated (i.e., education and social service) showed a tendency for females to be more effective. These findings are consistent with the meta-analysis on gender differences in leadership effectiveness conducted on studies published though 1989 (Eagly, et al, 1995). Interestingly, the magnitude of the effects seems to have waned over time. The Eagly, et al study (1995) showed a d=.42 for military studies, yet in the current meta-analysis, the d= .12 for this group. The largest effect for the current meta-analysis was for government organizations with a d of only .27. Also, in the Eagly et al, (1995) meta-analysis, studies conducted in government and social service industries were combined into one category which had a d= -.15. In the current study, these were considered separate groups due to the different percentages of women who make up each industry (37.3% for government and 85% for social service). Different effects were seen for each group, with men seen as more effective in government positions and women being seen as more effective in social service positions.

In the business industry (43% female)—with a close to equal representation of men and women—women were seen as slightly more effective than men. Despite this close to equal distribution of men and women in these roles, perhaps women are seen as more congruent with leadership roles in business than men. Future research should examine this possibility. Also, it is important to note that 68% of the studies that examined leaders in business settings were conducted after 1995. This is a higher association than any of the other industries included in this meta-analysis. Thus, it may be

that the weakening of stereotypes over time associating men with leadership played a role in gender differences in leadership effectiveness in the business industry.

In addition, it is important to consider the different effects of prejudice based on role incongruity, and actual differences in leadership effectiveness, on assessments of gender differences in leadership effectiveness. RCT explains that prejudice is to blame for perceptions that either men or women are more effective within a specific industry. However, it is also possible that there are actual behavioral differences between men and women leaders in these settings, leading to real differences in their effectiveness at leading others. When leader roles are particularly masculine (feminine), people may suspect that women (men) are not qualified for them, and they may resist women's (men's) authority (Eagly & Karau, 2002; Heilman, 2001). This resistance to authority may mean that leaders who are in gender-incongruent roles actually behave in a less effective manner.

When individuals have high status characteristics in a situation, they are likely to receive more opportunities to make contributions to the group task, receive higher evaluations for their contributions, and exert greater influence over the behaviors of others (Berger et al, 1977). These performance expectations can be self-fulfilling because those who are expected to perform will be given more opportunities to participate, and are therefore likely to make more contributions than others. When group members have unequal status, those with high status who are talkative and assertive are likely to be considered competent by others, and are thus able to influence the group, however, those with low status who are talkative and assertive are likely to be perceived as incompetent, and thus are typically ignored or rejected (Meeker & Weitzel-O'Neill, 1985).

Unfortunately, the current meta-analytic data are not informative about the relative influence of the masculinity of the industry on perceptions of effectiveness versus actual effective leadership behaviors displayed by the leader. If more primary studies had included objective indicators of effectiveness, it may have been possible to examine the effects of industry on subjective versus objective measures. Yet, it is important to note that another meta-analysis supports the conclusion that prejudiced perceptions may play a greater role than actual leadership behaviors on effectiveness ratings. A meta-analysis of lab studies that experimentally equalized leadership behaviors and performance of male and female leaders found that men were rated more favorably than women in male-dominated leadership roles (Eagly, et al, 1992). Future research should explore the influence of prejudicial attitudes versus actual differences in leader behavior on gender differences in leadership effectiveness.

The results for the effects of industry on gender differences in leadership effectiveness provide little support for the boomerang effect offered by Feldman (1981) and Weber and Crocker (1983). A considerable amount of research has been conducted on this effect—showing that when observed behavior contrasts sharply with a stereotypically derived expectation, the incongruence between the observed behavior and the expected behavior can cause a reversal of the traditional stereotype (i.e., Cantor & Wilson, 2003; Bailey & Vietor, 1996). However, the majority of these studies seem to use highly controlled lab settings in which very specific behaviors are manipulated via vignettes or videos. To my knowledge, the boomerang effect has not been tested in field settings or in any studies that involve participants assessing someone they interact with on a regular basis (as would be the case in most of the studies examined in the current

meta-analysis). Thus, it seems that the boomerang effect may only take place in very specific instances where an actor's behavior is obviously incongruent with an expected behavior. A woman simply making it into a highly masculine leadership position may not be enough of a cue to counteract the stereotypes associated with the incongruity between her female gender role and the leader role.

Nationality

RCT proposed an exploratory research question regarding how the nationality of leaders may moderate gender differences in their effectiveness. Results show that the nationality of leaders and the gender egalitarianism present within geographical regions did not have a major influence on the effect sizes. All of the effects showed women as being slightly more effective than men, regardless of nationality or gender egalitarianism. These findings support the idea that even though leadership styles and gender beliefs may vary among national contexts, organizational cultures may play a larger role in leadership effectiveness assessments. Even though research shows that in organizations located in countries with low GE, there tends to be more gender discrimination leading to fewer women in leadership positions (Emrich, Denmark, & Den Hartog, 2004), it seems that national GE may not negatively influence perceptions of women who are able to make it into leadership positions.

As proposed by Pothukuchi et al., (2002), it may be that organizational cultural dimensions are more proximal than national level variables, and may have a greater effect on employee perceptions. Thus, it is important for human resource practitioners to monitor the degree to which their organizational cultures and practices promote equality and fairness for all employees, especially if their organizations are located in national

contexts that promote gender inequality. However, it is important to note that there were only 12 studies conducted outside of the U.S. and Canada, thus more research is needed to better assess how nationality may moderate gender differences in leadership effectiveness.

Support for Competing Hypotheses

Level of Leadership

The current meta-analytic findings were consistent with 3 of the hypotheses that compete with RCT, and partially consistent with 3 hypotheses developed from RCT. First, there was partial support for the hypothesis proposed by RCT regarding the influence of the level of leadership on gender differences in leadership effectiveness. Consistent with the hypothesis proposed by Eagly and Karau (2002), men were seen as more effective in lower level positions, and women were seen as more effective in middle management positions. Counter to their hypothesis that men would be seen as more effective than women in the highest leadership levels, the results showed that women were seen as slightly more effective than men at the top. This finding is consistent with the competing hypothesis supported by Foschi (2000) who argues that a woman's presence in a top leadership position provides information about her abilities to others in the organization—that she must be exceptionally competent to have made it in such a high status and agentic leadership position.

However, Eagly and Karau (2002) proposed the opposite, that men would be more effective at the highest levels. They based this idea on a body of research that has shown the higher the level of leadership is, the more masculine and agentic are the expected behaviors for the leader (Hunt, Boal, & Sorenson, 1990; Lord & Maher, 1993;

Martell, Parkers, Emrich, & Crawford, 1998). Yet, other research has shown that leaders in these roles describe the behaviors needed for success as requiring a greater range of task and relational skills (e.g., Alexander, 1979; Gomez-Mejia et al., 1985; Paolillo, 1981; Pavett & Lau, 1983). It seems that senior leaders may need to enact both agentic and communal behaviors in order to successfully complete the complex variety of tasks in such elite leadership levels. Results for the current meta-analysis support this idea, showing that the gender difference in leadership effectiveness in the highest levels of leadership is very small, slightly favoring women. Future research is needed to clarify the behaviors needed to be successful in senior leadership positions in order to better understand how the masculinity or femininity of those behaviors may affect men and women's' congruity in such roles. Additionally, given the small size of this effect, it is difficult to conclude whether women are truly seen as more effective than men (perhaps due to perceptions of women's exceptional competence) or whether there are no real gender differences at these levels.

These findings fit with the leadership labyrinth paradigm (Eagly & Carli, 2007) which explains that women may be under-represented in top leadership positions due to barriers and obstacles all throughout their careers, rather than just at the top (at the glass ceiling). The current study shows that women may be seen as less effective than men in lower-level supervisor positions but not at middle level or higher level leadership positions. Thus, the masculine behaviors associated with lower level management positions, and the prejudice they may produce against female leaders, may serve as one of those lower level barriers women face on their way through the leadership labyrinth. Future research should expand upon these findings to determine other barriers to

women's progress that may exist at lower level management positions within organizations. It is also important for human resource managers to be aware of the potential for prejudice against women to be stronger at lower management levels. Training and development programs could be developed to help reduce incongruity biases against women in these positions, as well as to aid female supervisors in adopting the most effective behaviors for lower level leaders to enact.

Sex of Raters

The competing hypothesis regarding the effect of sex of raters on gender differences in leadership effectiveness was generally supported by the pattern of effects, despite the overall effect being non-significant. Consistent with the hypothesis developed from self categorization theory (Turner, 1985; Turner, et al., 1987) and social identity theory (Tajfel & Turner, 1979), raters generally perceived same-sex leaders as more effective than opposite sex leaders. RCT proposes that because women have a more androgynous definition of leadership than men (Schein, 2001), that they would be less likely than men to rely on gender-leader congruency biases, and would be more likely to see male and female leaders similarly. Yet, I found more support for the argument that people strongly identify with members of their own gender social group, and this in-group favoritism and bias towards same-sex others seems to play a role in leadership effectiveness ratings.

This finding has important implications for human resource practices. If a leader's performance appraisal includes perceptions of their leadership effectiveness by only members of one sex category, this could potentially bias the appraisal that the leader receives. To reduce the influence of in-group favoritism and out-group prejudice based

on sex similarity and difference, human resource managers may want to encourage performance management assessments to come from multiple evaluators from both sex categories. However, it is important to consider that researchers have not consistently found support for in-group favoritism based on demographic similarity (Shore, Cleveland, & Goldberg, 2003; Tsui & O'Reilly, 1989; Tsui, Porter, and Egan, 2002). This in-group favoritism effect may only for both sexes in settings where the gender of the leader is not strongly related to higher performance expectations.

Goldberg, Riordan, and Zhang (2008) recently proposed that men may enhance their social identities by favoring in-group members, because this demographic category generally has high status in organizations. On the other hand, there may be some settings in which women are low status individuals, and in these settings, in-group favoritism may offer little status benefits to females. In these cases, the low status employee may attempt to dissociate from other members of their sex category by showing favoritism towards leaders who are members of the higher status demographic group, despite not being a member of that group (Ellemers,Van Knippenberg, DeVries, & Wilke, 1988; Ellemers,Wilke, & Van Knippenberg, 1993). I attempted to test this idea in the current meta-analysis by examining the effects of sex of rater on gender differences in leadership effectiveness in different industries and levels of leadership, however there were not enough primary studies in each category to conduct hierarchical sub-group analyses.

However, there were enough studies across different time periods for me to conduct such analyses. Time period should act in a similar manner as industry or level of leadership such that in older studies, men had higher status as leaders than women, and in newer studies this is not the case. Thus, if men have higher status, they should be seen as

more effective by both male and female raters, whereas if women have close to the same status as men do, in-group favoritism should occur for both male and female raters. Indeed, I found that in older studies (published before 1981), men were favored by both female raters (d= .07, K=3), and by male raters (d= .06, K=3), whereas in studies published after 1981, women rated women leaders as more effective than male leaders (d= -.15, K=8), and the opposite occurred for male raters (d= .04, K=8).

Thus, in leadership settings where women or men have a status advantage, it is likely that in-group and out-group favoritism will occur, leading to an effectiveness advantage for members of the higher status group. If the leadership setting does not provide a status advantage for either sex, it is likely that there will be in-group favoritism observed for each sex respectively (as was seen in this meta-analysis). Regardless, such biases are likely to be reduced if performance management systems include evaluations of leadership effectiveness from both men and women raters, as well members of other diverse demographic groups.

