

Journal of Applied Psychology

Gender and Perceptions of Leadership Effectiveness: A Meta-Analysis of Contextual Moderators

Samantha C. Paustian-Underdahl, Lisa Slattery Walker, and David J. Woehr

Online First Publication, April 28, 2014. <http://dx.doi.org/10.1037/a0036751>

CITATION

Paustian-Underdahl, S. C., Walker, L. S., & Woehr, D. J. (2014, April 28). Gender and Perceptions of Leadership Effectiveness: A Meta-Analysis of Contextual Moderators. *Journal of Applied Psychology*. Advance online publication. <http://dx.doi.org/10.1037/a0036751>

Gender and Perceptions of Leadership Effectiveness: A Meta-Analysis of Contextual Moderators

Samantha C. Paustian-Underdahl
Florida International University

Lisa Slattery Walker and David J. Woehr
University of North Carolina at Charlotte

Despite evidence that men are typically perceived as more appropriate and effective than women in leadership positions, a recent debate has emerged in the popular press and academic literature over the potential existence of a female leadership advantage. This meta-analysis addresses this debate by quantitatively summarizing gender differences in perceptions of leadership effectiveness across 99 independent samples from 95 studies. Results show that when all leadership contexts are considered, men and women do not differ in perceived leadership effectiveness. Yet, when other-ratings only are examined, women are rated as significantly more effective than men. In contrast, when self-ratings only are examined, men rate themselves as significantly more effective than women rate themselves. Additionally, this synthesis examines the influence of contextual moderators developed from role congruity theory (Eagly & Karau, 2002). Our findings help to extend role congruity theory by demonstrating how it can be supplemented based on other theories in the literature, as well as how the theory can be applied to both female and male leaders.

Keywords: gender, leadership, leader effectiveness, gender roles

Supplemental materials: <http://dx.doi.org/10.1037/a0036751.supp>

Although the proportion of women in the workplace has increased remarkably within the past few decades, women remain vastly underrepresented at the highest organizational levels (U.S. Bureau of Labor Statistics, 2011). Women occupy a mere 3.8% of Fortune 500 chief executive officer seats (Catalyst, 2012b) and represent only 3.2% of the heads of boards in the largest companies of the European Union (European Commission, 2012). The numbers are only slightly better in the political arena. In 2012 women held only 90 of the 535 seats (16.8%) in the U.S. Congress (Center for American Women and Politics, 2012b) and 19.1% of parliamentary seats globally (Inter-Parliamentary Union, 2012). Why are women so underrepresented in the most elite leadership positions? For decades researchers have proposed several explanations (e.g., Berger, Fisek, Norman, & Zelditch, 1977; Eagly & Karau, 2002; Heilman, 2001; Schein, 1973, 2007).

One explanation for women's underrepresentation in elite leadership positions points to the undervaluation of women's effectiveness as leaders. This explanation is supported by several theoretical perspectives including lack of fit theory (Heilman, 2001),

role congruity theory (RCT; Eagly & Karau, 2002), expectation states theory (Berger et al., 1977; Ridgeway 1997, 2011), and the think manager–think male paradigm (Schein, 1973, 2007). Despite research showing that men may be perceived as better suited for and more effective as leaders than women (e.g., Carroll, 2006; Eagly, Makhijani, & Klonsky, 1992), some popular press publications have reported the opposite: that there may be a female gender advantage in modern organizations that require a “feminine” type of leadership (e.g., Conlin, 2003; R. Williams, 2012).

The *New York Times* concluded that “no doubts: women are better managers” (Smith, 2009), and an article in the *Daily Mail* agreed (“Women in Top Jobs Are Viewed as ‘Better Leaders’ Than Men,” 2010). An article published in *Psychology Today* reported new data exploring “why women may be better leaders than men. [Is] women’s leadership style more suited to modern organizations?” (R. Williams, 2012). The arguments for a “female advantage” in leadership generally stem from the belief that women are more likely than men to adopt collaborative and empowering leadership styles, while men are disadvantaged because their leadership styles include more command-and-control behaviors and the assertion of power. Yet, an academic discussion among leadership and gender researchers criticized the simplicity of these arguments, proposing that studies should not be asking whether there is a perceived gender difference in leadership but rather when and why there may be gender differences in perceived leadership effectiveness (Eagly & Carli, 2003a, 2003b; Vecchio, 2002, 2003).

Vecchio (2002) discussed the importance of examining the leadership context, proposing that “advocates of a gender advantage perspective offer a simplistic, stereotypic view that largely ignores the importance of contextual contingencies” (p. 655).

Samantha C. Paustian-Underdahl, Department of Management and International Business, Florida International University; Lisa Slattery Walker, Department of Sociology, University of North Carolina at Charlotte; David J. Woehr, Department of Management, University of North Carolina at Charlotte.

Correspondence concerning this article should be addressed to Samantha C. Paustian-Underdahl, Department of Management and International Business, Florida International University, 11200 Southwest 8th Street, Miami, FL 33174. E-mail: spaustia@fiu.edu

Eagly and Carli (2003a) agreed, stating that “contemporary journalists, while surely conveying too simple a message . . . must approach these issues with sophisticated enough theories and methods that they illuminate the implications of gender in organizational life” (p. 808). These arguments support the use of meta-analysis to address gender differences in perceptions of leadership effectiveness because of its ability to summarize a large body of studies while taking into account the influence of contextual moderators.

Such contextual moderators were discussed in Eagly and Karau’s RCT (2002), which suggested that, in general, prejudice toward female leaders follows from the perceived incongruity between the characteristics of women and the requirements of leader roles. Eagly and Karau (2002) also proposed that prejudice toward female leaders can vary depending on features of the leadership context as well as characteristics of leaders’ evaluators. An early meta-analysis on gender differences in leadership effectiveness conducted by Eagly, Karau, and Makhijani (1995) reflected this RCT that two of its authors later published (Eagly & Karau, 2002). The current meta-analysis also uses RCT as a theoretical framework, yet we believe that our study can extend this theory in two ways. First, due to the recent cultural shifts supporting a possible female advantage in leadership, we argue that RCT can be applied beyond female leaders, to also explain perceptions of incongruence affecting male leaders.

Second, we aim to supplement RCT by considering how aspects of the double standards of competence model (Foschi, 2000), the cognitive load paradigm (Macrae, Hewstone, & Griffiths, 1993), and tokenism (Kanter, 1977a) offer theoretical explanations that contrast with or add to what is proposed by RCT. For instance, RCT proposes that men may be seen as more effective leaders in male-dominated or senior leadership positions, due to the masculine nature of those roles. Yet, Foschi (1996, 2000) argued that a woman’s presence in a top leadership role or a male-dominated position provides information about her abilities to others in the organization (i.e., that she must be exceptionally competent to have made it in such a high-status and challenging leadership role). Thus, in this study we are able to examine which perspective is empirically supported by meta-analytic data, and, thus, we can provide insight into ways RCT can be supplemented by other theoretical perspectives.

In addition to examining these contributions to RCT, we address a call in the literature to examine the unique effects of self-ratings versus other-ratings of leadership effectiveness. Vecchio (2002) criticized Eagly and Johnson (1990) for including self-ratings of leadership effectiveness in their meta-analysis, as he believed “this type of assessment is presently regarded as highly suspect in the field of leadership research” (p. 650). Eagly and Carli (2003a) refuted his point by arguing that ignoring self-ratings “violates the meta-analytic principle of including a wide range of methods and disaggregating based on method” (p. 814). Yet, in their meta-analysis, Eagly et al. (1995) examined how the summary effect for gender differences in leadership effectiveness varied based on rater source, but they did not conduct hierarchical subgroup analyses to examine the effects of each moderator separately per rater group.

We believe that by examining the effects of moderators on gender differences in self-ratings compared to other-ratings of leadership effectiveness, we can clarify how role incongruity may vary depending on aspects of the context as well as the source of

the rater. As such, the current comprehensive meta-analysis answers a call in the literature to clarify gender advantage and disadvantage through “systematic research integration” that examines gender differences between self-ratings and other-ratings of leadership effectiveness across a variety of leadership contexts (Eagly & Carli, 2003b, p. 851).

Our meta-analysis makes three primary contributions to the literature on gender and perceptions of leadership effectiveness. First, we expand upon and update an early meta-analysis conducted by Eagly et al. (1995), which integrated relevant research conducted in the United States until 1989. In the past 23 years, not only have new perspectives and research appeared but also more rigorous meta-analytical methods have been developed, which we use in the current study (i.e., random- and mixed-effects models; Borenstein, Hedges, Higgins, & Rothstein, 2009; Hedges & Vevea, 1998; Lipsey & Wilson, 2001). Second, we aim to extend RCT by applying it to both men and women and by examining how other theories can supplement aspects of RCT. Finally, we address an important point of contention raised in the academic discussion of gender advantages in leadership effectiveness: the importance of examining self-reported and other-reported leadership effectiveness (Eagly & Carli, 2003a, 2003b; Vecchio, 2002, 2003).

Theory and Hypotheses

RCT was developed in part from social role theory, which argues that individuals develop descriptive and prescriptive gender role expectations of others’ behavior based on an evolutionary sex-based division of labor (Eagly, 1987; Eagly & Wood, 2012). This division of labor has traditionally associated men with breadwinner positions and women with homemaker positions (Eagly & Wood, 2012). Based on these social roles, women are typically described and expected to be more communal, relations-oriented, and nurturing than men, whereas men are believed and expected to be more agentic, assertive, and independent than women. The agentic characteristics associated with men are consistent with traditional stereotypes of leaders (Schein 1973, 2007). RCT builds upon social role theory by considering the congruity between gender roles and leadership roles and proposing that people tend to have dissimilar beliefs about the characteristics of leaders and women and similar beliefs about the characteristics of leaders and men (Eagly & Karau, 2002).

According to the theory, when occupying leadership positions, women likely encounter more disapproval than men due to perceived gender role violation (Eagly & Karau, 2002). Yet, Eagly and Karau (2002) also proposed that perceptions of incongruity can vary depending on features of the leadership context as well as characteristics of leaders’ evaluators. Indeed, Eagly et al. (1995) surveyed respondents and found that a leadership role requiring behaviors consistent with encouraging participation and open consideration was considered to be feminine, while a role requiring the ability to direct and control people was rated as masculine in nature. On the basis of this notion, we argue that RCT can also be applied to men when occupying certain leadership positions that may be seen as incongruent with the agentic characteristics associated with the male gender role. As organizations have become faster paced, globalized environments, some organizational scholars have proposed that a more feminine style of leadership is

needed to emphasize the participative and open communication needed for success (e.g., Hitt, Keats, & DeMarie, 1998; Volberda, 1998).