Percent of Male Raters

Finally, the competing hypothesis regarding the effect of the percent of male raters on gender differences in leadership effectiveness was also supported. Consistent with expectation states theory (Ridgeway, 2001), the percentage of male raters was curvilinearly related to gender differences in leadership effectiveness such that when there are an extreme number of male or female raters, men are seen as more effective leaders than women. This is partially consistent with RCT that proposed that when there was a high percent of male raters, male leaders would be seen as more effective than female leaders. However, counter to the suggestion by Eagly and Karau (2002) developed

from tokenism (Kanter, 1977a), when there were a majority of female raters present among the group of raters, male leaders were still seen as more effective than women. Additionally, when there were close to equal numbers of men and women raters in the group, there were negligible differences in assessments of men and women's leadership effectiveness. These findings support expectation states theory's (Ridgeway, 2001) proposition that when gender becomes salient to the leadership task (through a majority of team members being male or female), performance expectations will increase for men, leading to men being seen as more effective leaders than women.

Organizations should be made aware of the effects of numerical underrepresentation on men and women's leadership effectiveness ratings. If possible, leadership teams should be created to be gender balanced in order to reduce the salience of gender to the task, and to reduce the effect of gender-based expectation states on leadership effectiveness assessments. Further, research has shown that gender diversity facilitates team performance in male-oriented tasks without negatively impacting team cohesion or decision making time (Rogelberg & Rummery, 1996). Indeed, several studies have found positive effects of gender diversity on team performance, showing that balanced cross-gender teams may be more advantageous than all male or all female teams (Martin, 1985; Orlitzky & Benjamin, 2003; Smith-Lovin, Skvoretz, & Hudson, 1986). Thus, not only may gender diversity in leadership teams reduce the influence of gendered performance expectations favoring male leaders, but it may also positively enhance team functioning and performance.

Exploratory Moderators

Effectiveness Criteria

It is practically important to understand how the type of leadership effectiveness criteria used to assess leaders may influence gender differences. Consistent with the summary effect for effectiveness ratings overall, results of this meta-analysis show that the effects for objective versus subjective measures were also quite small—with females scoring slightly higher on average for subjective assessments, and males scoring slightly higher on average for objective assessments. This finding goes against literature that has encouraged the increased use of specific and objective job performance criteria, in order to reduce bias against minority groups (Frear & Paustian-Underdahl, 2011; Eagly, 2007). These authors propose that subjective behavioral rating scales used in performance management assessments allow for the subconscious biases of raters to negatively influence ratings of female leaders. Yet, in this meta-analysis, there was a slight tendency for men to be seen as more effective than women when objective, rather than subjective criteria were used to assess their performance.

This effect may be explained in part by the time-period of the studies. Among the 5 studies that used objective indicators of effectiveness, 3 of them were conducted prior to 1995, when men were likely to be seen as more effective than women. In the remaining 2 studies conducted after 1995, women were seen as more effective. Women's status within society has improved within the past several decades (Twenge, 2001), likely leading to increased perceptions of women's competence and legitimacy. The majority of studies examining objective leadership effectiveness measured such effectiveness via the groups' productivity or output (i.e., Bartol, 1973; Jacobson & Effertz, 1974; Schneier & Bartol, 1980). Thus, followers would need to accept the leader's authority and direction in order for the leader to be able truly affect the behaviors of others in order to reach unit-

level objective effectiveness. It seems that this is more likely to occur in time periods with greater equality between men and women.

These findings suggest that the extent to which a gender bias exists for either type of effectiveness criteria may depend upon aspects of the leadership context. In order to be an effective leader objectively or subjectively, the leader has to be perceived as competent and legitimate in the eyes of followers. If the time of the study, the industry, or the task at hand leads to a male or female status advantage, then it is likely that this advantage will manifest itself in gender differences in leadership effectiveness—regardless of whether those measures are subjective or objective. Human resource managers should develop training programs and socialization activities for their leaders, in order to ensure that male and female leaders have equal status and legitimacy within their organizations.

Rater Type

The type of rater affected gender differences in leadership effectiveness such that women were seen as more effective than men when rated by bosses, subordinates, judges/ trained observers, and by a mixed group of raters, while men were seen as more effective when rated by themselves or by an objective counting device. There were no gender differences when leaders were rated by their peers. These findings contrast somewhat with findings from the 1995 meta-analysis examining gender differences in leadership effectiveness (Eagly et al, 1995). In the previous study, not only were men seen as more effective than women by themselves (as in this study), but also by their bosses, peers, and mixed groups of raters. Women were seen as more effective than men by their subordinates, judges, and objective counting devices (Eagly et al, 1995). Thus, it appears

that women are seen as more effective than men by a more diverse variety of raters now than in the past, and men are seen as more effective by fewer rater groups now than before.

The largest effect size found for the type of rater moderator was for self-reported effectiveness perceptions; men rated themselves as significantly more effective than women rated themselves (d= .27). This finding is consistent with literature examining how cultural beliefs about gender and task competence (Ridgeway 1997; West and Zimmerman, 1987) may bias performance feedback individuals use in assessing their own competence. Correll (2001) examined men and women's perceptions of their mathematics ability, an academic area that is generally considered to be male-typed. Correll (2001) reasoned that females who receive positive feedback about their mathematical ability should be less likely to perceive that they are skilled at mathematics since this perception is incongruent with widely shared beliefs about gender and mathematics. Consistent with this idea, Correll (2001) found that controlling for performance feedback about mathematical ability, males' assessments of their own mathematical competence was higher than females' assessments. This same effect was observed in the present study examining leader's self evaluations of their effectiveness.

This finding has important implications for performance management practices. If self reports are used as the primary measure of a leader's effectiveness within an organization, it is likely that men may receive higher performance evaluations than women, due to their tendency to assess themselves as more competent than women do. It is important to note that this may still occur even when women receive very positive performance feedback from others (Correll, 2001). Thus, it is important for performance

management systems to include ratings from a variety of sources, and for human resource managers to be aware of the possibility for male leaders to evaluate themselves considerably higher than female leaders, regardless of the kind of feedback they may receive from others in the workplace.

Publication Type

Publication type also exhibited a non-significant moderating effect on gender differences in leadership effectiveness. Yet, the effect sizes did change somewhat based on the publication type with studies written in books and 'other' sources such as whitepapers showing women as being more effective than men. Studies written as published journal articles also favored women, but to a smaller extent, while studies written as unpublished theses or dissertations showed no gender differences. Consistent with Rosenthal's (1979) discussion of the "file drawer" problem, the unpublished theses or dissertations included in this meta-analysis had very small effect sizes. Interestingly, non-published white papers and studies in published books had quite large effects favoring women. There may be pressures by the popular press and media for consulting companies and/ or authors of books to report findings which reveal a 'female advantage' in leadership (e.g. Helgesen, 1990; Rosener, 1995). These pressures may explain in part why these non-peer reviewed sources of research have large effects favoring women leaders. Yet, it is important to note that despite the size of the effects being significantly larger in non peer-reviewed sources, the direction of effects (favoring females) in these sources remains consistent with findings found in more scientifically rigorous, peer-reviewed journal publications.

Summary of Contributions and Implications

Theory

A primary goal of this meta-analysis was to examine the extent to which hypotheses proposed by RCT were supported compared to some competing theoretical perspectives proposed in the literature on gender and leadership. Overall, the findings were consistent with 2 RCT hypotheses, and were partially consistent with 3 additional hypotheses proposed by RCT. I also found support for competing hypotheses proposed by the double standards of competence model (Foschi, 2000), self categorization theory (Turner, 1985; Turner et al., 1987), social identity theory (Tajfel & Turner, 1979), and expectations states theory (Ridgeway, 2001). Taken all together, it appears that the primary tenets of RCT are well supported with the data in this meta-analysis—that prejudice against female leaders can vary depending on a variety of features of the leadership context and characteristics of the perceiver(s) of the leader's effectiveness. However, based on the partial support found for 3 of the RCT hypotheses and the support provided for some of the competing hypotheses, I believe there are certain aspects of RCT that may need to be further clarified and explored.

I propose that instead of competing with RCT, each of the theories that were supported in this study align well with the core tenets of RCT, and can serve to inform the development of RCT. For instance, the RCT hypotheses regarding lower and middle level leadership positions were supported, yet at the highest levels, the direction of effects contradicted that proposed by RCT. The direction proposed by the double standards of competence model (Foschi, 2000) was supported instead—that women were seen as slightly more effective than men at the highest levels. It makes sense based on RCT that

the effects would slightly favor females if there is a slightly greater demand for stereotypically feminine characteristics in the highest leadership positions. Additionally, if women in these positions are seen as being exceptionally competent by having overcome significant obstacles in order to reach these senior positions, there would be an increased congruency between their gender role and leader roles. Yet, the effect size was quite small and could be considered to show a negligible gender difference. This finding still fits with what is core to RCT—if the characteristics needed for success at the highest levels are both masculine and feminine, then the incongruity between the female gender role and the leader role may not be very large, leading to small gender differences in effectiveness.

The findings for sex of raters can also be explained in ways that fit with RCT. I found support for the arguments proposed by self categorization theory (Turner, 1985; Turner, et al., 1987) and social identity theory (Tajfel & Turner, 1979) that people strongly identify with members of their own gender social group, and this in-group favoritism and bias towards same-sex others seems to play a role in leadership effectiveness ratings. These ideas are in line with RCT, which proposes that because women have a more androgynous definition of leadership than men (Schein, 2001), they would be less likely than men to rely on gender-leader congruency biases. It is likely that men have a more masculine view of leadership, which leads to them perceiving men to be more effective than women due to a greater congruence between the male gender role and leader roles, and to in-group favoritism. However, because women have a more androgynous view of leadership, they are less inclined to perceive incongruence between

the female gender role and leader roles, and instead they have an in-group favoritism which leads to perceptions that women are more effective leaders than men.

The curvilinear effect of the percent of male raters on gender differences in leadership effectiveness also fits with RCT. The findings in this study support expectation states theory's (Ridgeway, 2001) proposition that when gender becomes salient to the leadership task (through a majority of team members being male or female), performance expectations will increase for men, leading to men being seen as more effective leaders than women. When performance expectations increase for men, this would lead to a greater congruence between the male gender role and the leader role, and an incongruence for females in leader roles, leading to male leaders being favored in such settings. Thus, based on the findings of this meta-analysis, RCT can be updated to explain how the theoretical perspectives provided by the double standards of competence model (Foschi, 2000), self categorization theory (Turner, 1985; Turner, et al., 1987), social identity theory (Tajfel & Turner, 1979), and expectations states theory (Ridgeway, 2001) may affect the influence of moderators like leadership level, sex of raters, and percent of male raters in a different way than what was originally proposed by Eagly and Karau (2002).

There was only one hypothesis based on RCT that did not receive any support in this study—the effect of study setting on gender differences in leadership effectiveness. Interestingly, the competing hypothesis based on individuating information (Reskin, 2000) was not supported either. Both of these hypotheses were based on the idea that overall gender differences in leader effectiveness would favor men instead of women, and that this effect would be moderated by the study setting. However, this meta-analysis

found the opposite effect, such that women were seen as more effective than men in both organizational and laboratory studies. It seems that the time of the study, and the sex-type of the task being conducted in the study, may serve as more proximal variables (than the setting) influencing gender's effect on leadership effectiveness. Thus, this moderator proposed by RCT may need to be reconsidered as the theory is developed and utilized in future research.

Practice

This meta-analysis also has important implications for human resource management and performance management practices. I found that women and men are fairly equally effective in leading and managing (with women being seen as slightly more effective). This finding has important implications for organizations and employees because discriminatory barriers limiting women's access to leadership roles are likely to reduce an organization's productivity by removing a substantial proportion of its managerial talent. This study helped to identify some of the characteristics that moderate tendencies for men or women to be seen as more effective, which should facilitate progress in understanding the social and psychological processes that may account for gender differences in perceptions of leadership effectiveness.

It is important for human resource managers to be aware of the potential for perceptions of incongruity between leaders' gender roles and leader roles to lead to prejudicial perceptions of gender differences in effectiveness. These effects seem to favor women in business, education, and social service organizations, while they favor men in government and military settings. Additionally, women are likely to be seen as more effective in middle management positions, while men are seen as more effective in lower

level supervisor positions. Steps can be taken to reduce biases and prejudice in the performance management process within organizations.

In this study, the sex of the rater influenced effectiveness ratings such that men tend to see male leaders as more effective than women, while women raters tend to assess women leaders as more effective than male leaders. To reduce the influence of in-group favoritism based on sex similarity, human resource managers may want to encourage performance management assessments to come from multiple evaluators from both sex categories. The percent of males amongst the group of raters also influenced gender differences in leadership effectiveness ratings such that when there was a majority of females or males in the group, men were seen as more effective leaders than women. Leadership teams should be created to be moderately gender-balanced in order to reduce the salience of gender to the task, and to reduce the effect of gender-based expectation states on leadership effectiveness assessments that favor men. Finally, this study found that the type of rater evaluating the leader's effectiveness plays a role in gender differences such that self ratings strongly favored male leaders over female leaders. Thus, it is important for performance management systems to include ratings from a variety of sources in order to reduce the strong self-assessment biases that exist for male leaders.

Limitations and Directions for Future Research

Like all research efforts, this study has some limitations. First, although the meta-analytic sample size was fairly large with K = 99 studies, the use of moderator analyses led to relatively small numbers for some analyses (e.g., social service settings, Australian studies). These small numbers limit the generalizability of the findings of these sub-group comparisons. In addition, sample size constraints did not allow for many hierarchical

examinations of multiple subgroup moderators. Although the results highlight a variety of substantive moderators of gender differences in leadership effectiveness, considerable variability in effect sizes typically remained when examining each moderator separately. Nevertheless, to the extent that moderators are correlated with one another (e.g., if studies examining senior leaders are primarily conducted in education settings), observed moderation effects may be confounded with one another. Yet, the approach taken here of examining sub-group moderators separately is consistent with standard practices when conducting meta-analysis (Hunter and Schmidt, 2004).

The potential confounding of sub-group moderators is another limitation of this meta-analysis. In examining the relationships among moderators in this study, I found 10 significant correlations. Study setting was included in half of these relationships, likely due to the large number of studies being conducted in organizations (N= 82) rather than lab settings (N=13). It should be noted that among these relationships, the largest example of overlapping between the setting and another moderator, was that 63% of the studies conducted in organizations were published in academic journals. The next largest overlap for organizational setting was only 43% (overlap with studies using leader performance as measure of effectiveness). While these are examples of potential confounding, they are not practically significant given the relatively small amounts of overlap. Among the lab studies, however, the examples of overlap were larger with 92% of these studies being conducted with leaders from mixed industries and 92% being conducted with lower level supervisors. Based on these large levels of overlap, it is difficult to disentangle the effects of the lab setting from those of lower level leader

positions, and mixed industries. Yet, these examples of confounding do not significantly affect any of the primary conclusions drawn from this study.

Among the additional 5 significant relationships among moderators, there were two examples of confounding that may influence how the primary findings of this study are interpreted. The first involved the overlap between studies examining senior leaders and those conducted in education settings (K= 29), 79%. Thus, it is difficult to know the relative influences of incongruity between the gender role and leader role in senior leadership positions versus that of educational settings on gender differences in leadership effectiveness. Future research should examine the extent to which the female gender role and the leader role align in educational settings with leaders of different levels in order to tease apart the influences of these two moderators. The second important example of confounding in this study involved the large amount of studies that examined lower level supervisors, between 1962 and 1995 (73%). Thus, it is difficult to know whether the time period was the most important factor in men being seen as more effective than women, or if being in lower level positions was the more important moderator. Future research is needed to examine the feminine and masculine behaviors needed for success at different leadership levels in order to better understand how level moderates gender differences in leadership effectiveness.

Another limitation of this meta-analysis is that it does not examine how leader behaviors play a role in the relationship between gender and leadership effectiveness. As I discussed in the introduction section of this dissertation, research has shown that there are gender differences in leadership behaviors. Indeed, a meta-analytic review of gender and leadership styles found that women are more likely than men to engage in

transformational leadership behaviors, while men are more likely to engage in transactional and laissez-faire leadership (Eagly, et al., 2003). Transformational leadership is increasingly associated with leadership effectiveness moreso than other leadership styles (Eagly & Johannesen-Schmidt, 2001; Eagly & Johnson, 1990; Eagly et al., 2003). This body of research may lead one to hypothesize that gender may be related to leadership effectiveness through the effects of leadership behaviors, such that women are more likely to adopt transformational leadership behaviors, and thus they may be seen as more effective than men.

Yet, other research supports a different model—that leader behaviors and the leader's gender interact to affect leadership effectiveness. Gender stereotypes play a very important role in how leader behaviors are perceived in the workplace. Even when male and female leaders behave similarly, they are not necessarily evaluated similarly (Jago & Vroom, 1982; Nieva & Gutek, 1982; Powell, 1988). Research has shown that both women and men are evaluated less favorably when they behave in ways that are incongruent with their gender roles than when they behave in a gender-congruent manner (Haccoun, et al., 1978; Petty & Lee, 1975; Schein, 1973; Watson, 1988; Wiley & Eskilson, 1982). When women enact a transformational leadership style, they are likely to be seen as more effective than men because of the perceived alignment of this style with female characteristics and the requirements of leader roles. Transactional and laissez-faire leadership behaviors may be seen as more congruent with masculine characteristics, and thus there is likely to be a more positive relationship between these leader behaviors and effectiveness for men than women. This literature may lead one to

hypothesize that gender will moderate the relationship between leader behaviors and leadership effectiveness.

However, these competing models could not be tested in the current study. Including studies that examined gender differences in leadership effectiveness and leadership behaviors was beyond the scope of the current dissertation which was aimed at examining the effects of contextual moderators proposed by RCT. Yet, future research should examine the competing models discussed above to determine how gender and leader behaviors influence leadership effectiveness. On a related note, the current meta-analysis was unable to examine why the leadership industry and level of leadership moderated gender differences in leadership effectiveness. This could be explained by the enactment of different leadership behaviors seen as necessary in that industry or at that leadership level, and perceptions of the congruity of those behaviors with the leader's gender. For example, if some settings require transformational leadership behaviors (e.g. middle management positions, or education settings) and leaders adopt such behaviors, women may be seen as more effective than men, due to the congruity between their behaviors and their gender role. However, when leaders are in settings that require transactional or laissez-faire leader behaviors (e.g. lower level supervisor positions, or the military) and they adopt such behaviors, men may be considered to be more effective than women due to the congruity between these behaviors and their gender role. These are additional avenues for future research to explore.

Another limitation of this meta-analysis is that perceptions of the incongruity between certain industries and levels of leadership, and gender roles were not measured. Instead, I used statistics on the percent of females in different industries as well as

research examining the masculine and feminine behaviors believed to be necessary at different leadership levels, as proxies for the congruity between the female gender role and leader roles. While Heilman (1983) suggests that one way to conceptualize the masculinity or femininity of a job is by the percentage of men and women occupying that job, and Eagly and Karau (2002) argue that congruity can result from extent to which feminine and masculine behaviors are needed in a leadership role, it would have been useful to receive ratings from a general population sample on the perceived congruity between the female gender role and leader roles in various leader roles. This would have provided an additional source of evidence regarding the extent to which there is incongruity between the female gender role and leader roles across different contextual moderators.

Eagly and Johnson (1990) completed a questionnaire survey which was reported by Eagly et al, (1995), in which respondents rated a wide range of leadership roles on a number of gender-relevant dimensions. This survey included several measures to assess the degree to which the leadership roles were perceived as congenial for women and men. Sex differences were calculated for respondents' self-reported competence to perform each role and their interest in performing each role. Additionally, respondents' reported their beliefs about differences in average men's and women's interest in performing each role. Other measures in the survey assessed respondents' judgments of the extent to which each role requires interpersonal abilities and task-oriented abilities. This kind of survey study should be replicated today in order to gauge how perceptions of congruity may have shifted over time, and to more thoroughly assess the affect of incongruity on gender differences in leadership effectiveness.

This meta-analysis included a variety of moderators based on those presented in Eagly and Karau's (2002) review of RCT. However, a potential limitation is that some moderators proposed in the 2002 review paper were not included in the current analysis because they were not reported in enough studies examining gender differences in leadership effectiveness. These excluded moderators include leader race, the dress, grooming, and overall physical appearance of leaders, leadership style, and perceivers' personal endorsement of traditional gender roles. Given the conceptual importance of these issues, future research that reports the influence of one or more of these variables on gender differences in leadership effectiveness would be valuable.

Finally, many of the effect sizes reported in this meta-analysis are fairly small. While these small effects are consistent with findings of other meta-analyses examining gender differences in mathematics performance (Lindberg, Hyde, Petersen, & Linn, 2010), job performance (Roth, Purvis, & Bobko, 2012), leadership effectiveness (Eagly et al, 1995), and leadership evaluations (Eagly et al, 1992), some critics have suggested that such small effects are unimportant (Vecchio, 2002). However, many researchers have disagreed with this perspective, arguing that even quite small effects can have practical importance in real-life settings (see Abelson, 1985; Bushman & Anderson, 2001). For example, Eagly and Carli (2003b, pg 825) provide an important example regarding a study examining the effects of aspirin, "the relation between taking aspirin and the prevention of heart attacks in a randomized double-blind experiment was only r =.034, yet this effect corresponded to 3.4% fewer people experiencing heart attacks, a drop meaningful enough to induce researchers to end the experiment prematurely because it was deemed unethical to deny the benefits of the treatment to the individuals in the

control group (Rosnow & Rosenthal, 1989)". Similarly, small biases against women in assessments of leadership effectiveness in certain contexts, when repeated over individuals and occasions, can produce large consequences in terms of women's ability to reach and succeed in the most elite leadership positions (Martell, 1999; Martell, Lane, & Emrich, 1996; Eagly & Carli, 2003b).

Conclusion

This meta-analysis summarized a body of literature examining gender differences in leadership effectiveness, and found that when 99 studies from different time periods and different settings were combined, there was a slight tendency for women to be seen as more effective than men. Additionally, I examined the influence of contextual moderators developed from role congruity theory (Eagly & Karau, 2002) and some competing theoretical frameworks. Overall, I found that the core tenets of role congruity theory are supported; however the theory can be updated and expanded by taking into account findings supported by other theories presented in the literature on gender and leadership. Finally, this meta-analysis provides important practical implications for reducing the barriers women may face throughout the leadership labyrinth as they pursue the most elite leadership positions.

REFERENCES

Abelson, R. P. (1985). A variance explanation paradox: When a little is a lot. Psychological Bulletin, 97, 129–133.

Aguinis, H. (2001). Estimation of sampling variance of correlations in meta-analysis. Personnel Psychology, 54, 569–590.

Aguinis, H., & Pierce, C. A. (1998). Testing moderator variable hypotheses meta-analytically. Journal of Management, 24, 577-592.

Aguinis, H., & Whitehead, R. (1997). Sampling variance in the correlation coefficient under indirect range restriction: Implications for validity generalization. Journal of Applied Psychology, 82, 528–538.

Aguinis, H., Gottfredson, R. K., & Wright, T. A. (2010). Best-practice recommendations for estimating interaction effects using meta-analysis. Journal of Organizational Behavior, 31, 776- 786.