Helgesen (1990) and Rosener (1995) proposed that female leaders are more inclined to fill this leadership need than men, by drawing upon characteristics they are encouraged to uphold as part of their femininity including an emphasis on cooperation rather than competition and equality rather than a supervisor–subordinate hierarchy. More recently, Koenig, Eagly, Mitchell, and Ristikari (2011) conducted a meta-analysis examining the extent to which stereotypes of leadership are culturally masculine and determined that “leadership now, more than in the past, appears to incorporate more feminine relational qualities, such as sensitivity, warmth, and understanding” (p. 634). To the extent that organizations shift away from a traditional masculine view of leadership and toward a more feminine and transformational outlook, women should experience reduced prejudice, while men may be seen as more incongruent with leadership roles. We propose, based on RCT, that key aspects of the leadership context will affect the extent to which leadership roles are seen as congruent or incongruent with both male and female gender roles, which may help to explain whether men or women are seen as more effective leaders in different situations. To address these contextual moderators of gender differences in leadership effectiveness, we undertook a quantitative synthesis of studies that compared men and women on measures of leadership effectiveness.

Time of Study as a Moderator

RCT proposes that it is important to consider how time may moderate gender differences in perceptions of leadership effectiveness (Eagly & Karau, 2002). Over the past several decades, more women have entered the paid labor force and increased their representation in many leadership roles. Women’s participation in the U.S. labor force has increased from 33% in 1950 to 59.2% in 2012 (U.S. Department of Labor, 2012). Additionally, women currently represent 14.3% of corporate officers in America’s 500 largest companies (vs. 8.7% in 1995; Catalyst, 2012b). 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 American Women and Politics, 2012b) and holding 12% of governor seats (vs. 0% in 1950; Center for American Women and Politics, 2012a). The same pattern is being observed in countries other than the United States as well (see European Commission, 2012; Inter-Parliamentary Union, 2012).

This increase in female participation in leadership roles may be associated with a weakening of the perceived incongruity between women and leadership. Powell, Butterfield, and Parent (2002) proposed 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; Weber & Crocker, 1983). Thus, as more women have entered into and succeeded in leadership positions, it is likely that people’s stereotypes associating leadership with masculinity have been dissolving slowly over time. Additionally, researchers have proposed that definitions of effective managerial behaviors have changed in response to features of modern organizational environments, to include less masculine practices

(relations-oriented and team-focused practices; McCauley, 2004). Koenig et al. (2011) conducted a meta-analysis on the stereotypes of leadership and masculinity and found that leadership perceptions became less masculine over time. 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 be more similar now than in the past.

Hypothesis 1: Publication date will moderate gender differences in perceptions of leadership effectiveness such that there will be greater gender differences (favoring men) seen among older studies and smaller gender differences or differences that favor women seen among newer studies.

Type of Organization as a Moderator

RCT highlights the importance of the fit between gender roles and the requirements of leader roles, and it 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, organizations that are highly male dominated or culturally masculine in their demands present particular challenges to women because of the incompatibility of these demands with people’s expectations about women. This incompatibility not only restricts women’s access to such organizations but also can compromise perceptions of women’s effectiveness. When leader roles are particularly masculine, people may suspect that women are not qualified for them and may resist women’s authority (Eagly & Karau, 2002; Heilman, 2001). Although leadership positions are typically considered to be masculine and male typed, they can vary widely in these aspects. Some types of organizations are considered to be feminine and are occupied by more women than men (e.g., social service and educational organizations; United States Government Accountability Office, 2010).

Studies have found support for the effect of organization on perceptions of gender differences in leadership effectiveness. A meta-analysis that integrated the results of 76 effect sizes found that male leaders were seen as more effective than female leaders in organizations that were male dominated or masculine in other ways (i.e., numerically male-dominated organizations; military roles; Eagly et al., 1995). Additionally, female leaders were seen as more effective than male leaders in less male-dominated or less masculine organizations (i.e., educational, governmental, and social service organizations). Thus, the extent to which organizations are male dominated is one moderator proposed by RCT.

Hypothesis 2: In organizations that are male dominated, women will be considered to be less effective leaders than men, and in organizations that are female dominated, men will be considered to be less effective leaders than women.

Hierarchical Level as a Moderator

Eagly and Karau (2002) reviewed research pertaining to the kinds of leader behaviors associated with different hierarchical levels of leadership (e.g., Martell, Parker, Emrich, & Crawford, 1998; Pavett & Lau, 1983). Research has supported the idea that different hierarchical levels require different types of behaviors (e.g., McCauley, 2004; Paolillo, 1981). Lower level managers have reported relying on abilities involving direct supervision of

employees' task involvement, such as monitoring potential problems and managing conflict. Eagly et al. (1995) argued that such characteristics may be considered masculine in nature, leading to greater perceived congruity for men than women in lower level supervisory positions. However, more recent studies have shown that the characteristics associated with lower level leadership positions may be considered to be gender neutral in nature. Mumford, Campion, and Morgeson (2007) discussed that the majority of skills needed in lower level supervisor roles involve "cognitive" skills including effective communication, active learning, and critical thinking. Given the gender neutrality of these skills, at the lowest hierarchical levels there may not be a gender advantage for men or women in leadership effectiveness.

In contrast, middle level managers believed that their roles required greater relational skills, such as fostering cooperative effort and motivating and developing subordinates. In middle management positions, more relational and transformational leadership behaviors are needed, and women are considered to be more likely than men to engage in such behaviors; thus, women may be seen as more effective in middle management than men (Eagly & Karau, 2002).

RCT also proposes that perceptions of leadership are likely to be the most masculine for higher status, senior leadership positions, thereby increasing role incongruity for women in these positions (Eagly & Karau, 2002). Indeed, research has shown that the higher the level of leadership, the more masculine and agentic are the expected behaviors for the leader (Eagly & Karau, 2002; Hunt, Boal, & Sorenson, 1990; Lord & Maher, 1993; Martell et al., 1998). Eagly and Karau (2002) proposed that as the incongruity between the female gender role and the leadership roles increases, so should the discrepancy in gender differences of perceived leadership effectiveness. However, recent research on double standards of competence (Foschi, 1996, 2000) proposes a different hypothesis.

Though 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) argued 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 vignettes supported this prediction (Rosette & Tost, 2010). Rosette and Tost found that top female leaders received more positive evaluations than top male leaders, as well as mid-level managers who were male or female, because women at the top were perceived to have faced higher standards than their male counterparts or those at lower levels.

We propose, based on RCT, that there will not be a gender difference in perceptions of men and women's congruity and effectiveness in lower level leadership positions; in middle management positions, however, women will be seen as more congruent and effective than men. Finally, on the basis of the double standards of competence model, we argue that women who reach

and succeed at the very top of organizations may be evaluated favorably than men. They have demonstrated that they have overcome double standards both to arrive in their top position and to excel in that top position that is dominated by men and perceived to be particularly masculine.

Hypothesis 3: Hierarchical level will moderate gender differences in perceived leadership effectiveness, with female leaders being rated as more effective than male leaders at middle and upper hierarchical levels (with no differences in ratings expected at lower hierarchical levels).

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). Research in social cognition has demonstrated that when individuals experience high cognitive load, feel tired, or are under time pressure, they tend to rely more on stereotypes to form impressions of others (e.g., 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., Eskilson, 1975; Jacobson & Effertz, 1974; York, 2005). The group members are typically 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, these individuals may be less likely to rely upon think manager-think male stereotypes and biases in making judgments of leadership effectiveness.

Hypothesis 4: Gender differences in perceptions of leadership effectiveness will be greater (favoring men) in organizational settings than in laboratory settings.

Percent of Male Raters as a Moderator

Many studies of assessments of leaders' performance occur in lab or organizational group settings where the members of a leader's work group assess the leader's effectiveness. RCT proposes that sex ratios in work groups should moderate gender differences in perceived leadership effectiveness based on the concept of tokenism (Kanter, 1977a as cited in Eagly & Karau, 2002). The theory argues that as the percentage of male raters increases, the female-stereotypical qualities of women leaders become more salient; 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. Kanter found that tokenism has several consequences for minority group within the workplace: higher visibility (and increased scrutiny), exaggeration of

differences from majority group members, exclusion from informal workplace interactions, and assimilation.

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), and by the first policewomen on patrol (Martin, 1980). These findings are explained with the concept of numeric gender imbalance. Kanter (1977b) proposed that the unique effects of numerical underrepresentation will take place for both male and female tokens; yet, for high-status tokens (men), the outcomes may differ. A high-status token might receive more positive than negative attention, and perceptions of his behaviors may be distorted such that he is seen as more rather than less competent. Similarly, research on the *glass escalator* effect has shown that men are likely to be promoted to leadership positions more quickly than women in female-dominated groups or organizations (e.g., Maume, 1999; C. Williams, 1995).

Thus, we argue that as the percentage of male raters becomes very high, women may be seen as less effective due to the increased perceptions of their femininity and lessened leadership abilities. Yet, when the percentages of male and female raters are close to equal, gender-related characteristics should become less salient to the group, and there should be small or nonexistent gender differences in perceptions of leadership effectiveness (Eagly & Carli, 2007). Finally, when there is a majority of female raters in the group, men may be seen as more effective than women, due to the increased perceptions of their masculinity, competence, and leadership abilities. 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.

Hypothesis 5: There will be a nonlinear relationship between percentage of male raters and gender differences in perceptions of 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. At the extremes, men will be seen as more effective.

Rating Source as a Moderator

Use of self-ratings versus other-ratings of leadership effectiveness has consistently been a point of contention in the literature on gender and leadership (e.g., Eagly & Carli, 2003a, 2003b; Vecchio, 2002, 2003). Additionally, there is considerable literature highlighting the importance of gender to self-evaluations of leadership. Social role theory argues that individuals develop descriptive and prescriptive gender role expectations based on an evolutionary sex-based division of labor in which women were historically homemakers and men were breadwinners (Eagly, 1987; Eagly & Wood, 2012). As such, gender may play a critical role in self ratings of performance in work settings, such that men may see themselves as more suited for and effective in leadership roles than women may consider themselves to be.