Aguinis, H., Dalton, D. D., Bosco, F. A., Pierce, C. A., & Dalton, C. M. (2011). Meta-analytic choices and judgment calls: Implications for theory building and testing, obtained effect sizes, and scholarly impact. Journal of Management, 37: 5-38.

Alexander, L. D. (1979). The effect level in the hierarchy and functional area have on the extent Mintzberg's roles are required by managerial jobs. Academy of Management Proceedings, 39, 186–189.

Allport, G. W. (1954). The nature of prejudice. Reading, MA: Addison-Wesley.

Arthur, W., Bennett, W., & Huffcutt, A. I. (2001). Conducting meta- analysis using SAS. Mahwah, NJ: Erlbaum.

Ashmore, R. D. (1981). Sex stereotypes and implicit personality theory. In D. L. Hamilton (Ed.), Cognitive processes in stereotyping and intergroup behavior. Hillsdale, NJ: Erlbaum.

Avolio, B. J. (1999). Full leadership development: Building the vital forces in organizations. Thousands Oaks, CA: Sage.

Avolio, B. J., Sosik, J. J., Jung, D. I., & Berson, Y. (2003). Leadership models, methods, and applications: Small steps and giant leaps. In W. C. Borman, R. Klimoski, D. R. Ilgen & B. Weiner (Eds.), Handbook of psychology, vol. 12, 277-307. New York: John Wiley & Sons.

Avolio, B. J., Bass, B. M., & Jung, D. I. (1999). Re-examining the components of

transformational and transactional leadership using the Multifactor Leadership Questionnaire. Journal of Occupational and Organizational Psychology, 72, 441-462.

Bailey, R. C., & Vietor, N. A. (1996). A religious female who engages in casual sex: Evidence of a boomerang effect. Social Behavior and Personality, 24, 215-220.

Banks, C. G., & Murphy, K. R. (1985). Toward narrowing the research-practice gap in performance appraisal. Personnel Psychology, 38, 335-345.

Bass, B. M. (1990). Bass and Stogdill's handbook of leadership (3 ed.). New York: Free Press.

Becker, B. J. (2005). Failsafe N or file-drawer number. In H. R. Rothstein, A. J. Sutton, & M. Borenstein (Eds.), Publication bias in meta-analysis: Prevention, assessment and adjustments (pp. 111-125). West Sussex, England: John Wiley.

Beckett, N. E., & Park, B. (1995). Use of category versus individuating information: Making base rates salient. Personality and Social Psychology Bulletin, 21, 21-31.

Berger, J., Cohen, B. P., Zelditch, M. Jr. (1966). Status characteristics and expectation states. In Sociological Theories in Progress, Vol. 1, ed. J. Berger, M. Zelditch Jr., B. Anderson, pp. 29-46. Boston: Houghton Mifflin.

Berger, J., Fisek, M. H., Norman, R. Z., & Zelditch, M. (1977). Status characteristics and social interaction. New York: Elsevier Scientific Pub. Co.

Berger, J., Ridgeway, C. L., Fisek, M. H.,&Norman, R. Z. (1998). The legitimation and delegitimation of power and prestige orders. American Sociological Review, 63, 379–405.

Berger, J., Rosenholtz, S. J., & Zelditch, M. (1980). Status organizing processes. Annual Review of Sociology, 6: 479-508.

Berger, J., Wagner, D. G., & Zelditch, M. (1985). Introduction: Expectation states theory. In J. Berger & M. Zelditch (Eds.), Status, rewards, and influence: 1-72. San Francisco: Jossey-Bass.

Biddle, B. J. (1979). Role theory: Expectations, identities, and behavior. New York: Academic Press.

Bigoness, W. J. (1976). Effect of applicant's sex, race, and performance on employers' performance ratings: Some additional findings. Journal of Applied Psychology, 62, 301-311.

Blanton, J. (1996). A study of the relationship between leadership behavior components

and demographic variables among middle school principals. (Doctoral dissertation.) Texas Southern University, TX.

Bolman, L. G. & Deal, T. E., (1991). Images of leadership. National Center for Educational Leadership. (Occasional Paper No. 7.) Cambridge, MA.

Borenstein, M., Hedges, L.V., Higgins, J.P.T., & Rothstein, H.R. (2009). Introduction to Meta-analysis. Wiley: Chichester, 2009.

Brenner,O.C., Tomkiewicz, J.,&Schein,V. E. (1989). The relationship between sex-role stereotypes and requisite management characteristics revisited. Academy of Management Journal, 32, 662–669.

Brewer, M.B., & Kramer, R.M. (1985). The psychology of intergroup attitudes and behavior. In M.R. Rosenweig & L.W. Porter (Eds.), Annual review of Psyhcology (Vol. 36, pp. 219-244). Palo Alto, CA: Annual Review.

Browne, K. (1999). Divided labours: An evolutionary view of women at work. New Haven, CT: Yale University Press.

Budig, M. J. (2002). Male advantage and the gender composition of jobs: Who rides the glass escalator? Social Problems, 49, 258-277.

Bushman, B. J., & Anderson, C. A. (2001). Media violence and the American public: Scientific facts versus media misinformation. American Psychologist, 56, 477–489.

Byrne, D. (1971). The attraction paradigm. New York: Academic.

Cantor, J., & Wilson, B. J. (2003). Media Violence: Intervention strategies for reducing aggression. Media Psychology, 5, 363-403.

Carless, S. A. (1998). Gender differences in transformational leadership: an examination of supervisor, leader, and subordinate perspectives. Sex Roles: A Journal of Research, (311-912), 887-888.

Carson, M. A. (2011). Antecedents of effective leadership: The relationships between social skills, transformational leadership, leader effectiveness, and trust in the leader. (Doctoral dissertation.) University of North Carolina at Charlotte. Charlotte, NC.

Catalyst. (2012a). Statistical Overview of Women in the Workplace. Retrieved May 21, 2012 from http://www.catalyst.org/publication/219/statistical-overview-of-women-in-the-workplace

Catalyst. (2012b). Women in Management in the United States, 1960-Present. Retrieved

May 30, 2012 from http://www.catalyst.org/publication/207/women-in-management-in-the-united-states-1960-present

Center for American Women and Politics. (2012a). Fact sheet: Women in the U.S. Congress 2011. Retrieved from http://www.cawp.rutgers.edu/fast_facts/levels_of_office/documents/cong.pdf

Center for American Women and Politics. (2012b). Fact sheet: Women in Elective Office. Retrieved from http://www.cawp.rutgers.edu/fast_facts/levels_of_office/documents/elective.pdf

Chemers, M. M. (1997). An integrative theory of leadership. Mahwah, NJ: Erlbaum.

Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

Colquitt, J. A., Conlon, D. E., Wesson, M. J., Porter, C. O. L. H., & Ng, K. Y. (2001). Justice at the millennium: A meta-analytic review of 25 years of organizational justice research. Journal of Applied Psychology, 86: 425-445.

Conlin, M. (2003). The new gender gap: From kindergarten to grad school, boys are becoming the second sex. BusinessWeek. May 26, 2003.

Correll, S. J. (2001). Gender and the Career Choice Process: The Role of Biased Self-Assessments. American Journal of Sociology, 106, 1691–1730.

Correll, S. J., & Ridgeway, C. L. (2003). Expectation states theory. In J. Delamater (Ed.), Handbook of social psychology: 29-51. New York: Kluwer Academic/Plenum.

Cortina, J. M. (2003). Apples and oranges (and pears, oh my!): The search for moderators in meta-analysis. Organizational Research Methods, 6: 415-439.

Cota, A.,&Dion, K. (1986). Salience of gender and sex composition of ad hoc groups: An experimental test of distinctiveness theory. Journal of Personality and Social Psychology, 50, 770–776.

Crocker, J., & Major, B. (1989). Social stigma and self-esteem: The self protective properties of stigma. Psychological Review, 96, 608–630.

Davison, H. K., & Burke, M. J. (2000). Sex discrimination in simulated employment contexts: A meta-analytic investigation. Journal of Vocational Behavior, 56, 225–248.

DeRue, D. S., Nahrang, J. D., Wellman, N., & Humphrey, S. E. (2011). Trait and behavioral theories of leadership: An integration and meta-analytic test of their relative validity. Personnel Psychology, 64, 7–52.

Devine, P. G. (1989). Stereotypes and prejudice: Their automatic and controlled components. Journal of Personality and Social Psychology, 56, 5–18.

Douglas, C. & Ammeter, A. P. (2004). An examination of leader political skill and its effect on ratings of leader effectiveness. The Leadership Quarterly. 15, 537-550.

Eagly, A. H. (1987). Sex differences in social behavior: A social-role interpretation. Hillsdale, NJ: Erlbaum.

Eagly, A. H. (2007). Female leadership advantage and disadvantage: Resolving the contradictions. Psychology of Women Quarterly 31: 1–12.

Eagly, A. H., & Carli, L. L. (2003a). The female leadership advantage: An evaluation of the evidence. Leadership Quarterly, 14, 807–834.

Eagly, A. H., & Carli, L. L. (2003b). Finding gender advantage and disadvantage: Systematic research integration is the solution. The Leadership Quarterly, 14, 851–859.

Eagly, A. H. & Carli, L.L. (2007). Through the Labyrinth. Boston, MA: Harvard Business School Press.

Eagly, A.H. & Johannesen-Schmidt, M.C. (2001). "The Leadership Style of Women and Men." Journal of Social Issues, 57:781-97.

Eagly, A. H., Johannesen-Schmidt, M. C., & van Engen, M. I. (2003). Transformational, transactional, and laissez-faire leadership styles: a meta-analysis comparing women and men. Psychological Bulletin, 129, 569–591.

Eagly, A. H., & Johnson, B. T. (1990). Gender and leadership style: a meta-analysis. Psychological Bulletin, 108, 233–256.

Eagly, A. H., & Karau, S. J. (2002). Role congruity theory of prejudice toward female leaders. Psychological Review, 109, 573–598.

Eagly, A., Makhijani, M., & Klonsky, B. (1992). Gender and the evaluation of leaders: A meta-analysis. Psychological Bulletin, 111, 3–22.

Eagly, A. H., Karau, S. J., & Makhijani, M. G. (1995). Gender and the effectiveness of leaders: A meta-analysis. Psychological Bulletin, 117, 125-145.

Egger, M., Smith, D. G., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. British Medical Journal, 315, 629-634.

Ellemers, N., Van Knippenberg, A., DeVries, N., & Wilke, H. (1988). Social

identification and permeability of group boundaries. European Journal of Social Psychology, 18, 497-513.

Ellemers, N.,Wilke, H., & Van Knippenberg, A. (1993). Effects of the legitimacy of low group or individual status on individual and collective status-enhancement strategies. Journal of Personality and Social Psychology, 64, 766-778.

Emrich, C. G., Denmark, F. L., & Den Hartog, D. N. (2004). Cross-cultural differences in gender egalitarianism: Implications for societies, organizations, and leaders. In R. J. House, P. J. Hanges, M. Javidan, P. W. Dorfman, & V. Gupta (Eds.), Culture, leadership, and organizations: The GLOBE study of 62 societies (pp. 343–394). Thousand Oaks, CA: Sage.

European Commission. (2010). Gender balance in decision making. Retrieved from http://ec.europa.eu/social/main.jsp?catId_764&langId_en

Fairhurst, G. T., & Snavely, B. K. (1981). An examination of the communication between high status tokens and their dominant colleagues: Kanter's theory reexamined. Academy of Management Journal, 26, 353-361.

Fairhurst, G. T., & Snavely, B. K. (1983). Majority and token minority group relationships: Power acquisition and communication. Academy of Management Review, 8, 292-300.

Farh, J. L., A. S. Tsui, K. Xin, B. S. Cheng. (1998). The influence of relational demography and guanxi: The Chinese case. Organization Science, 9 471-488.

Feldman, J. M. (1981). Beyond attribution theory: Cognitive processes in performance appraisal. Journal of Applied Psychology, 66, 127-148.

Fiske, S. T., & Neuberg, S. L. (1990). A continuum of impression formation, from category-based to individuating processes: Influences of information and motivation on attention and interpretation. In M. P. Zanna (Ed.), Advances in experimental social psychology (Vol. 23, pp. 1-74). San Diego, CA: Academic Press.

Floge, L., & Merrill, D. (1985). Tokenism reconsidered: Male nurses and female physicians in a hospital setting. Social Forces, 64, 925–947.

Foschi, M., (1996). Double standards in the evaluation of men and women. Social Psychology Quarterly, 59:237-54

Foschi, M. (2000). Double standards for competence: Theory and research. Annual Review of Sociology 26, 21–42.

Frear, K. A., & Paustian-Underdahl, S. C. (2011). From elusive to obvious: Improving

performance management through specificity. Industrial and Organizational Psychology: Perspectives on Science and Practice, 4, 198–200.

Gergen, D. (2005). Women leading in the twenty-first century. In L. Coughlin, E. Wingard, & K. Hollihan (Ed.), Enlightened power: How women are transforming the practice of leadership. San Francisco: Jossey-Bass.

Geyskens, I., Krishnan, R., Steenkamp, J. E. M., & Cunha, P. V. (2009). A review and evaluation of meta-analysis practices in management research. Journal of Management, 35, 393–419.

Glaman J, Jones A, & Rozelle R. (1996). The effects of co-worker similarity on the emergence of affect in work teams. Group and Organization Management, 21, 192–215.

Goldberg, S. (1993). Why men rule: A theory of male dominance. Chicago: Open Court.

Goldberg, C., Riordan, C. M., & Zhang, L. (2008). Employees' perceptions of their leaders: Is being similar always better? Group & Organization Management, 33, 330-355.

Gomez-Mejia, L. R., McCann, J. E., & Page, R. C. (1985). The structure of managerial behaviors and rewards. Industrial Relations, 24, 147–154.

Greenhaus, J. H., & Parasuraman, S. (1999). Research on work, family, and gender: Current status and future directions. In G. N. Powell (Ed.), Handbook of gender & work (pp. 391–412). Thousand Oaks, CA: Sage.

Haccoun, D. M., Haccoun, R. R., & Sallay, G. (1978). Sex differences in the appropriateness of supervisory styles: A nonmanagement review. Journal of Applied Psychology, 63, 124-127.

Hamner, W. C., Kim, J. S., Baird, L., & Bigoness, W. J. (1974). Race and sex as determinants of ratings by potential employers in a simulated work-sampling task. Journal of Applied Psychology, 59, 705-711.

Harrison, D., Newman, D., & Roth, P. L. (2006). How important are job attitudes? Meta-analytic comparisons of integrative behavioral outcomes and time sequences. Academy of Management Journal, 49, 305-326.

Hedges, L. V., & Olkin, I. (1985). Statistical methods for meta-analysis. Orlando, FL: Academic Press.

Hedges, L. V., & Vevea, J. L. (1998). Fixed and random effects models in meta-analysis. Psychological Methods, 3, 486–504.

Hefferman, M. (2002). The female CEO ca. 2002, August. Fast Company, 61.

Heilman, M. E. (1983). Sex bias in work settings: The lack of fit model. Research in Organizational Behavior, 5, 269–298.

Heilman, M. E. (1995). Sex stereotypes and their effects in the workplace: What we know and what we don't know. Journal of Social Behavior and Personality, 10, 3–26.

Heilman, M. E. (2001). Description and prescription: How gender stereotypes prevent women's ascent up the organizational ladder. Journal of Social Issues, 57, 657–674.

Heilman, M. E., Block, C. K., Martell, R. F., & Simon, M. C. (1989). Has anything changed? Current characteristics of men, women, and managers. Journal of Applied Psychology, 74, 935–942. doi:10.1037/0021-9010.74.6.935

Heilman, M. E., Martell, R., & Simon, M. (1988). The vagaries of bias: Conditions regulating the undervaluation, equivaluation and overvaluation of female job applicants. Organizational Behavior and Human Decision Processes, 41, 98–110.

Heilman, M. E., & Okimoto, T. G. (2007). Why are women penalized for success at male tasks?: The implied communality deficit. Journal of Applied Psychology, 92, 81–92.

Helgesen, S. (1990). The female advantage: Women's ways of leadership. New York: Doubleday.

Hofstede, G., Neuijen, B., Ohayv, D., & Sanders, G. (1990). Measuring Organizational Cultures: A Quantitative study Across Twenty Cases. Administrative Science Quarterly, 35: 286-316.

Hofstede, G. (2001) Culture's Consequences: Comparing Values, Behaviors, Institutions and Organizations Across Nations, Sage: Thousand Oaks, CA.

Hogan, R., Curphy, G. J., & Hogan, J. (1994). What we know about leadership. American Psychologist, 49, 493–504.

House, R.J., Hanges, P.J., Javidan, M., Dorfman, P.W., Gupta, V. (2004) (Eds.), Leadership, Culture, and Organizations: The GLOBE study of 62 societies (pp. 437–512). Thousand Oaks, CA: Sage

House, R., Javidan, M., Hanges, P., & Dorfman, P. (2002). Understanding cultures and implicit leadership theories across the globe: An introduction to project GLOBE. Journal of World Business, 37: 3–10.

Huddy, L., & Terkildsen, N. (1993a). The consequences of gender stereotypes for women

candidates at different levels and types of office. Political Research Quarterly, 46, 503–525.

Hunt, J. G. (1991). Leadership: A new synthesis. Newbury Park, CA: Sage.

Hunt, J. G., Boal, K. B., & Sorenson, R. L. (1990). Top management leadership: Inside the black box. Leadership Quarterly, 1, 41–65.

Hunter, J. E., & Schmidt, F. L. (2004). Methods of meta-analysis: Correcting error and bias in research findings (2nd ed.). New York, NY: Academic Press.

Inter-Parliamentary Union. 2011. Women in national parliaments. http://www.ipu.org/wmn-e/classif.htm

Jacobson, M. B., & Effertz, J. (1974). Sex roles and leadership: Perceptions of the leaders and the led. Organizational Behavior and Human Performance, 12, 383-396.

Jago, A. G., & Vroom, V. H. (1982). Sex differences in the incidence and evaluation of participative leader behavior. Journal of Applied Psychology, 67, 776-783.

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.

Jurik, N. (1985). An officer and a lady: Organizational barriers to women working as correctional officers in men's prisons. Social Problems, 32, 375–388.

Kanter, R. (1977a). Men and Women of the Corporation. New York: Basic.

Kanter, R. (1977b). Some effects of proportions on group life: Skewed sex ratios and responses to token women. American Journal of Sociology, 82, 965–990.

King, E. B., Hebl., M. R., George, J. M., & Matusik, S. F. (2010). Understanding tokenism: Antecedents and consequences of a psychological climate of gender inequity. Journal of Management, 36: 482-510.

Koenig, A. M., Eagly, A. H., Mitchell, A. A., & Ristikari, T. (2011). Are leader stereotypes masculine? A meta-analysis of three research paradigms. . Psychological Bulletin, 137(4), 616-642.

Kryger, B.R., & Shikiar, R. (1978). Sexual discrimination in the use of letters of recommendation: A case of reverse discrimination. Journal of Applied Psychology, 63, 309-314.

Lauterbach, K. E. & Weiner, B. J. (1996). Dynamics of upward influence: How male and female managers get their way. Leadership Quarterly 7, (1), 87-107.

Lindberg, S. M., Hyde, J. S., Petersen, J. L., & Linn, M. C. (2010). New trends in gender and mathematics performance: A meta-analysis. Psychological Bulletin, 136, 1123–1135.

Lipman-Blumen, J. (2000). Connective leadership: Managing in a changing world. New York: Oxford University Press.

Lipsey, M.W. (2003). Those confounded moderators in meta-analysis: The good, bad, and ugly. The Annals of the American Academy of Political and Social Science, 587, 69-81.

Lipsey, M. W., & Wilson, D. B. (2001). Practical meta-analysis. Thousand Oaks, CA: Sage.

Locksley, A . , Hepbum, C . , Ortiz, V. (1982). Social stereotypes and judgments of individuals: An instance of the base-rate fallacy. Journal of Experimental Social Psychology. 1 8:23-42.

Lord, R. G., & Maher, K. J. (1993). Leadership and information processing: Linking perceptions and performance. New York: Routledge.

Lyness, K. S., & Thompson, D. E. (2000). Climbing the corporate ladder: do female and male executives follow the same route? Journal of Applied Psychology, 85(1), 86–101.

Macrae, C. N., Bodenhausen, G. V., Milne, A. B., & Ford, R. L. (1997). On the regulation of recollection: The intentional forgetting of stereotypical memories. Journal of Personality and Social Psychology, 72, 709– 719.

Martell, R. F. (1999, May). Sensitive dependence in organizations: How just a little gender bias can cause the ‘glass ceiling’ in organizations. Paper presented at the meeting of the Midwestern Psychological Association, Chicago.

Martell, R. F., Lane, D. M., & Emrich, C. G. (1996). Male–female differences: A computer simulation. American Psychologist, 51, 157–158.

Martell, R. F., Parker, C., Emrich, C. G., & Crawford, M. S. (1998). Sex stereotyping in the executive suite: “Much ado about something.” Journal of Social Behavior and Personality, 13, 127–138.

Martin, S.H. (1980). Breaking and entering: Policewomen on patrol. Berkeley: University of California Press.

Martin, P. Y. (1985). Group sex composition in work organizations: A structural-normative model. Research in the Sociology of Organizations, 4, 311-349.

McCauley, C. D. (2004). Successful and unsuccessful leadership. In J. Antonakis, A. T. Cianciolo, & R. J. Sternberg (Eds.), The nature of leadership (pp. 199-221). Thousand Oaks, CA: Sage.

McDaniel, M. A., Rothstein, H. R. & Whetzel, D. L. (2006). Publication bias: A case study of four test vendors. Personnel Psychology, 59: 927-953.

Meeker, B. F, & Weitzel-O'Neill, P. A. (1977). Sex roles and interpersonal behavior in task-oriented groups. American Sociological Review, 91-105.

Mumford, M. D. (1986). Leadership in the organizational context: A conceptual approach and its applications. Journal of Applied Social Psychology, 16, 508-531.

Neter J, Wasserman W, Kutner M. (1989). Applied linear regression models (2nd ed.). Homewood, IL: Irwin.

Nieva, V. F., & Gutek, B. A. (1980). Sex effects on evaluation. Academy of Management Review, 5: 267–276.

Nieva, V. E., & Gutek, B. A. (1981). Women and Work: A Psychological Perspective. New York: Praeger.

Nieva, V. F., & Gutek, B. A. (1982). Women and work." A psychological perspective. New York: Praeger.

Orlitzky, M., & Benjamin, J. D. (2003). The effects of sex composition on small-group performance in a business school competition. Academy of Management Learning and Performance, 2, 128-138.

Paolillo, J. G. (1981). Manager's self assessments of managerial roles: The influence of hierarchical level. Journal of Management, 7, 43–52.

Pavett, C. M., & Lau, A. W. (1983). Managerial work: The influence of hierarchical level and functional specialty. Academy of Management Journal, 26, 170–177.