Consistent with this notion, research using 360-degree performance evaluations found that male managers were more likely to overestimate their effectiveness, while female managers were

more likely to rate themselves consistently with ratings by their peers and subordinates (Brutus, Fleenor, and McCauley, 1999; Vecchio and Anderson, 2009). A study on causal attributions suggested that women have a greater tendency to underrate their performance because they tend to attribute their success to external factors more than men do (Parsons, Meece, Adler, & Kaczala, 1982). Men, on the other hand, have been shown to have higher self-esteem than women do (Kling, Hyde, Showers, & Buswell, 1999), which may explain why men tend to have higher self-evaluations than women. For these reasons, we argue that gender differences in perceived leadership effectiveness may depend on the source of the rating.

Hypothesis 6a: The source of the rating will moderate gender differences in perceptions of leadership effectiveness such that there will be greater gender differences favoring men seen among self-ratings than among other-ratings.

Yet, we contend that the degree to which self-ratings reflect a gender advantage in leadership effectiveness for men may be stronger in male-typed settings or when leaders are engaged in male-typed tasks (Beyer, 1990, 1992). Eagly and Karau (2002) proposed that, in addition to being affected by contextual moderators impacting perceptions others have regarding the incongruity between the female gender role and leadership roles, women in leadership positions may be impacted by contextual factors such that they themselves can exhibit diminished self-confidence (Lenny, 1977) and expectancy-confirming behavior (Geis, 1993) in certain environments. Thus, despite the tenets of RCT being primarily focused on other-ratings of leaders, we argue that self-ratings will also be moderated by the contextual variables presented above.

An empirical study examining gender differences in self-rated job performance found that women significantly underrated their performance and recalled more task failure than had actually occurred when they had engaged in masculine tasks but not when they had engaged in gender neutral or feminine tasks (Beyer, 1990, 1992). Additionally, Correll (2001) examined men's and women's perceptions of their mathematics ability, an academic area that is generally considered to be male-typed. She found that, controlling for positive performance feedback about mathematical ability, men's assessments of their own mathematical competence were higher than women's assessments. When comparing self-assessments of verbal ability, an academic area that is generally considered to be female-typed, women rated themselves as more competent than men (Correll, 2001). We propose, based on these arguments and findings, that the contextual moderators described above may moderate the extent to which self-ratings as well as other-ratings of leadership effectiveness favor males versus females.

Hypothesis 6b: Gender differences in self-ratings and other-ratings of perceived leadership effectiveness will be moderated by the contextual moderators presented above.

Method

Literature Search and Inclusion Criteria

To gather primary studies to include in this meta-analysis, we conducted an extensive literature search to select studies published

through 2011, initially using the keywords *leadership performance* and *leadership effectiveness*. Studies found using these keywords were manually searched for data on gender differences in these 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*. We also searched through numerous review articles, books, and recent Academy of Management and Society for Industrial and Organizational Psychology conference proceedings (2010–2012), as well as the reference lists of other related meta-analyses (Eagly et al., 1992, 1995). In addition, we 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 *Psychological Bulletin*. Finally, e-mails 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 data sets. The search yielded 270 potential articles and dissertations, which were reviewed for their ability to meet the specific inclusion criteria discussed below.

Consistent with Eagly et al. (1995), our criteria for including studies in the meta-analysis consisted of the following: (a) the study compared male and female leaders, executives, managers, directors, supervisors, principals, or administrators; (b) participants were at least 18 years old; (c) the study assessed the effectiveness of at least five leaders of each sex; and (d) measures of leaders' effectiveness included one of the following: performance or leadership ability; ratings of satisfaction with leaders or satisfaction with leaders' performance; coding or counting of effective leadership behaviors; or measures of organizational productivity or group performance. Authors were contacted for more information if the appropriate data appeared to have been collected but were not reported in the paper.

If a study reported data separately for different countries or different types of organizations, the samples of leaders were treated as independent. The first author carefully tracked authors of multiple studies in order to determine if the same sample and data may have been reported in multiple studies. If the same sample was used in more than one study, only data from one of the studies were coded (e.g., Bolman & Deal, 1991, 1992). Lab studies that involved leaderless group discussions in which no one was designated to fill a leadership role were excluded. Also, studies of nonsupervisory 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. Application of these criteria resulted in a final sample of 95 studies and 99 independent effect sizes. Citations of studies considered but excluded from the meta-analysis are available as online supplemental materials.

Coding the Studies

Each of the studies included in the analyses was coded with respect to the moderator variables described above as well as general study characteristics. A thorough coding manual including instructions for coding articles and abstracting appropriate effect sizes was developed in order to aid in the coding process. All studies were coded by at least two researchers. The researchers

agreed on 89.5% of the initial codes, and disagreements were resolved via discussion. Additionally, reliability estimates (alpha) of the effectiveness measure from each study were recorded. Finally, all relevant information was coded to aid in the calculation of the standardized mean difference effect size.

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). 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 value of the male and female group standard deviations. If means and standard deviations were not available, the effect size was computed from other statistics, such as *t*, *F*, or *r*, according to formulas provided by Lipsey and Wilson (2001).

To reduce computational error, we calculated effect sizes with the aid of a computer program (Borenstein et al., 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 & Marin-Martinez, 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). Recommended transformation calculations were used 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).

Moderator Analyses

Multiple methods—the chi-square-based *Q* statistic and the 75% rule—were used to assess for the need to test for moderators. Categorical moderator variables were examined with subgroups analyses, and continuous moderators (percentage of male raters and time of study) were examined with meta-analytic regression analyses. Calculating the categorical models results in the between-class goodness-of-fit statistic Q_b , which is equivalent to a main effect in an analysis of variance and indicates whether the categorical moderator fully explains variance in the data (Cortina, 2003).

Results

In total, 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 ($SD = 9.72$). These studies were conducted

between 1962 and 2011. Thirty-six of the 99 studies included in the present meta-analysis were also included in the Eagly et al. (1995) meta-analysis, representing a 37% overlap.

Magnitude of Gender Differences

The distribution of effect sizes was approximately normal and centered around zero. The overall analysis of effectiveness measures resulted in a mean corrected d of $-.05$ ($K = 99$, $N = 101,676$), which is not significantly different from zero (see Table 1). We examined the data for any extreme outliers (± 3 SD) and found two effect sizes that met this criteria ($d = 1.44$, $N = 30$ and $d = 1.52$, $N = 40$). Hunter and Schmidt (2004) argued that, when sample sizes of outliers are small to moderate, extreme outliers can occur due to sampling error. They noted that such outliers should not be removed from the data, because removing them could result in an overcorrection of sampling error. We reanalyzed the data with these two effect sizes removed from the sample, 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, they were not eliminated from the data. The Q test of homogeneity for the summary effect indicated that moderation is likely (i.e., $Q = 415.3$, $p < .01$), suggesting that there is substantial variation in estimated population values and that in some cases males are more effective (positive values) and that in some cases females are more effective (negative values).

Moderator Analyses

Publication date as a moderator. To test Hypothesis 1, that there will be greater gender differences (favoring men) seen among older studies and smaller gender differences or differences that favor women seen among newer studies, we utilized both subgroup analyses and meta-analytic regression techniques. Theoretically meaningful subgroups were created with data on the percent of women in management in the United States from 1960 to the present day (Catalyst, 2012a). The subgroup 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 occupied close to 25% or less of management positions. The approximate percentages of women in management per each time-based subgroup can be seen in Table 1.

The time categories exhibited a nonsignificant moderating effect on gender differences in overall leadership effectiveness ($Q_b = 3.349$, $p = .34$). Although the effects for each time period were not significantly different from zero, the direction of effect sizes indicates that men were seen as more effective leaders in the oldest group of studies and that women were seen as more effective leaders in subsequent years.

We used weighted least squares analysis (Neter, Wasserman, & Kutner, 1989) in SPSS/PASW 18 to examine the effect of the continuous variable of time on overall gender differences in leadership effectiveness following Steel and Kammeyer-Mueller

(2002) and Geyskens, Krishnan, Steenkamp, and Cunha (2009), who found it the most accurate method. The weights were the same as used in the subgroup meta-analyses, the inverse variance of each effect size. The unstandardized beta term for time was not statistically significant ($\beta = .001$, $p > .05$, $R^2 = .005$); however, the pattern of the effect was consistent with Hypothesis 1 (see Figure 1). Overall, Hypothesis 1 was not supported, although results were in the predicted direction such that male leaders are seen as more effective in older studies and female leaders are seen as more effective in newer studies.

Organization as a moderator. To test Hypothesis 2, regarding the extent to which the organization is male dominated influences gender differences in leadership effectiveness, we used secondary data from the U.S. Bureau of Labor Statistics (2011) as well as the "Statistics on Women in the Military" report (Women in Military Service for America Memorial Foundation, 2011). Heilman (1983) suggested that one way to conceptualize the masculinity or femininity of an organization is by the percentage of men and women occupying that organization. The U.S. Bureau of Labor Statistics database reports the average percent of men and women in different types of organizations, and the "Statistics on Women in the Military" report includes the percent of women in the U.S. military. Organization exhibited a significant moderating effect on gender differences in leadership effectiveness ($Q_b = 11.72$, $p < .05$), providing preliminary support of Hypothesis 2.

Consistent with RCT (Eagly & Karau, 2002), organizations that are more masculine in nature and male dominated numerically tended to show that male leaders are more effective (see Table 1). Government organizations (37.3% female) exhibited a significant effect, $d = .27$ ($K = 5$, $N = 1,113$, 95% CI [.02, .51]). The other types of masculine organizations exhibited nonsignificant differences; yet, the pattern of effects generally supports the RCT hypothesis. The effect size for military organizations was positive (14.5% female), $d = .12$ ($K = 6$, $N = 2,505$, 95% CI [-.09, .32]). In addition, as expected based on RCT, organizations that are more feminine and female dominated had negative effect sizes, indicating that women were seen as more effective than men. The direction of the (nonsignificant) effect sizes indicated that females were seen as more effective in social service organizations (85% female), with a d of $-.23$ ($K = 2$, $N = 369$, 95% CI [-.58, .13]), and as slightly more effective in education organizations (68.4% female), with a d of $-.03$ ($K = 36$, $N = 4,051$, 95% CI [-.13, .06]). The former effect should be interpreted cautiously given the small number of studies. Overall, Hypothesis 2 was partially supported in that men were seen as more effective in male-dominated organizations such as the government. There were nonsignificant gender differences in the female-dominated organizations. Of interest, in business settings, which are 42.5% female, women were seen as significantly more effective than men, with a d of $-.12$ ($K = 25$, $N = 28,440$, 95% CI [-.19, -.02]).

Hierarchical level as a moderator. Consistent with Hypothesis 3, hierarchical level exhibited a significant moderating effect on gender differences in leadership effectiveness ($Q_b = 10.71$, $p < .05$). The results of a subgroup analysis are partially consistent with the hypothesis proposed by RCT (see Table 1). Women were rated as significantly more effective than men in middle manage-

Table 1
 Meta-Analysis of Overall Gender Differences and by Time of Study, Organization, Leader Level, and Setting

Variable	Mean <i>d</i>	Cor <i>d</i>	<i>K</i>	<i>N</i>	Var.	% Art.	95% CI	<i>Q</i>	<i>Q_b</i>
Overall leader effectiveness	-.047	-.05	99	101,676	.001	23.6	[-.102, .001]	415.3**	
Time of study								410.88**	3.35
1962–1981 (15.6%–26.2%)	.044	.046	23	4,363	.004	41.75	[-.08, .17]	52.70**	
Self	.395	.422	5	862	.046	15.38	[.01, .84]	26.01**	
Other	-.049	-.055	16	3,429	.004	79.43	[-.19, .08]	18.89	
1982–1995 (26.2%–38.7%)	-.053	-.055	25	5,651	.003	39.40	[-.16, .05]	60.91**	
Self	.247	.267	5	1,035	.045	46.06	[-.15, .68]	8.69**	
Other	-.125	-.134	20	4,616	.003	61.36	[-.24, -.03]	30.97*	
1996–2003 (38.7%–50.6%)	-.058	-.064	25	7,183	.003	19.76	[-.17, .04]	121.43**	
Self	.016	.022	5	2,131	.045	5.68	[-.40, .44]	70.42**	
Other	-.113	-.122	20	5,052	.003	66.99	[-.23, -.02]	28.36**	
2004–2011 (50.6%–51.4%)	-.091	-.097	26	84,478	.003	14.22	[-.20, .00]	175.83**	
Self	.087	.095	4	683	.054	95.0	[-.36, .55]	3.16**	
Other	-.128	-.137	22	83,795	.002	12.39	[-.23, -.05]	169.54**	
Organization								283.15**	11.72**
Social service (85% female)	-.192	-.225	2	369	.033	49.06	[-.58, .13]	2.04	
Self	-.070	-.086	1	304	.148	100.00	[-.84, .67]	0.00	
Other	-.490	-.527	1	65	.093	100.00	[-1.12, .07]	0.00	
Education (68.4% female)	-.030	-.033	36	4,051	.002	47.99	[-.13, .06]	72.92**	
Self	.333	.365	4	516	.044	82.16	[-.05, .78]	3.65	
Other	-.095	-.103	32	3,535	.002	73.47	[-.20, -.01]	42.20	
Business (42.5% female)	-.098	-.106	25	28,441	.002	23.66	[-.19, -.02]	101.45**	
Self	.160	.174	6	686	.03	29.34	[-.18, .53]	17.04**	
Other	-.150	-.162	19	27,754	.002	26.72	[-.24, -.08]	67.37**	
Government (37.3% female)	.253	.265	5	1,113	.016	16.73	[.02, .51]	23.91**	
Self	.302	.324	2	1,002	.072	6.03	[-.20, .85]	16.59**	
Other	.090	-.095	3	111	.060	100.00	[-.58, .39]	1.02	
Military (14.6% female)	.108	.117	6	2,505	.011	11.43	[-.09, .32]	43.75**	
Self	.315	.341	3	467	.065	8.40	[-.16, .84]	23.82**	
Other	.065	.064	3	2,038	.013	10.07	[-.16, .29]	19.87**	
Mixed	-.076	-.08	25	65,197	.003	61.41	[-.18, .02]	39.08	
Self	-.004	-.005	3	1,736	.048	100.00	[-.44, .43]	1.58	
Other	-.094	-.099	20	63,389	.003	43.55	[-.20, .00]	52.83	
Level of leadership								377.72**	10.71**
Lower level supervisors	.066	.069	37	7,421	.002	40.99	[-.03, .17]	87.81**	
Self	.345	.375	8	1,330	.019	48.26	[.11, .64]	14.51*	
Other	-.028	-.032	27	6,019	.003	50.89	[-.13, .07]	51.09**	
Middle managers	-.163	-.172	12	4,570	.005	21.13	[-.31, -.03]	52.07**	
Self	-.096	-.097	5	839	.032	51.06	[-.45, .25]	7.83	
Other	-.223	-.235	7	3,731	.005	16.74	[-.38, -.10]	35.83**	
Upper level leaders	-.038	-.042	28	12,364	.003	24.32	[-.15, .07]	111.02**	
Self	.329	.355	3	1,192	.040	9.28	[-.04, .75]	21.54**	
Other	-.123	-.133	25	11,172	.003	87.44	[-.24, -.03]	27.45	
Mixed	-.116	-.123	22	77,321	.003	16.56	[-.23, -.02]	126.82**	
Self	.052	.052	3	1,350	.049	11.57	[-.38, .49]	17.29**	
Other	-.122	-.130	19	75,971	.002	16.89	[-.22, -.04]	16.89**	
Setting								412.04**	1.73
Laboratory	-.021	-.023	13	1,425	.010	100.00	[-.20, .16]	11.49	
Self	.050	.054	1	550	.159	100.00	[-.73, .84]	0.00	
Other	-.062	-.064	10	803	.010	84.90	[-.27, .14]	10.60	
Organizational	-.043	-.046	82	99,769	.000	20.36	[-.10, .01]	397.88**	
Self	.199	.216	18	4,161	.011	15.62	[.02, .42]	108.93**	
Other	-.110	-.119	64	95,608	.000	24.87	[-.17, -.06]	253.28**	
Self-ratings vs. other-ratings								415.32**	25.41**
Self-ratings	.181	.206	19	4,711	.009	16.35	[.09, .31]	110.07**	
Other-ratings	-.100	-.120	78	96,893	.001	28.56	[-.18, -.06]	269.62**	
Raters								273.67**	24.46**
Boss	-.159	-.172	9	13,273	.005	13.53	[-.32, -.03]	59.13**	
Peers	.015	.017	2	58	.108	91.76	[-.63, .66]	1.09	
Subordinates	-.077	-.082	32	63,450	.003	70.80	[-.19, .02]	43.79*	
Self	.181	.206	19	4,711	.009	16.35	[.09, .31]	110.07**	
Judges/trained observers	-.167	-.177	5	882	.015	100.00	[-.42, .06]	0.26	
Objective counting device	.137	.147	2	72	.082	100.00	[-.41, .71]	0.07	
Mixed/unclear	-.109	-.119	30	19,229	.002	48.93	[-.21, -.03]	59.27**	

Note. The self and other ratings may not add up to equal the general category statistics because these findings include ratings from additional sources beyond self and other ratings (i.e., objective ratings). The bold font indicates that the 95% CI does not include zero. 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; *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 United States (Catalyst, 2012a).

* $p < .05$. ** $p < .01$.

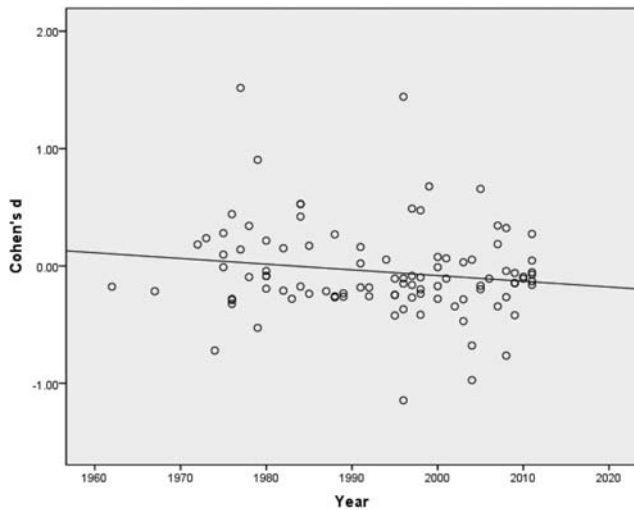


Figure 1. Scatterplot of time of study and effect sizes.

ment positions, with a d of $-.17$ ($K = 12$, $N = 4,570$, 95% CI $[-.31, -.03]$). There was a nonsignificant gender difference in effectiveness for leaders in upper level leadership positions, with a d of $-.04$ ($K = 28$, $N = 12,364$, 95% CI $[-.15, .07]$), and in lower hierarchical levels/supervisor positions, with a d of $.07$ ($K = 37$, $N = 7,421$, 95% CI $[-.03, .17]$). Overall, Hypothesis 3 was partially supported in that women were more effective in middle management positions, although there were not gender differences in either lower or higher level positions.

Setting as a moderator. Study setting exhibited a nonsignificant moderating effect on gender differences in leadership effectiveness ($Q_b = 1.73$, $p = .42$; see Table 1). Thus, Hypothesis 4 was not supported when all rating sources combined were examined. The effects of self- and other-ratings are reported separately in Table 1 and summarized in Table 2, providing additional information about this moderator (as discussed below).

Percent of male raters as a moderator. Meta-analytic weighted least squares regression analyses were used to exam-

ine the centered continuous variable. Upon viewing the scatterplot of percent of male raters on Cohen's d , we identified extreme outliers at the $d = 1.52$ and $d = .43$ points. Although these were not removed for the complete meta-analysis, they were removed for this specific analysis given the degree to which they differed from the normal distribution for the 58 studies that included the variable of percent of male raters. The linear effect for percent of male raters was significant (unstandardized $\beta = -.003$, $R^2 = .09$, $p < .01$), while the squared term (testing the curvilinear effect) was not statistically significant.