Pendry, L. F., & Macrae, C. N. (1994). Stereotypes and mental life: The case of the motivated but thwarted tactician. Journal of Experimental Social Psychology, 30, 303–325.

Petty, M. M., & Lee, G. IC (1975). Moderating effects of sex of supervisor and subordinate on relationships between supervisory behavior and subordinate satisfaction. Journal of Applied Psychology, 60, 624-628.

Phillips, J. S., & Lord, R. G. (1982). Schematic information processing and perceptions

of leadership in problem-solving groups. Journal of Applied Psychology, 67,486-492.

Pothukuchi, V., Damanpour, F., Choi, J., Chen, C.C. and Park, S.H. (2002) National and organizational culture differences and international joint venture performance, Journal of International Business Studies, 33(2): 243-265.

Powell, G. N. (1988). Women & men in management. Newbury Park, CA: Sage.

Powell, G. N., & Butterfield, D. A. (1979). The "good manager": Masculine or androgynous? Academy of Management Journal, 22, 395–403.

Powell, G. N., Butterfield, D. A., & Parent, J. D. (2002). Gender and managerial stereotypes: Have the times changed? Journal of Management, 28, 177–193.

Reskin, B. (1988). Bringing men back in: Sex differentiation and the devaluation of women's work. Gender & Society, 2, 58-81.

Reskin, B. (2000). The Proximate Causes of Employment Discrimination. Contemporary Sociology, 29:319-28.

Reskin, B. F., & Ross, C. E. (1995). Jobs, authority, and earnings among managers: The continuing significance of sex. In J. A. Jacobs (Ed.), Gender inequality at work (pp. 127–151). Thousand Oaks, CA: Sage.

Ridgeway, C.L. (1997). Interaction and the Conservation of Gender Inequality: Considering Employment. American Sociological Review 62, 218–35.

Ridgeway, C. L. (2001). Gender, status, and leadership. Journal of Social Issues, 57, 637–655.

Ridgeway, C. L., & Berger, J. (1986). Expectations, legitimation, and dominance behavior in groups. American Sociological Review, 51, 603–617.

Riggio, R. E., Riggio, H. R., Salinas, C., & Cole, E. J. (2003). The role of social and emotional communication skills in leader emergence and effectiveness. Group Dynamics: Theory, Research, and Practice. 7, (2), 83-103.

Robertson, T. M. (1986). Fundamental strategies for wage and salary administration. Personnel Journal, 65(11): 120-132.

Rogelberg, S. G., & Rummery, S. M. (1996). Gender diversity, team decision quality, time on task, and interpersonal cohesion. Small Group Research, 27(1), 79-90.

Rosenthal, R. (1979). The ''file drawer problem'' and tolerance for null results. Psychological Bulletin, 86, 638-641.

Rosenthal, R., & Rubin, D. B. (1986). Meta-analytic procedures for combining studies with multiple effect sizes. Psychological Bulletin, 99, 400−406.

Rosener, J. B. (1995). America's competitive secret: Utilizing women as a management strategy. New York: Oxford University Press.

Rosette, A. and Tost, L. (2010). Agentic women and communal leadership: how role prescriptions confer advantage to top women leaders, Journal of Applied Psychology, Vol. 95, pp. 221-35.

Rosnow, R. L., & Rosenthal, R. (1989). Statistical procedures and the justification of knowledge in psychological science. American Psychologist, 44, 1276–1284.

Roth, P. L., Purvis, K. L., & Bobko, P. (2012). A meta-analysis of gender group differences for measures of job performance in field studies. Journal of Management, 38, 2, 719-739.

Rothbart, M. (1981). Memory processes and social beliefs. In D. L. Hamilton (Ed.), Cognitive processes in stereotyping and intergroup behavior: 145–181. Hillsdale, NJ: Erlbaum.

Rustad, M. (1982). Women in khaki: The American enlisted woman. New York: Praeger.

Sackett, P. R., Harris, M. M., & Orr, J. M. (1987). On seeking moderator variables in the meta-analysis of correlation data: A Monte Carlo investigation of statistical power and resistance to Type I error. Journal of Applied Psychology, 71, 302-310.

Sanchez-Meca, J., & Marin-Martinez, F. (1998). Weighting by inverse variance or by sample size in meta-analysis: A simulation study. Educational and Psychological Measurement, 58, 211-220.

Sarbin, T. R., & Allen, V. L. (1968). Role theory. In G. Lindzey & E. Aronson (Eds.), Handbook of social psychology (2nd ed., Vol. 1, pp. 488–567). Reading, MA: Addison-Wesley.

Schein, V. E. (1973). The relationship between sex role stereotypes and requisite management characteristics. Journal of Applied Psychology, 57, 95–100.

Schein, V. E. (1975). The relationship between sex role stereotypes and requisite management characteristics among female managers. Journal of Applied Psychology, 60, 340–344.

Schein, V. E., Mueller, R., Lituchy, T., & Liu, J. (1996). Think manager–think male: A global phenomenon? Journal of Organizational Behavior, 17, 33–41.

Schein, V. E. (2001). A global look at psychological barriers to women's progress in management. Journal of Social Issues, 57, 675–688.

Schoorman, F. D., Mayer, R. C., & Davis, J. H. (2007). An integrative model of organizational trust: Past, present, and future. Academy of Management Review, 32, 344-354.

Sharpe, R. (2000). As leaders, women rule: New studies find that female managers outshine their male counterparts in almost every measure. Business Week, 74, November 20, 2000.

Shore, L. M., Cleveland, J. N., & Goldberg, C. B. (2003). Work attitudes and decisions as a function of manager age and employee age. Journal of Applied Psychology, 88, 529–537.

Smith-Lovin, L., Skvoretz, J. V., & Hudson, C. G. (1986). Status and participation in six-person groups: A test of Skvoretz's comparative status model. Social Forces, 64, 992-1005.

Statistics on Women in the Military (2011). Retrieved May 30, 2012 from http://www.womensmemorial.org/PDFs/StatsonWIM.pdf

Steel, P. D., & Kammeyer-Mueller, J. D. (2002). Comparing meta-analytic moderator estimation techniques under realistic conditions. Journal of Applied Psychology, 87: 96-111.

Tajfel H, & Turner JC. (1979). An integrative theory of intergroup conflict. In Austin WG, Worchel S (Eds.), The social psychology of intergroup relations (pp. 33–47). Monterey, CA: Brooks/Cole.

Terrin, N., Schmid, C. H., Lau, J., & Olkin, I. (2003). Adjusting for publication bias in the presence of heterogeneity. Statistics in Medicine, 22, 2113-2126.

Tsui, A. S. (1984). A multiple-constituency framework of managerial reputational effectiveness. In J. G. Hunt, D. Hosking, C. A. Schriesheim, & R. Stewart (Eds.), Leaders and managers: International perspectives on managerial behavior and leadership (pp. 28-44). New 'York: Pergamon Press.

Tsui, A. S., & Gutek, B. A. (1984). A role set analysis of gender differences in performance, affective relationships, and career success of industrial middle managers. Academy of Management Journal, 27, 619-635.

Tsui, A. S., & O'Reilly, C. A. (1989). Beyond simple demographic effects: The importance of relational demography in superior-subordinate dyads. Academy of Management Journal, 32, 402-423.

Tsui, A. S., Porter, L. W., & Egan, T. D. (2002). When both similarities and dissimilarities matter: Extending the concept of relational demography. Human Relations, 55, 899–929.

Tucker, M. L., McCarthy, A. M., & Jones, M. C. (1999). Women and men politicians: Are some of the best leaders dissatisfied? Leadership & Organization Development Journal, 20, (6), 285-290.

Turner, J.C. (1985). Social categorization and the self-concept: A social cognitive theory of group behavior. In Lawler EJ (Ed.), Advances in group processes: Theory and research (pp. 77–121). Greenwich, CT: JAI Press.

Turner, J.C., Hogg M.A., Oakes, P.J., Reicher, S.D., Wetherell, M.S. (1987). Rediscovering the social group: A self-categorization theory. Oxford: Blackwell.

United States Government Accountability Office. (2010). Women in Management: Female Managers' Representation, Characteristics, and Pay. Retrieved July 21, 2011 from http://www.gao.gov/new.items/d101064t.pdf

U.K. Equality and Human Rights Commission. (2008). Sex and power 2008. London: U.K. Equality and Human Rights Commission.

U.S. Bureau of Labor Statistics. (2010). Current Population Survey, Annual Averages: Household Data. (Characteristics of the employed, Table 9: Employed persons by occupation, sex and age). Retrieved March 26, 2010, from http://www.bls.gov/cps/cpsaat9.pdf

U.S. Bureau of Labor Statistics. (2011). Labor force statistics from the Current Population Survey. Retrieved from http://www.bls.gov/cps/wlf-table11-2011.pdf

van Engen, M. L., van der Leeden, R., & Willemsen, T. M. (2001). Gender, context and leadership styles: A field study. Journal of Occupational and Organizational Psychology, 74, 581– 598.

Vecchio, R. P. (2002). Leadership and gender advantage. The Leadership Quarterly, 13(6), 643–671.

Vecchio, R. P. (2003). In search of gender advantage. The Leadership Quarterly, 14, 835–850.

Vilkinas, T. (2000). The gender factor in management: How significant others perceive effectiveness. Women in Management Review. 15, (5/6), 261-271.

Wagner, D. G., & Berger, J. (1997). Gender and interpersonal task behaviors: Status expectation accounts. Sociological Perspectives, 40, 1–32.

Watson, C. (1988). When a woman is the boss: Dilemmas in taking charge. Group and Organization Studies, 13, 163-181.

Weber, R., & Cracker, J. (1983). Cognitive processes in the revision of stereotypic beliefs. Journal of Personality and Social Psychology, 45, 961-977.

Weber, Y., Shenkar, O. & Raveh, A. (1996). National and Corporate Culture Fit in Mergers/Acquisitions: An Exploratory Study. Management Science, 42, 1215-1227.

West, C. & Zimmerman, D. (1987). Doing Gender. Gender and Society, 1, 125–51.

Whitener, E. M. 1990. Confusion of confidence intervals and credibility intervals in meta-analysis. Journal of Applied Psychology, 75: 315-321.

Wiley, M. G., & Eskilson, A. (1982). The interaction of sex and power base on perceptions of managerial effectiveness. Academy of Management Journal, 25, 671-677.

Willemsen,T.M.(2002). Gender typing of the successful manager— A stereotype reconsidered. Sex Roles, 46, 385–391.

Williams, C. L. (1992). The glass escalator: Hidden advantages for men in "female" professions. Social Problems, 39, 253- 267.

Woehr, D.J., & Roch, S.G. (1996). Context effects in performance evaluation: the impact of ratee sex and performance level on performance ratings and behavioral recall. Organizational Behavior and Human Decision Processes, 66(1):31.41

Yoder, J. D. (1991). Rethinking tokenism: Looking beyond numbers. Gender & Society, 5, 178-192.

Yoder, J. D., Adams, J., & Prince, H. T. (1983). The price of a token. Journal of Political and Military Sociology, 11, 325– 337.

Young, C. J., Mackenzie, D. L., & Sherif, C. W. (1980). In search of token women in academia. Psychology of Women Quarterly, 4, 508–525.

Yukl, G. (ed.) (2002). Leadership in Organizations, 5th edn. Upper Saddle River, NJ: Prentice Hall.

Yukl, G. (2006). Leadership in organizations. Upper Saddle River, NJ.

Yukl, G., & Van Fleet, D. D. (1992). Theory and research on leadership in organizations.