Inconsistent with our RCT hypothesis developed based on the concept of tokenism being helpful for high-status individuals, when rated by a majority of female raters, female leaders were seen as more effective than men. Consistent with our hypothesis, this difference became smaller in gender-balanced groups (see Figure 2). Surprisingly, in male-dominated rater groups, men were not seen as more effective leaders than women. Rather, the effect approached zero as the percent of male raters increased. Thus, Hypothesis 5 was partially supported (i.e., gender differences were quite small in gender-balanced groups).

Rating source as a moderator. When we analyzed the summary effect using self-ratings of leadership effectiveness only, men rated themselves as significantly more effective than women rated themselves ($d = .21$, $n = 19$, 95% CI $[.09, .31]$). This significant effect is driven by a large effect size favoring men in self-ratings from before 1996 (see Table 1). When we analyzed other-ratings of leadership effectiveness (from peers, subordinates, bosses, judges/trained observers, and/or mixed raters), women were rated as significantly more effective than men (particularly after 1982; $d = -.12$, $n = 78$, 95% CI $[-.18, -.06]$; see Table 1). These findings support Hypothesis 6a. To test Hypothesis 6b, we examined the effects of each moderator separately for the self-ratings and other-ratings. Overall, the moderators affected gender differences in leadership effectiveness in both self-ratings and other-ratings, providing initial support for Hypothesis 6b. The full results per moderator group are reported in Table 1 and summarized in Table 2.

Table 2
Summary of Findings Based on Rating Source

Variable	Self-ratings	Other-ratings	Combined ratings
Overall (summary effect size)	Men rated themselves as significantly more effective than women.	Women were rated as significantly more effective than men.	Overall, there was not a significant gender difference.
Time of study	Men rated themselves as significantly more effective than women rated themselves prior to 1982.	Women were rated as significantly more effective than men between 1982 and 2011.	There were no significant differences across time periods.
Organization type	There were no significant differences in self-ratings of effectiveness across organization type.	Women were rated as significantly more effective than men in business and educational organizations.	Women were rated as significantly more effective than men in business; men were rated as more effective in government organizations.
Leadership level	Men rated themselves as significantly more effective than women rated themselves in lower level positions.	Women were rated as significantly more effective than men in mid-level and upper level positions.	Women were rated as significantly more effective than men in mid-level positions.
Study setting	Men rated themselves as significantly more effective than women rated themselves in organizational studies.	Women were rated as significantly more effective than men in organizational settings; there were no differences in laboratory settings.	There were no significant differences across study settings.

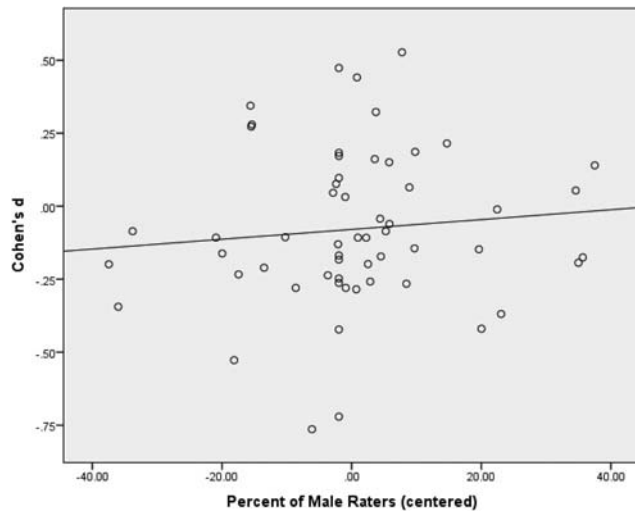


Figure 2. Scatterplot of percent of male raters and effect sizes.

Discussion

First and foremost our goal in the present study was to update and extend our understanding of the relationship between gender and leadership effectiveness. Toward this end, we quantitatively review 49 years of research pertaining to the relationship between gender and leadership effectiveness as well as to a number of theoretically based moderators of this relationship. In doing so, we update and extend Eagly et al.'s (1995) meta-analysis of gender differences in leadership effectiveness. Moreover, we expand role congruity theory (RCT) by reframing it such that it applies regardless of gender (i.e., to both men and women). Finally, we highlight the importance of rating source in any examination of gender advantages in leadership effectiveness (i.e., the importance of examining self-reported and other-reported assessments of leadership effectiveness; Eagly & Carli, 2003a, 2003b; Vecchio, 2002, 2003).

With respect to the overall impact of gender on leader effectiveness, the literature to date seems to oversimplify gender advantages in leadership. Our examination of potential moderators of this relationship suggests that it may not be so simple. Rather, our findings indicate a number of factors that moderate this relationship. One of our key findings is that very different patterns of results occur depending on whether self- or other-ratings serve as the measure of leader effectiveness. Moreover, results are further moderated by a number of other contextual variables. These results (quantitatively presented in Table 1) are summarized in Table 2. Below, we briefly discuss these findings with respect to implications for RCT, consistency with the Eagly et al. (1995) meta-analysis, and implications for gender-focused research and practice.

Consistent with RCT, the extent to which the organization being examined was male or female dominated significantly moderated gender differences in effectiveness. The direction of effects overall indicated that organizations that were male dominated (i.e., government) showed a tendency for men to be perceived as more effective. These findings are consistent with the 1995 meta-analysis of gender differences in leadership effectiveness (Eagly et

al., 1995). However, the magnitude of the effects seems to have waned over time. The Eagly et al. study showed a d of .42, 95% CI [.32, .52], for military studies, yet in the current meta-analysis, $d = .12$, 95% CI [-.09, .32], for this group.

When examining the effects of type of organization separately by rating source, we found that for other-ratings, women were rated as significantly more effective leaders than men in business and education organizations. Alternately, self-ratings were not significantly different from zero. These findings highlight how perceptions of incongruity may negatively impact men, as well as women, in certain situations. To date, however, RCT has primarily been used to explain the impact of incongruity negatively affecting women in leadership roles. Yet, our findings show that certain leadership roles (i.e., education and business) may also be seen as incongruent with men's gender role, negatively affecting perceptions of their effectiveness.

In addition, consistent with the hypothesis proposed by Eagly and Karau (2002), women are seen as more effective than men in middle management positions ($d = -.17$, 95% CI [-.31, -.03]). These findings are comparable to findings of the 1995 meta-analysis, which found that women were seen as more effective in mid-level leadership ($d = -.18$, 95% CI [-.24, -.12]). We also found a nonsignificant difference favoring men in lower level positions ($d = .07$, 95% CI [-.03, .17]), which is considerably smaller than the effect found by Eagly et al. (1995) for first-level or line leadership positions ($d = .19$, 95% CI [.13, .26]). Perhaps more important, we found this effect was significantly moderated by rating source. For self-ratings, men rated themselves as significantly more effective than women rated themselves in lower level supervisor positions and senior leader positions. There were nonsignificant gender differences in self-ratings for middle management positions. For other-ratings of leadership effectiveness, women were rated as significantly more effective as middle managers and as senior leaders. There was a nonsignificant gender difference in other-ratings of leadership effectiveness for lower level supervisors.

As noted above, our study provides preliminary support for RCT as it relates to men being seen as less effective leaders than women in middle management positions, possibly because of those positions being considered to be feminine in nature. We also found a nonsignificant gender difference in effectiveness for lower level positions, perhaps due to the gender-neutral characteristics associated with such positions. Yet, contrary to hypotheses proposed by RCT, we found that women were seen as more effective (by others) when they held senior-level management positions. Foschi's (2000) double standards of competence model proposes that women may be seen as more effective than men in top leadership positions due to perceptions of their extra competence. This notion was supported in a recent laboratory study by Rosette and Tost (2010). Thus, RCT may be supplemented by this model to explain that perceptions of extra competence can potentially override perceptions on women's incongruity, increasing assessments of their effectiveness at the top.

Similarly, the role of study setting was also moderated by rating source. That is, despite our finding of no significant impact of study setting overall, we did find significant differences when we broke down the analysis by rating source. We expected that men would be seen as more effective leaders than

women when comparing results from organizational data to those of laboratory data, due to the effects of cognitive load in increasing the use of think manager–think male stereotypes in organizations. However, we found that, in 64 studies, women were rated (by others) as significantly more effective than men in organizational settings, while the gender difference was not significantly different in 10 laboratory studies. Thus, perhaps the notion that cognitive load leads to the increased use of stereotypes holds true; however, rather than relying on stereotypes of men’s greater leadership effectiveness, organizational members may rely upon a different, newer stereotype: that women are more effective leaders.

In their meta-analysis, Eagly et al. (1995) found a significant linear effect for the percent of male subordinates on gender differences such that as the proportion of males increased, effectiveness ratings favored male leaders. We found, counter to Eagly et al.’s results, that gender differences in effectiveness approached zero as the percent of male raters increased. Additionally, we found that as the percent of female raters increased, female leaders were seen as more effective than men. This difference became smaller in gender-balanced groups. This finding is inconsistent with the suggestion that Eagly and Karau (2002) developed from tokenism (Kanter, 1977a) and research on the glass escalator effect (Maume, 1999). These authors suggest that when the percent of female raters increases (i.e., a male leader would be a high-status token), men should be seen as more congruent and thus as more effective leaders than women (Eagly & Karau, 2002). Rather, our study found that the percent of female raters positively affects ratings of women’s effectiveness. Again, these findings highlight the applicability of RCT to men as well as women depending on the context or situation. If stereotypes have shifted such that women may be considered to be better leaders than men, men may suffer consequences from being tokens in the same way women have historically suffered from numerical underrepresentation in a group. Additional research on the evolving effects of tokenism on men and women in work groups is needed to clarify the mechanisms of this possible reversal of effects.

Implications for Theory

A primary goal in this meta-analysis was to examine the extent to which the framework of contextual moderators proposed by RCT could be supplemented by additional theoretical perspectives proposed in the literature on gender and leadership. Overall, our findings indicate that certain aspects of RCT may need to be further clarified and explored. For instance, the RCT-based hypothesis regarding middle-level leadership positions was supported, yet at the highest levels, the direction of effects contradicted that proposed by RCT. The effect was small and negative when we examined other-ratings, providing support for the direction proposed by the double standards of competence model (Foschi, 2000) that women may be perceived to be slightly more effective than men at the highest levels. 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. More research should be conducted to better understand how

perceptions of competence, as well as expectations of leader behaviors that are needed for success at the top, affect perceptions of congruity and effectiveness.