In M. D. Dunnette & L. M. Hough (Eds.), Handbook of industrial and organizational psychology (2nd ed., Vol. 3, pp. 147-197). Palo Alto, CA: Consulting Psychologists Press.

Zimmer, L. (1986).Women guarding men. Chicago, IL: University of Chicago Press.

APPENDIX A: CODING MANUAL

Coding Manual

Carefully read the method and results sections of each study. The first stage of coding will be to obtain information about the sample and setting of the study. You will also need to determine how many effect sizes will be coded from each study. We will be coding separate effect sizes for rater type, for effectiveness measure type, and rater sex (if reported separately within each study). To be safe, if you think there are multiple forms/ perspectives of effectiveness, code it all.

Study/ moderator Data Coding

Study ID: A unique identification number will be assigned to each study. If a study reports two or more independent studies (i.e., studies with different sets of participants), then a decimal point will be added to the study ID number to distinguish each data set within a study, and to code each one separately. Additionally, once you get to the effect size coding portion of the excel sheet, you will indicate how many effect sizes can be extracted from the article.

Authors

Year

Sex of 1st author

% male authors

Title

Journal Name

Source: 1) Journal, 2) Dissertation/ thesis, 3) Book, 4) Other

N male leaders

N female leaders

Industry: 1) business, 2) educational, 3) government (legislators, government-operated facilities, elected office holders, etc) 4) military, 5) social service (social workers), or6) mixed or other

List industry

Leader role: briefly describe the leadership role (i.e. coach, manager, executive, principal)

% men in role (if listed)

Nationality: 1) American/ Canadian, 2) European, 3) Australian/New Zealander, 4) Asian, 5) African, 6) Latin American, 7) Middle Eastern, or 8) Mixed

Nationality: List the specific country

Mean age (if reported for men and women separately, compute weighted average = (age for men) (N men) + (age for women) (N for women)/ Total N

Mean age for men

Mean age for women

Leader level: 1) 1st level/supervisor, 2) middle/ manager, 3) upper/ senior/ executive, 4) other/ mixed

Study setting: 1) Organization, 2) Laboratory, 3) other/ mixed

List how effectiveness is assessed (generally)

Effect Size Coding

Effect size number: Assign each effect size within a study a unique number. Number multiple effect sizes within a study sequentially.

Type of data effect size based on: 1) means and SDs, 2) t-value or F-value, 3) correlation, 4) chi square (df=1)

Page number where data was found (use the original article pages, not the .pdf pages)

Rater type: (1) self, (2) subordinates, (3) boss(es), (4) peers, (5) judges or trained observers, (6) objective counting device, (e.g., organizational productivity), (7) mixed or unclear

Types of leadership effectiveness:

- **content**:

  (1) objective effectiveness (i.e group/ unit/ organizational performance; Examples of objective measures include groups' speed of completing a geometric puzzle (Eskilson, 1975), test of managers' tacit knowledge of business (Horgan & Simeon, 1988), workers' productivity defined by the proportion of their production goals they met, and seasonal win-loss records of athletic team.)

  (2) subjective effectiveness (i.e., perceptions of performance, effectiveness, or leader abilities; satisfaction with the leader),

(3) an overall assessment of effectiveness/ mixed (measures that don't seem to fit elsewhere)

If coded as 2) subjective-- further code the data into one of five classes: (some examples come from Eagly et al, 1995 meta-analysis: 1) effectiveness, 2) ability (sub-code in next column as task, relational, or mixed), 3) effort/ motivation to perform, 4) satisfaction with leader, and 5) other or mixed

1. Effectiveness: Examples include (a) a rating scale anchored by poor leader and outstanding leader (Day & Stogdill, 1972, p. 355), b) a rating of the leader's overall performance

2. Ability: Examples include (a) ratings of 29 leadership skills such as "training officers/committee members" and "planning club/group activities," given on scales anchored by excellent and poor (Couch, 1980, pp. 114-115), and (b) ratings of nine managerial abilities such as decisiveness, leadership, and organizing and planning, given on nine scales anchored by outstanding and poor (Friedman, 1981, p. 67).

- **If coded as 2) ability, further code these items into 1) task-related, 2) relational (other-oriented) abilities, or 3) mixed**

3. Effort or motivation to perform well: Examples include (a) ratings of the effort managers "put forth on the job" on a scale anchored by slightly below that of others \or far exceeds that of others (Renwick, 1977, p. 407).

4. Satisfaction with the leader: Examples of the measures of satisfaction with leaders include (a) the satisfaction with supervisors scale from the Job Descriptive Index (Adams, 1978), (b) ratings of the extent to which subordinates would "like to continue working for this supervisor" and (c) ratings of the extent to which subordinates were satisfied with the "helpfulness of your supervisor in getting the job done" (Gupta, Jenkins, & Beehr, 1983, p. 178).

5. Other or mixed: Examples include (a) frequency and intensity of feeling reported on the personal accomplishment subscale of the Maslach Burnout Inventory (Maslach & Jackson, 1981), which assesses "feelings of competence and successful achievement in one's work with people" (Caccese & Mayerberg, 1984, p. 281); (b) subordinates' rating of their own performance and (c) supervisors' ratings of their subordinates' performance in areas such as "motivation to work hard," "potential for promotion," and "overall performance" on scales anchored by outstanding and very poor.

- **levels of analysis**: (1) individual, (2) group, and (3) organizational

% men amongst raters (may need to be estimated)

Sex of raters: (1) reported together, (2) reported separately (report M/ SDs separately)

Reliability of effectiveness measure (alpha):

Effect size data- Gender and Effectiveness

Mean for men

SD for men

N for men

Mean for women

SD for women

N for women

Mean for men

SD for men

Mean for women

SD for women

t-value

Correlation between gender and effectiveness

total N

Rater identity groups:

Report the effect size data separately for different raters groups (if it is reported separately within the study).

(1) self, (2) subordinates, (3) boss(es), (4) peers, (5) judges or trained observers, (6) objective counting device, (e.g., organizational productivity), (7) mixed or unclear

Rater sex groups:

Report the effect size data separately for different raters sex groups (if it is reported separately within the study).

When the raters who provided the leadership effectiveness measure are different from the leaders or managers rated (e.g., they were subordinates) and these ratings were reported separately for male and female raters, effect sizes were calculated separately for the male and female raters (and the significance and direction of the sex comparisons will be recorded separately). These additional effect sizes will supplement the effect sizes that were combined over both sexes of raters and used in most analyses.

Male raters

Female raters

APPENDIX B: CODING AGREEMENT

Coding Agreement Across 3 Time Points

Coder A and B	Average agreement
3/12/2012-3/26/2012	0.865789
3/27/2012-4/2/2012	0.857895
4/3/2012-4/23/2012	0.9385
Coder A and C	
3/12/2012-3/26/2012	0.950526
3/27/2012-4/4/2012	0.875
4/5/2012-4/18/2012	0.8805

APPENDIX C: REFERENCES FOR PRIMARY STUDIES

Adams, T., & Keim, M. (2000). Leadership practice and effectiveness among Greek student leaders. College Student Journal, 34, 259-270.

Agezo, C.K., & Hope, W.C. (2011). Gender leadership in Cape Coast Municipality primary schools. International Journal of Leadership in Education, 14 (2), 181-201.

Antonaros, M.E. (2010). Gender differences in leadership style: A study of leader effectiveness in higher education. (Doctoral dissertation.) University of Michigan, Ann Arbor, MI.

Atwater, L., & Roush, P. (1994). An investigation of gender effects on followers' ratings of leaders, leaders' self-ratings and reactions to feedback. The Journal of Leadership Studies, 1 (4), 37-52.

Ayman, R., Korabik, K., & Morris, S. (2009). Is transformational leadership always perceived as effective? Male subordinates' devaluation of female transformational leaders. Journal of Applied Social Psychology, 39 (4), 852-879.

Barker, C. J., (1982). The effect of selected demographic variables on Hollander's theory of idiosyncracy credit. (Doctoral dissertation.) University of Maryland, MD.

Bartol, K. M., (1973). Male and female leaders in small work groups: An empirical study of satisfaction, performance, and perceptions of leader behavior. East Lansing Division of Research, Michigan State University, East Lansing, MI.

Bartone, P. T. & Snook, S.A. (2002). Cognitive and personality predictors of leader performance in West Point Cadets. Military Psychology, 14 (4), 321-338.

Bass, B. M., Avolio, B. J., & Atwater, L., (1996). The transformational and transactional leadership of men and women. Journal of Applied Psychology, 45 (1), 5-34.

Bess, M. M., (2001). Perceived leadership effectiveness of male and female directors of schools in West and East Tennessee. (Doctoral dissertation.) Tennessee State University, TN.

Blanton, J. (1996). A study of the relationship between leadership behavior components and demographic variables among middle school principals. (Doctoral dissertation.) Texas Southern University, TX.

Bolman, L. G. & Deal, T. E., (1991). Images of leadership. National Center for Educational Leadership. (Occasional Paper No. 7.) Cambridge, MA.

Bolman, L. G. & Deal, T. E., (1992). Leading and managing: Effects of context, culture, and gender. Education Administration Quarterly. 28, (3), 314-329.

Burke, S. & Collins, K. M., (2001). Gender differences in leadership styles and management skills. Women in Management Review, 16, (5), 244-256.

Burns, G. & Martin, B. N. (2010). Examination of the effectiveness of male and female educational leaders who made use of the invitational leadership style of leadership. Journal of Invitational Theory and Practice. 16, 29-55.

Caccese, T. M. & Mayerberg, C. K., (1984). Gender differences in perceived burnout of college coaches. Journal of Sport Psychology. 6, 279-288.

Carless, S. A., (1998). Gender differences in transformational leadership: An examination of superior, leader, and subordinate perspectives. Sex Roles. 39, (11/12), 887-902.

Carson, M. A. (2011). Antecedents of effective leadership: The relationships between social skills, transformational leadership, leader effectiveness, and trust in the leader. (Doctoral dissertation.) University of North Carolina at Charlotte. Charlotte, NC.

Couch, M. E. (1980). Self-assessment of leadership skills by 4-H volunteer leaders in the southern region of the United States. (Doctoral dissertation.) Texas Tech University, TX.

Daughtry, L. H. & Finch, C. R. (1997). Effective leadership of vocational administrators as a function of gender and leadership style. Journal of Vocational Education Research. 22, 3, 173-186.

Davidson, B. J. (1995). Leadership styles of successful male and female college choral directors. (Doctoral dissertation.) Arizona State University, Tempe, AZ.

Davy, J. & Shipper, F. (2004). A cross gender study of managerial skills, employees' attitudes and managerial performance. (Unpublished paper.)

Day, D. R. & Stogdill, R. M. (1972). Leader behavior of male and female supervisors: A comparative study. Personnel Psychology 25, 353-360.

Deaux, K. (1979). Self-Evaluations of male and female managers. Sex Roles. 5, 571-580.

Douglas, C. (2011). The moderating role of leader and follower sex in dyads on the leadership behavior-leader effectiveness relationships. The Leadership Quarterly, 23, 1, 163–175.

Douglas, C. & Ammeter, A. P. (2004). An examination of leader political skill and its effect on ratings of leader effectiveness. The Leadership Quarterly. 15, 537-550.

Duehr, E. E., (2006). Personality, gender, and transformational leadership: Investigating differential prediction for male and female leaders. (Doctoral dissertation.) University of Minnesota, MN.

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, 3, 259-278.

Elesser, K. M. & Lever, J. (2011). Does gender bias against female leaders persist? Quantitative and qualitative data from a large-scale survey. Human Relations. 64, (12) 1555-1578.