Our findings with respect to rating source also have important implications for RCT. The theory seems to highlight how perceptions of incongruity from “other” evaluators can influence gender differences in leaders’ effectiveness. Yet, it is also important to consider how self-evaluations can be affected by perceptions of incongruity and congruity between leadership and gender roles. Our findings indicate that men may see themselves as congruent with most leadership contexts (even though other evaluators may disagree). Women, on the other hand, may see themselves as incongruent with many leadership contexts, even though others evaluate them as more effective as middle and senior managers, as well as in business and education organizations. More research should be conducted to examine how self-assessments of leadership effectiveness vary based on perceptions on incongruity and how self-perceptions can influence men and women’s career experiences and success.

Another aim in this study was to understand how the contextual moderators derived from RCT could negatively impact perceptions of men’s as well as women’s leadership effectiveness. We found that there are certain contexts in which there may be a greater perceived congruity between the female gender role and leadership roles (i.e., middle management, business and education organizations, settings with a high percent of female raters, and in organizational settings rather than laboratory settings), and, in these contexts, women are seen as more effective leaders than men. The findings regarding middle management and educational organizations fit with propositions based on RCT. However, the findings regarding the percent of female raters (tokenism for men) and organizational settings (settings with higher cognitive loads for participants) were unexpected.

Such findings highlight the shifting stereotypes surrounding gender and leadership (e.g., Koenig et al., 2011), which may be increasing perceptions of men’s incongruity (and ineffectiveness) in leadership positions. To the extent that organizations shift toward a more feminine and transformational outlook, women should experience reduced prejudice, while men may be seen as more incongruent with leadership roles. Yet, men on average continue to earn more and advance into higher managerial levels than women (Blau & Kahn, 2007; Catalyst, 2012a). Researchers argue that performance may matter less for female leader’s pay than male leader’s pay. Indeed, Kulich, Ryan, and Haslam (2007) found that company performance more strongly related positively to bonus pay for male CEOs than female CEOs, whereas perceptions of the leader’s charisma more strongly impacted bonus pay for women than men. Thus, future research needs to examine the other factors that may help to explain why women are seen as equally (or more) effective leaders than men, yet they are not being rewarded in the same ways. Additionally, future meta-analytic research could gather and examine studies that reflect leader ratings used for employment decisions (pay raises, promotions, etc.) versus those used for research purposes only in order understand whether a greater bias emerges when ratings are being used for employment decisions.

Limitations and Directions for Future Research

Although the number of studies included in this meta-analysis is relatively large ($K = 99$), our use of moderator analyses led to relatively small numbers and low power for some analyses. For instance, we found that the analysis of time periods was not statistically significant; however, the pattern of the effects did show support for the hypothesis proposed by RCT. We found that men were considered slightly more effective leaders in older studies, and women were favored in studies conducted more recently. This is consistent with a recent meta-analysis that found that leadership definitions were seen as less masculine over time (Koenig et al., 2011). A post hoc power analysis was conducted to determine if a low sample size (99 studies) could explain the nonsignificance of the effect. Consistent with this idea, the observed power was only .35. Nonetheless, Hunter and Schmidt (2004) have indicated that although a small number of studies in a meta-analysis may increase the likelihood of secondary sampling error, if the overall sample size (N) across those studies is fairly large, this should reduce the second order sampling error. Fortunately, in our meta-analysis, even the analyses with a low number of studies include fairly large overall sample sizes.

Many researchers overlook the importance of testing for 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. We found two instances 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%). 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 rating themselves as more effective than women rated themselves, or whether being in lower level positions was the more important moderator. Future research should examine the feminine and masculine behaviors needed for success at different hierarchical levels in order to better understand how level moderates gender differences in leadership effectiveness.

We also recognize that many of the effect sizes reported in this meta-analysis are fairly small. Though 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). Similarly, small biases against men and women in assessments of leadership effectiveness in certain contexts, when repeated over individuals and occasions, can produce large conse-

quences in terms of their ability to reach and succeed in leadership positions.

Finally, it is important to note that all but two of our significant findings (i.e., CIs that do not include zero) have significant Q values suggesting the presence of moderators. This in and of itself is an important finding. That is, despite our explicit examination of a variety of potential moderators, our results suggest that there may be others. This question certainly represents an avenue for future research.

Conclusion

This meta-analysis contributes to a recent debate in the literature regarding gender advantages in leadership effectiveness by showing that when all leadership contexts are considered together, there is a nonsignificant gender difference in leadership effectiveness. More important, this study answers a call in the literature to examine the influence of contextual moderators developed from role congruity theory (Eagly & Karau, 2002) and some additional theoretical frameworks.

References

References marked with an asterisk indicate studies included in the meta-analysis.

- Abelson, R. P. (1985). A variance explanation paradox: When a little is a lot. *Psychological Bulletin*, *97*, 129–133. doi:10.1037/0033-2909.97.1.129
- *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*, 181–201. doi:10.1080/13603124.2010.537448
- *Antonaros, M. E. (2010). *Gender differences in leadership style: A study of leader effectiveness in higher education* (Doctoral dissertation). University of Michigan, Ann Arbor.
- *Atwater, L., & Roush, P. (1994). An investigation of gender effects on followers' ratings of leaders, leaders' self-ratings and reactions to feedback. *Journal of Leadership Studies*, *1*, 37–52. doi:10.1177/107179199400100405
- *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*, 852–879. doi:10.1111/j.1559-1816.2009.00463.x
- Banks, C. G., & Murphy, K. R. (1985). Toward narrowing the research-practice gap in performance appraisal. *Personnel Psychology*, *38*, 335–345. doi:10.1111/j.1744-6570.1985.tb00551.x
- *Barker, C. J. (1982). *The effect of selected demographic variables on Hollander's theory of idiosyncrasy credit* (Doctoral dissertation). University of Maryland.
- *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.
- *Bartone, P. T., & Snook, S. A. (2002). Cognitive and personality predictors of leader performance in West Point cadets. *Military Psychology*, *14*, 321–338. doi:10.1207/S15327876MP1404_6
- *Bass, B. M., Avolio, B. J., & Atwater, L. (1996). The transformational and transactional leadership of men and women. *Journal of Applied Psychology*, *45*, 5–34. doi:10.1111/j.1464-0597.1996.tb00847.x
- Berger, J., Fisek, M. H., Norman, R. Z., & Zelditch, M. (1977). *Status characteristics and social interaction*. New York, NY: Elsevier Scientific.

- *Bess, M. M. (2001). *Perceived leadership effectiveness of male and female directors of schools in west and east Tennessee* (Doctoral dissertation). Tennessee State University.
- Beyer, S. (1990). Gender differences in the accuracy of self-evaluations of performance. *Journal of Personality and Social Psychology*, *54*, 5–12.
- Beyer, S. (1992, August). *Self-consistency and gender differences in the accuracy of self-evaluations*. Paper presented at the meeting of the American Psychological Association, Washington, DC.
- *Blanton, J. (1996). *A study of the relationship between leadership behavior components and demographic variables among middle school principals* (Doctoral dissertation). Texas Southern University.
- Blau, F. D., & Kahn, L. M. (2007). The gender pay gap: Have women gone as far as they can? *Academy of Management Perspectives*, *21*, 7–23. doi:10.5465/AMP.2007.24286161
- *Bolman, L. G., & Deal, T. E. (1991). *Images of leadership* (Occasional Paper No. 7). Cambridge, MA: National Center for Educational Leadership.
- *Bolman, L. G., & Deal, T. E. (1992). Leading and managing: Effects of context, culture, and gender. *Education Administration Quarterly*, *28*, 314–329. doi:10.1177/0013161X92028003005
- Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. Chichester, West Sussex, United Kingdom: Wiley.
- Brutus, S., Fleenor, J. W., & McCauley, C. D. (1999). Demographic and personality predictors of congruence in multi-source ratings. *Journal of Management Development*, *18*, 417–435. doi:10.1108/02621719910273569
- *Burke, S., & Collins, K. M. (2001). Gender differences in leadership styles and management skills. *Women in Management Review*, *16*, 244–257. doi:10.1108/09649420110395728
- *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.
- Bushman, B. J., & Anderson, C. A. (2001). Media violence and the American public: Scientific facts versus media misinformation. *American Psychologist*, *56*, 477–489. doi:10.1037/0003-066X.56.6-7.477
- *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*, 887–902. doi:10.1023/A:1018880706172
- Carroll, J. (2006). *Americans prefer male boss to a female boss*. Retrieved from <http://brain.gallup.com>
- *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.
- Catalyst. (2012a). *Statistical overview of women in the workplace*. Retrieved from <http://www.catalyst.org/publication/219/statistical-overview-of-women-in-the-workplace>
- Catalyst. (2012b). *U.S. women in business*. Retrieved from <http://www.catalyst.org/knowledge/us-women-business-0>
- Center for American Women and Politics. (2012a). *Fact sheet: Women in Elective Office*. Retrieved from http://www.cawp.rutgers.edu/fast_facts/levels_of_office/documents/elective.pdf
- Center for American Women and Politics. (2012b). *Fact sheet: Women in the U.S. Congress 2011*. Retrieved from http://www.cawp.rutgers.edu/fast_facts/levels_of_office/documents/cong.pdf
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
- Conlin, M. (2003, May 26). The new gender gap: From kindergarten to grad school, boys are becoming the second sex. *BusinessWeek*. Retrieved from http://www.businessweek.com/@roCLFYUQ7MXa*w8A/magazine/content/03_21/b3834001_MZ001.htm
- Correll, S. J. (2001). Gender and the career choice process: The role of biased self-assessments. *American Journal of Sociology*, *106*, 1691–1730. doi:10.1086/321299
- Cortina, J. M. (2003). Apples and oranges (and pears, oh my!): The search for moderators in meta-analysis. *Organizational Research Methods*, *6*, 415–439. doi:10.1177/1094428103257358
- *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.
- Crocker, J., & Major, B. (1989). Social stigma and self-esteem: The self-protective properties of stigma. *Psychological Review*, *96*, 608–630. doi:10.1037/0033-295X.96.4.608
- *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*, 173–186.
- *Davidson, B. J. (1995). *Leadership styles of successful male and female college choral directors* (Doctoral dissertation). Arizona State University, Tempe.
- *Davy, J., & Shipper, F. (2004). *A cross gender study of managerial skills, employees' attitudes and managerial performance*. Unpublished manuscript.
- *Day, D. R., & Stogdill, R. M. (1972). Leader behavior of male and female supervisors: A comparative study. *Personnel Psychology*, *25*, 353–360. doi:10.1111/j.1744-6570.1972.tb01110.x
- *Deaux, K. (1979). Self-evaluations of male and female managers. *Sex Roles*, *5*, 571–580. doi:10.1007/BF00287661
- *Douglas, C. (2012). The moderating role of leader and follower sex in dyads on the leadership behavior–leader effectiveness relationships. *Leadership Quarterly*, *23*, 1, 163–175. doi:10.1016/j.leaqua.2011.11.013
- *Douglas, C., & Ammeter, A. P. (2004). An examination of leader political skill and its effect on ratings of leader effectiveness. *Leadership Quarterly*, *15*, 571–580. doi:10.1016/j.leaqua.2004.05.006
- *Duehr, E. E. (2006). *Personality, gender, and transformational leadership: Investigating differential prediction for male and female leaders* (Doctoral dissertation). University of Minnesota.
- Eagly, A. H. (1987). *Sex differences in social behavior: A social-role interpretation*. Hillsdale, NJ: Erlbaum.
- Eagly, A. H., & Carli, L. L. (2003a). The female leadership advantage: An evaluation of the evidence. *Leadership Quarterly*, *14*, 807–834. doi:10.1016/j.leaqua.2003.09.004
- Eagly, A. H., & Carli, L. L. (2003b). Finding gender advantage and disadvantage: Systematic research integration is the solution. *Leadership Quarterly*, *14*, 851–859. doi:10.1016/j.leaqua.2003.09.003
- Eagly, A. H., & Carli, L. L. (2007). *Through the labyrinth*. Boston, MA: Harvard Business School Press.
- Eagly, A. H., & Johnson, B. T. (1990). Gender and leadership style: A meta-analysis. *Psychological Bulletin*, *108*, 233–256. doi:10.1037/0033-2909.108.2.233
- Eagly, A. H., & Karau, S. J. (2002). Role congruity theory of prejudice toward female leaders. *Psychological Review*, *109*, 573–598. doi:10.1037/0033-295X.109.3.573
- Eagly, A. H., Karau, S. J., & Makhijani, M. G. (1995). Gender and the effectiveness of leaders: A meta-analysis. *Psychological Bulletin*, *117*, 125–145. doi:10.1037/0033-2909.117.1.125
- Eagly, A., Makhijani, M., & Klonsky, B. (1992). Gender and the evaluation of leaders: A meta-analysis. *Psychological Bulletin*, *111*, 3–22. doi:10.1037/0033-2909.111.1.3
- Eagly, A. H., & Wood, W. (2012). Social role theory. In P. van Lange, A. Kruglanski, & E. T. Higgins (Eds.), *Handbook of theories in social psychology* (pp. 458–476). Thousand Oaks, CA: Sage.

- *Eckert, R., Ekelund, B. Z., Gentry, W. A., & Dawson, J. F. (2010). "I don't see me like you see me, but is that a problem?" Cultural influences on rating discrepancy in 360-degree feedback instruments. *European Journal of Work and Organizational Psychology, 19*, 259–278. doi:10.1080/13594320802678414
- *Elsesser, K. M., & Lever, J. (2011). Does gender bias against female leaders persist? Quantitative and qualitative data from a large-scale survey. *Human Relations, 64*, 1555–1578. doi:10.1177/0018726711424323
- *Eskilson, A. (1975). *Sex composition and leadership* (Doctoral dissertation). University of Illinois at Chicago.
- European Commission. (2012). *Women in economic decision-making in the EU: Progress report*. Retrieved from http://ec.europa.eu/justice/gender-equality/files/women-on-boards_en.pdf
- *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*, 135–144. doi:10.1177/014920638000600203
- Foschi, M. (1996). Double standards in the evaluation of men and women. *Social Psychology Quarterly, 59*, 237–254. doi:10.2307/2787021
- Foschi, M. (2000). Double standards for competence: Theory and research. *Annual Review of Sociology, 26*, 21–42. doi:10.1146/annurev.soc.26.1.21
- *Francke, C. A. (1975). *Perceived performance differences between women and men supervisors and implications for training* (Doctoral dissertation). Michigan State University.
- *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.
- Geis, F. L. (1993). Self-fulfilling prophecies: A social psychological view of gender. In A. E. Beall & R. J. Sternberg (Eds.), *The psychology of gender* (pp. 9–54). New York, NY: Guilford Press.
- 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. doi:10.1177/0149206308328501
- *Goktepe, J. R., & Schneider, C. R. (1988). Sex and gender effects in evaluating emergent leaders in small groups. *Sex Roles, 19*, 29–36. doi:10.1007/BF00292461
- *Gore, S. (2000). *Perceived leadership effectiveness of male and female directors of schools* (Doctoral dissertation). Tennessee State University.
- *Gross, N., & Trask, A. E. (1976). *The sex factor and the management of schools*. New York, NY: Wiley.
- *Gupta, N., Jenkins, G. D., & Beehr, T. A. (1983). Employee gender, gender similarity, and supervisor-subordinate cross-evaluation. *Psychology of Women Quarterly, 8*, 174–184. doi:10.1111/j.1471-6402.1984.tb00627.x
- 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. doi:10.1037/1082-989X.3.4.486
- Heilman, M. E. (1983). Sex bias in work settings: The lack of fit model. In B. Staw & L. Cummings (Eds.), *Research in organizational behavior* (Vol. 5, pp. 269–298). Greenwich, CT: JAI Press.
- 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. doi:10.1111/0022-4537.00234
- Helgesen, S. (1990). *The female advantage: Women's ways of leadership*. New York, NY: Doubleday.
- *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, NY: Teachers College, Columbia University.
- Hitt, M. A., Keats, B. W., & DeMarie, S. M. (1998). Navigating in the new competitive landscape: Building strategic flexibility and competitive advantage in the 21st century. *Academy of Management Executive, 12*, 22–42.
- *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*, 28–33.
- *Hua, P. C. (2005). *The relationship between gender, presidential leadership, and the quality of community colleges* (Doctoral dissertation). University of Southern California.
- Hunt, J. G., Boal, K. B., & Sorenson, R. L. (1990). Top management leadership: Inside the black box. *Leadership Quarterly, 1*, 41–65. doi:10.1016/1048-9843(90)90014-9
- 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.
- *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.
- Inter-Parliamentary Union. (2012). *Women in national parliaments*. Retrieved from <http://www.ipu.org/wmm-e/classif.htm>
- *Jacobson, M. B., & Effertz, J. (1974). Sex roles and leadership perceptions of the leaders and the led. *Organizational Behavior & Human Performance, 12*, 383–396. doi:10.1016/0030-5073(74)90059-2
- *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. doi:10.1016/j.obhdp.2007.12.002
- 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. doi:10.2307/800759
- *Kabacoff, R. I. (1998, August). *Differences in organizational leadership: A large sample study*. Paper presented at the meeting of the American Psychological Association, San Francisco, CA.
- Kanter, R. (1977a). *Men and women of the corporation*. New York, NY: Basic Books.
- 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. doi:10.1086/226425
- *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.
- Kling, K. C., Hyde, J. S., Showers, C. J., & Buswell, B. N. (1999). Gender differences in self-esteem: A meta-analysis. *Psychological Bulletin, 125*, 470–500. doi:10.1037/0033-2909.125.4.470
- *Knott, K. B., & Nattale, E. J. (1997). Sex differences, organizational level, and superiors' evaluation of managerial leadership. *Management Communication Quarterly, 10*, 523–540. doi:10.1177/0893318997104005
- 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*, 616–642. doi:10.1037/a0023557
- *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*, 355–363. doi:10.1007/BF00288308
- Kulich, C., Ryan, M. K., & Haslam, S. (2007). Where is the romance for women leaders? The effects of gender on leadership attributions and performance-based pay. *Applied Psychology, 56*, 582–601. doi:10.1111/j.1464-0597.2007.00305.x
- *Lauterbach, K. E., & Weiner, B. J. (1996). Dynamics of upward influence: How male and female managers get their way. *Leadership Quarterly, 7*, 87–107. doi:10.1016/S1048-9843(96)90036-3