Eskilson, A. (1975). Sex composition and leadership. (Doctoral thesis.) University of Illinois at Chicago Circle, Chicago, IL.

Ezell, H. F., Odewahn, C. A., & Sherman, J. D. (1980). Perceived competence of women managers in public human service organizations: A comparative view. Journal of Management. 6, (2), 135-144.

Folkman, J. & Zenger, J. (2011). A study in leadership. Women do it better than men. Zenger-Folkman, Orem, UT. Retrieved from: http://www.zfco.com/media/articles/ZFCo.WP.WomenBetterThanMen.031312.pdf

Francke, C. A. (1975). Perceived performance differences between women and men supervisors and implications for training. (Doctoral dissertation.) Michigan State University, MI.

Friedman, M. J. (1980). Differences in assessment center performance as a function of the race and sex of rates and the race of assessors. (Doctoral dissertation.) University of Tennessee, Knoxville, TN.

Goktepe, J. R. & Schneier, C. R., (1988). Sex and gender effects in evaluating emergent leaders in small groups. Sex Roles, 19, (1/2), 29-36.

Gore, S., (2000). Perceived leadership effectiveness of male and female directors of schools. (Doctoral dissertation.) Tennessee State University, TN.

Gross, N., & Trask, A. E. (1976). The sex factor and the management of schools. New York: Wiley.

Gupta, N., Jenkins, G. D., & Beehr, T. A., (1983). Employee gender, gender similarity, and supervisor-subordinate cross-evaluation. Psychology of Women Quarterly, 8, (2), 174-184.

Hemphill, J. K., Griffiths, D. E., & Frederiksen, N. (1962). Administrative performance and personality: A study of the principal in a simulated elementary school. New York: Teachers College, Columbia University.

Horgan, D. D. & Simeon, R. J. (1988). Gender, mentoring, and tacit knowledge. (Report No. CG-021658). Memphis State University.

Hoyle, J. R. & Randall, R. S. (1967). Problem-attack behavior and sex, prior teaching experience, and college preparation of elementary school principals. Journal of Experimental Education. 36, (2) 28-33.

Hua, P. C. (2005). The relationship between gender, presidential leadership, and the quality of community colleges. (Doctoral dissertation.) University of Southern California, CA.

Hura, G. M. (2005). The effects of rater and leader gender on ratings of leader effectiveness and attributes in a business environment. (Doctoral dissertation.) University of Akron, OH.

Jacobson, M. B. & Effertz, J. (1974). Sex roles and leadership perceptions of the leaders and the led. Organizational Behavior and Human Performance. 12, 383-396.

Johnson, S. K., Murphy, S. E., Zewdie, S., & Reichard, R. J. (2008). The strong, sensitive type: Effects of gender stereotypes and leadership prototypes on the evaluation of male and female leaders. Organizational Behavior and Human Decision Processes. 106, 39-60.

Kabacoff, R. I. (1998). Differences in organizational leadership. A large sample study. Management Research Group. Presented at the 106th Annual Convention of the American Psychological Association in San Francisco, California, August 1998.

King, P. J. (1978). An analysis of teachers' perceptions of the leadership styles and effectiveness of male and female elementary school principals. (Doctoral dissertation.) University of Southern California, CA.

Knott, K. B. & Natalle, E. J. (1997). Sex differences, organizational level, and superiors' evaluation of managerial leadership. Management Communication Quarterly. 10, 523-540.

Komives, S. R. (1991). The relationship of same-and cross-gender work pairs to staff performance and supervisor leadership in residence hall units. Sex Roles. 24 (5/6) 355-363.

Lauterbach, K. E. & Weiner, B. J. (1996). Dynamics of upward influence: How male and female managers get their way. Leadership Quarterly 7, (1), 87-107.

Leithwood, K. & Jantzi, D. (1997). Explaining variation in teachers' perceptions of principals' leadership: a replication. Journal of Educational Administration. 35, (4), 312-331.

Love, A. (2007). Teacher's perception of the leadership effectiveness of female and male principals. (Doctoral dissertation.) Tennessee State University, TN.

Millard, R. J., & Smith, K. H. (1985). Moderating effects of leader sex on the relation between leadership style and perceived behavior patterns. Genetic, Social and General Psychology Monographs, 111, 305-316.

Moehlman, G.M. (1988). The relationship between leadership styles and faculty perceived effectiveness of principals in secondary vocational-technical schools in Michigan. (Doctoral dissertation.) The Ohio State University, OH.

Mohr, E. S. & Downey, R.G. (1977). Are women peers? Journal of Occupational Psychology. 50, 53-57.

Mohr, G. & Wolfram, H. J. (2008). Leadership and effectiveness in the context of gender: The role of leaders' verbal behavior. British Journal of Management 19, 4-16.

Morrison, A. M., White, R. P., & Van Velsor, E. (1987). Breaking the glass ceiling: Can women reach the top of America's largest corporations? Reading, MA: Addison-Wesley.

Moss, J. & Jensrud, Q. (1995). Gender, leadership, and vocational education. Journal of Industrial Teacher Education. 33, 1, 6-23.

Munson, C. E. (1979). Evaluation of male and female supervisors. Social Work, 104-110.

Nogay, K. (1995). The relationship of superordinate and subordinate gender to the perceptions of leadership behaviors of female secondary principals. (Doctoral dissertation.) Youngstown State University. Youngstown, OH.

Osborn, R. N. & Vicars, W. M. (1976). Sex stereotypes: An artifact in leader behavior and subordinate satisfaction analysis? Academy of Management Journal 19, (3), 439-449.

Pasteris, P. J. (1998). Investigating the differences in leadership styles and effectiveness between male and female public high-school principals in Illinois. (Doctoral dissertation.) Northern Illinois University. DeKalb, IL.

Patrick, J. F. (1978). The relationship between sex of supervisors, attitudes toward women, and the job satisfaction of female employees. (Doctoral dissertation.) University of Maryland, MD.

Peters, L. H., O'Connor, E. J., Weekley, J., Pooyan, A., Frank, B., & Erenkrantz, B. (1984). Sex bias and managerial evaluations: A replication and extension. Journal of Applied Psychology 69, (2), 349-352.

Petty, M. M. & Lee, G. K. (1975). Moderating effects of sex of supervisor and subordinate on relationships between supervisory behavior and subordinate satisfaction. Journal of Applied Psychology, 60, (5), 624-628.

Pratch, L. & Jacobowitz, J. (1996). Gender, motivation, and coping in the evaluation of leadership effectiveness. Consulting Psychology Journal: Practice and Research. 48, (4), 203-220.

Quinn, K. I. (1976). Self-perceptions of leadership behaviors and decisionmaking orientations of men and women elementary school principals in Chicago public schools. (Doctoral dissertation.) University of Illinois at Urbana-Champaign. Urbana, IL.

Ragins, B. R. (1989). Power and gender congruency effects in evaluations of male and female managers. Journal of Management, 15, (1), 65-76.

Ragins, B. R. (1991). Gender effects in subordinate evaluations of leaders: Real or artifact? Journal of Organizational Behavior, 12, (3), 259-268.

Randall, M. J. (1998). United States Air Force Nurse Captains' perceived leadership effectiveness. (Thesis.) Arizona State University, AZ.

Renwick, P. A. (1977). The effects of sex differences on the perception and management of superior-subordinate conflict: An exploratory study. Organizational Behavior and Human Performance. 19, 403-415.

Reyes, R. I. (1997). Leadership effectiveness of department chairpersons in the University of Los Andes. (Doctoral dissertation.) University of South Florida, Tampa, FL.

Rice, R. W., Bender, L. R., & Vitters, A. G. (1980). Leader sex, follower attitudes toward women, and leadership effectiveness: A laboratory experiment. Organizational Behavior and Human Performance 25, 46-78.

Rice, R. W., Yoder, J. D., Adams, J., Priest, R. F., & Prince, H. T. (1984). Leadership ratings for male and female and military cadets. Sex Roles. 10, (11/12), 885-901.

Riggio, R. E., Riggio, H. R., Salinas, C., & Cole, E. J. (2003). The role of social and emotional communication skills in leader emergence and effectiveness. Group Dynamics: Theory, Research, and Practice. 7, (2), 83-103.

Rivero, A.J. (2003) Sex and Gender Differences in Actual and Perceived Leader Effectiveness: Self and Subordinate Views (Thesis.) University of North Texas, TX.

Rohmann, A. & Rowold, J. (2009). Gender and leadership style. A field study in different organizational contexts in Germany. Equal Opportunities International 28, (7), 55-560.

Rosser, V. J. (2003). Faculty and staff members' perceptions of effectiveness leadership: Are there differences between women and leaders? Equity & Excellence in Education. 36, (1), 71-81.

Schneier, C. E. & Bartol, K. M. (1980). Short Notes. Sex effects in emergent leadership. Journal of Applied Psychology. 65, (3), 341-345.

Schulman, J. M. (1989). Power behavior, job satisfaction and leadership effectiveness in public school principals. (Doctoral dissertation.) University of Bridgeport, Bridgeport, CT.

Simpson, M. W. (1985). Teacher satisfaction with leadership based on principal sex role orientation. (Doctoral dissertation.) University of Georgia, GA.

Singh, S. K. (2007). Emotional intelligence and organizational leadership: A gender study in Indian context. International Journal of Indian Culture and Business Management. 1, (1/2), 48-62.

Skottheim, K. (2008). DI in relation to team and leaders evaluations. In Ekelund, B.Z.Langvik.E. (eds): Diversity Icebreaker. How to manage diversity processes. Oslo, Human Factors Publ.

Sparks & Kuhnert, (2011). Navigating the leadership labyrinth: Perceived outcomes for men and women. Paper presented at the Society of Industrial and Organizational Psychology Conference in Atlanta, GA, 2011.

Stephens, H. E. (2004). Gender-role orientation and leadership effectiveness. (Thesis.) University of Guelph, Ottawa, Canada.

Sutton, J. L. (1998). Effects of task content, task type, and leader gender on perceptions

of gender differences in leadership effectiveness. (Doctoral dissertation.) Texas Tech University, TX.

Thompson, M. D. (2000). Gender, leadership orientation, and effectiveness: Testing the theoretical models of Bolman & Deal and Quinn. Sex Roles, 42, (11/12), 969-992.

Tucker, M. L., McCarthy, A. M., & Jones, M. C. (1999). Women and men politicians: Are some of the best leaders dissatisfied? Leadership & Organization Development Journal, 20, (6), 285-290.

Vecchio, R. P. & Anderson, R. J., (2009). Agreement in self-other ratings of leader effectiveness: The role of demographics and personality. International Journal Of Selection and Assessment. 17, (2), 165-179.

Vilkinas, T. (2000). The gender factor in management: How significant others perceive effectiveness. Women in Management Review. 15, (5/6), 261-271.

Vilkinas, T., Shen, J., & Cartan, G. (2009). Predictors of leadership effectiveness for Chinese managers. Leadership & Organization Development Journal, 30, (6), 577-590.

Wexley, K. N. & Pulakos, E. D. (1982). Sex effects on performance ratings in manager-subordinate dyads: A field study. Journal of Applied Psychology, 67, (4), 433-439.

White, D. R. (2007). The relationship of leadership effectiveness to gender in Louisiana middle school principals. (Doctoral dissertation.) University of Louisiana at Monroe, LA.

Yoder, J. & Adams, J. (1984). Women entering nontraditional roles: When work demands and sex-roles conflict. The case of West Point. International Journal of Women's Studies. 7, 260-272.

York, C. D. (2005). Leadership effectiveness: Investigating the influences of leader sex, gender, and behaviors on self and other perceptions. (Doctoral dissertation.) University of North Texas, TX.