- *Leithwood, K., & Jantzi, D. (1997). Explaining variation in teachers' perceptions of principals' leadership: A replication. *Journal of Educational Administration, 35*, 312–331. doi:10.1108/09578239710171910
- Lenny, E. (1977). Women's self-confidence in achievement settings. *Psychological Bulletin, 84*, 1–13.
- 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. doi:10.1037/a0021276
- Lipsey, M. W. (2003). Those confounded moderators in meta-analysis: Good, bad, and ugly. *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.
- Lord, R. G., & Maher, K. J. (1993). *Leadership and information processing: Linking perceptions and performance*. New York, NY: Routledge.
- *Love, A. (2007). *Teacher's perception of the leadership effectiveness of female and male principals* (Doctoral dissertation). Tennessee State University.
- 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*, 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. doi:10.1037/0022-3514.72.4.709
- Macrae, C. N., Hewstone, M., & Griffiths, R. J. (1993). Processing load and memory for stereotype-based information. *European Journal of Social Psychology, 23*, 77–87. doi:10.1002/ejsp.2420230107
- 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.
- Maume, D. J. (1999). Glass ceilings and glass escalators: Occupational segregation and race and sex differences in managerial promotions. *Work and Occupations, 26*, 483–509.
- 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.
- *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). Ohio State University.
- *Mohr, E. S., & Downey, R. G. (1977). Are women peers? *Journal of Occupational Psychology, 50*, 53–57. doi:10.1111/j.2044-8325.1977.tb00358.x
- *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. doi:10.1111/j.1467-8551.2007.00521.x
- *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., Jr., & Jensrud, Q. (1995). Gender, leadership, and vocational education. *Journal of Industrial Teacher Education, 33*, 6–23.
- Mumford, T. V., Campion, M. A., & Morgeson, F. P. (2007). The leadership skills strataplex: Leadership skill requirements across organizational levels. *Leadership Quarterly, 18*, 154–166. doi:10.1016/j.leaqua.2007.01.005
- *Munson, C. E. (1979). Evaluation of male and female supervisors. *Social Work, 24*, 104–110.
- Neter, J., Wasserman, W., & Kutner, M. (1989). *Applied linear regression models* (2nd ed.). Homewood, IL: Irwin.
- *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.
- *Osborn, R. N., & Vicars, W. M. (1976). Sex stereotypes: An artifact in leader behavior and subordinate satisfaction analysis? *Academy of Management Journal, 19*, 439–449. doi:10.2307/255609
- Paolillo, J. G. (1981). Manager's self assessments of managerial roles: The influence of hierarchical level. *Journal of Management, 7*, 43–52. doi:10.1177/014920638100700104
- Parsons, J. E., Meece, J. L., Adler, T. F., & Kaczala, C. M. (1982). Sex differences in attributions and learned helplessness. *Sex Roles, 8*, 421–432. doi:10.1007/BF00287281
- *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.
- *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.
- Pavett, C. M., & Lau, A. W. (1983). Managerial work: The influence of hierarchical level and functional specialty. *Academy of Management Journal, 26*, 170–177. doi:10.2307/256144
- 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. doi:10.1006/jesp.1994.1015
- *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*, 349–352. doi:10.1037/0021-9010.69.2.349
- *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*, 624–628. doi:10.1037/0021-9010.60.5.624
- Powell, G. N., Butterfield, D. A., & Parent, J. D. (2002). Gender and managerial stereotypes: Have the times changed? *Journal of Management, 28*, 177–193. doi:10.1177/014920630202800203
- *Pratch, L., & Jacobowitz, J. (1996). Gender, motivation, and coping in the evaluation of leadership effectiveness. *Consulting Psychology Journal: Practice and Research, 48*, 203–220. doi:10.1037/1061-4087.48.4.203
- *Quinn, K. I. (1976). *Self-perceptions of leadership behaviors and decision making orientations of men and women elementary school principals in Chicago public schools* (Doctoral dissertation). University of Illinois at Urbana-Champaign.
- *Ragins, B. R. (1989). Power and gender congruency effects in evaluations of male and female managers. *Journal of Management, 15*, 65–76. doi:10.1177/014920638901500106
- *Ragins, B. R. (1991). Gender effects in subordinate evaluations of leaders: Real or artifact? *Journal of Organizational Behavior, 12*, 259–268. doi:10.1002/job.4030120308
- *Randall, M. J. (1998). *United States Air Force Nurse Captains' perceived leadership effectiveness* (Master's thesis). Arizona State University.
- *Renwick, P. A. (1977). The effects of sex differences on the perception and management of superior-subordinate conflict: An exploratory study. *Organizational Behavior & Human Performance, 19*, 403–415. doi:10.1016/0030-5073(77)90073-3
- *Reyes, R. I. (1997). *Leadership effectiveness of department chairpersons in the University of Los Andes* (Doctoral dissertation). University of South Florida.
- *Rice, R. W., Bender, L. R., & Vitters, A. G. (1980). Leader sex, follower attitudes toward women, and leadership effectiveness: A laboratory experiment. *Organizational Behavior & Human Performance, 25*, 46–78. doi:10.1016/0030-5073(80)90025-2

- *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*, 885–901. doi:10.1007/BF00288512
- Ridgeway, C. L. (1997). Interaction and the conservation of gender inequality: Considering employment. *American Sociological Review, 62*, 218–235. doi:10.2307/2657301
- Ridgeway, C. L. (2001). Gender, status, and leadership. *Journal of Social Issues, 57*, 637–655. doi:10.1111/0022-4537.00233
- *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*, 83–103. doi:10.1037/1089-2699.7.2.83
- *Rivero, A. J. (2003). *Sex and gender differences in actual and perceived leader effectiveness: Self and subordinate views* (Master's thesis). University of North Texas.
- *Rohmann, A., & Rowold, J. (2009). Gender and leadership style: A field study in different organizational contexts in Germany. *Equal Opportunities International, 28*, 545–560. doi:10.1108/02610150910996399
- Rosener, J. B. (1995). *America's competitive secret: Utilizing women as a management strategy*. New York, NY: Oxford University Press.
- Rosette, A. S., & Tost, L. P. (2010). Agentic women and communal leadership: How role prescriptions confer advantage to top women leaders. *Journal of Applied Psychology, 95*, 221–235. doi:10.1037/a0018204
- *Rosser, V. J. (2003). Faculty and staff members' perceptions of effectiveness leadership: Are there differences between women and leaders? *Equity & Excellence in Education, 36*, 71–81. doi:10.1080/10665680303501
- 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*, 719–739. doi:10.1177/0149206310374774
- Rothbart, M. (1981). Memory processes and social beliefs. In D. L. Hamilton (Ed.), *Cognitive processes in stereotyping and intergroup behavior* (pp. 145–181). Hillsdale, NJ: Erlbaum.
- Rustad, M. (1982). *Women in khaki: The American enlisted woman*. New York, NY: Praeger.
- 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. doi:10.1177/0013164498058002005
- Schein, V. E. (1973). The relationship between sex role stereotypes and requisite management characteristics. *Journal of Applied Psychology, 57*, 95–100. doi:10.1037/h0037128
- Schein, V. E. (2007). Women in management: Reflections and projections. *Women in Management Review, 22*, 6–18. doi:10.1108/09649420710726193
- *Schneier, C. E., & Bartol, K. M. (1980). Sex effects in emergent leadership. *Journal of Applied Psychology, 65*, 341–345. doi:10.1037/0021-9010.65.3.341
- *Schulman, J. M. (1989). *Power behavior, job satisfaction and leadership effectiveness in public school principals* (Doctoral dissertation). University of Bridgeport.
- *Simpson, M. W. (1985). *Teacher satisfaction with leadership based on principal sex role orientation* (Doctoral dissertation). University of Georgia.
- *Singh, S. K. (2007). Emotional intelligence and organizational leadership: A gender study in Indian context. *International Journal of Indian Culture and Business Management, 1*, 48–62. doi:10.1504/IJICBM.2007.014470
- *Skottheim, K. (2008). DI in relation to team and leaders evaluations. In B. Z. Ekelund & E. Langvik (Eds.), *Diversity Icebreaker: How to manage diversity processes*. Oslo, Norway: Human Factors.
- Smith, C. (2009, July 26). No doubts: Women are better managers. *The New York Times*. Retrieved from <http://www.nytimes.com>
- *Sparks, T. E., & Kuhnert, K. W. (2011). *Navigating the leadership labyrinth: Perceived outcomes for men and women*. Poster session presented at the meeting of the Society of Industrial and Organizational Psychology, Atlanta, GA.
- Steel, P. D., & Kammeyer-Mueller, J. D. (2002). Comparing meta-analytic moderator estimation techniques under realistic conditions. *Journal of Applied Psychology, 87*, 96–111. doi:10.1037/0021-9010.87.1.96
- *Stephens, H. E. (2004). *Gender-role orientation and leadership effectiveness* (Master's thesis). University of Guelph.
- *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.
- *Thompson, M. D. (2000). Gender, leadership orientation, and effectiveness: Testing the theoretical models of Bolman & Deal and Quinn. *Sex Roles, 42*, 969–992. doi:10.1023/A:1007032500072
- *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*, 285–290. doi:10.1108/01437739910292599
- United States Government Accountability Office. (2010). *Women in management: Female managers' representation, characteristics, and pay*. Retrieved from <http://www.gao.gov/new.items/d101064t.pdf>
- U.S. Bureau of Labor Statistics. (2011). *Household data annual averages: 11. Employed persons by detailed occupation, sex, race, and Hispanic or Latino ethnicity*. Retrieved December 10, 2012, from <http://www.bls.gov/cps/cpsaat11.pdf>
- U.S. Department of Labor. (2012). *Employment status of the civilian population by sex and age*. Retrieved from <http://www.bls.gov/news.release/empsit.t01.htm>
- Vecchio, R. P. (2002). Leadership and gender advantage. *Leadership Quarterly, 13*, 643–671. doi:10.1016/S1048-9843(02)00156-X
- Vecchio, R. P. (2003). In search of gender advantage. *Leadership Quarterly, 14*, 835–850. doi:10.1016/j.leaqua.2003.09.005
- *Vecchio, R. P., & Anderson, R. J. (2009). Agreement in self–other ratings of leader effectiveness: The role of demographics and personality. *International Journal of Selection and Assessment, 17*, 165–179. doi:10.1111/j.1468-2389.2009.00460.x
- *Vilkinas, T. (2000). The gender factor in management: How significant others perceive effectiveness. *Women in Management Review, 15*, 261–271. doi:10.1108/09649420010372922
- *Vilkinas, T., Shen, J., & Cartan, G. (2009). Predictors of leadership effectiveness for Chinese managers. *Leadership & Organization Development Journal, 30*, 577–590. doi:10.1108/01437730910981944
- Volberda, H. W. (1998). *Building the flexible firm: How to remain competitive*. New York, NY: Oxford University Press.
- Weber, R., & Crocker, J. (1983). Cognitive processes in the revision of stereotypic beliefs. *Journal of Personality and Social Psychology, 45*, 961–977. doi:10.1037/0022-3514.45.5.961
- *Wexley, K. N., & Pulakos, E. D. (1982). Sex effects on performance ratings in manager–subordinate dyads: A field study. *Journal of Applied Psychology, 67*, 433–439. doi:10.1037/0021-9010.67.4.433
- *White, D. R. (2007). *The relationship of leadership effectiveness to gender in Louisiana middle school principals* (Doctoral dissertation). University of Louisiana at Monroe.
- Williams, C. (1995). *Still a man's world*. Berkeley, CA: University of California Press.
- Williams, R. (2012, December 15). Why women may be better leaders than men. *Psychology Today*. Retrieved from <http://www.psychologytoday.com/blog/wired-success/201212/why-women-may-be-better-leaders-men>
- Women in Military Service for America Memorial Foundation. (2011). *Statistics on women in the military*. Retrieved from <http://www.womensmemorial.org/PDFs/StatsonWIM.pdf>

Women in top jobs are viewed as “better leaders” than men. (2010, May 14). *Daily Mail*. Retrieved from <http://www.dailymail.co.uk/sciencetech/article-1278009/Women-jobs-viewed-better-leaders-men.html#ixzz1quPyGPa>

*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.

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

*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.

Received January 24, 2013

Revision received March 18, 2014

Accepted March 25, 2014 ■