



‘Competent’ or ‘Considerate’? The Persistence of Gender Bias in Evaluation of Leaders¹

■ **Salin Denise²**

Professor, Hanken School of Economics, Department of Management and Organization, Finland

ABSTRACT

The aim of this article is to analyze the possible persistence of gender bias in the evaluation of leaders in Finland. Findings are based on two different studies. The first study confirmed that the perceived effectiveness and likeability ratings of fictive leaders ($n = 358$) varied as a function of leader gender. The second study, based on qualitative content analysis of subordinates' descriptions ($n = 119$) of good and poor leaders, pointed to gendered differences in the dimensions that subordinates paid attention: female leaders were both more likely than men to be praised for having, and criticized for not having, communal traits, whereas men were more likely than women to be judged on their expertise. As Finland has consistently been rated one of the most gender-equal countries in the world, these findings can be seen as particularly strong evidence of the persistence of gender bias in evaluations and of ongoing gendering of leadership.

KEYWORDS

Gender bias / gender role congruity / gender stereotypes / leadership / performance evaluations

‘Competent’ or ‘Considerate’? The Persistence of Gender Bias in Evaluation of Leaders

Research continues to provide evidence for the existence of subtle barriers that make it difficult for women to attain leadership positions in the same proportions as men (e.g. Carli & Eagly 2016). Different metaphors have been used to describe these barriers, including a glass ceiling and a leadership labyrinth, the latter drawing attention to the fact that it is not about a single obstacle but multiple hurdles along the way (Eagly & Carli 2007). Overall, both organizations themselves and leadership have been described as gendered (e.g. Acker 1991; Collinson & Hearn 1994) and recent studies suggest ‘think manager, think male’-assumptions continue to affect managerial recruitments (Holgerson 2013). The aim of this article is to study how gender affects evaluation of leaders. The article thus contributes to our understanding of gender bias in evaluations of leaders as a possible obstacle for women.

Earlier research suggests that there may be different performance standards for men and women and gender bias in evaluations (Heilman & Parks-Stamm 2007; Robertson et al. 2007). Several researchers point to the importance of gender stereotypes in the evaluation of men's and women's performance and of male and female leaders more specifically (Eagly & Karau 2002; Heilman 2001). While several studies have found

¹ You can find this text and its DOI at <https://tidsskrift.dk/njwls/index>.

² Corresponding author: denise.salin@hanken.fi.



support for the existence of gender bias in evaluations (Eagly et al. 1992; Moss-Racusin et al. 2012; Rudman & Glick 2001; Wennerås & Wold 1997), it remains unclear to what extent such biases are a thing of the past or still in place.

A recent meta-analysis revealed that women receive less favorable ratings than men in hiring and promotion situations (Koch et al. 2015). Their findings were based on experimental simulations and found that this tendency was particularly strong among male raters making decisions about male gender-typed jobs, that is, jobs traditionally perceived as appropriate for males. On the other hand, in another recent meta-analysis, Paustian-Underdahl et al. (2014) found that female leaders were rated as significantly more effective than male leaders in organizational settings. Furthermore, Koenig et al. (2011) found that leadership was less strongly associated with masculinity over time.

This article seeks to examine the possible persistence of gender bias in evaluations of men and women in a country ranked as one of the most gender equal in the world, that is, Finland (World Economic Forum 2018). The study is motivated by contradictory findings in the existing research on the persistence of gender bias, and because the public discourse in Finland is often characterized by an assumption that gender equality has already been achieved (Korvajärvi 2002).

This study makes several contributions to the existing research on gender bias in leader evaluation. First, it finds support for the continued persistence of such biases in one of the most gender-equal countries in the world, and in a sample of young, well-educated, urban participants, often associated with more progressive attitudes. Second, the article finds support not only for the assumption that men and women are rated differently for the same behavior but also that raters still pay attention to different aspects of behavior when evaluating female compared to male leaders. A specific strength of the study is that it employs both fictive examples, allowing leader behavior and performance to be held constant while manipulating leader gender, and qualitative evaluations of real-world examples of leaders perceived to be good and less so.

Gender Bias in Evaluation

Leadership as gendered

Although gender has often been invisible and neglected in discussions about leadership, traditionally there has been a persistent and pervasive assumption that 'leadership is synonymous with men' and an implicit conflation of men and masculinities with management and authority (Collinson & Hearn 1994; 4). Already back in the 1970s, Schein (1973; 1975) drew attention to the 'think leader, think male' phenomenon, by showing that people's images of successful managers were much more congruent with descriptions of men than of women. Moreover, the role congruity theory of prejudice against female leaders (Eagly & Karau 2002) suggests that the perceived incongruity between the female gender role and leadership roles means women are evaluated less favorably than men for and in leadership positions.

This is also in line with the assertion that organizations themselves are gendered rather than gender-neutral arenas (cf. Acker 1991). Collinson and Hearn (1994) highlight how masculine values and assumptions are embedded in and permeate the structure, culture, and practices of organizations, but how this often happens in taken-for-granted

ways. The gendered nature of organizational life was also demonstrated by Kanter (1977), who showed how the uncertainty of managerial work created pressures for social certainty and conformity, ‘homosocial reproduction’ leading to a preference for male candidates.

Recent research from Sweden points to the ongoing gendering of leadership and how unreflexive practice and perceptions that women are deficient as managers still can lead to a preference for men for leadership positions (Holgersson 2013; Wahl 2010). By interviewing chairmen of the board and closely studying the recruitment process of three managing directors, Holgersson (2013) obtained detailed insight into how recruitment processes were carried out and how successful candidates were chosen. The results showed that throughout the recruitment process competence was often redefined in order to motivate the recruitment of a male candidate and thereby maintain the gender order, typically without the decision-makers reflecting upon this themselves.

This article seeks to contribute to the understanding of the gendering of leadership, by specifically focusing on gender bias in evaluations of male and female leaders. The article will start by reviewing existing evidence and then continue to examine whether these findings can be supported also by contemporary research evidence. In the next section of the article, the tendency for raters to be affected by ratee gender will be discussed in more detail. Empirical studies showing how the same behavior may elicit different evaluations and interpretations will be reviewed and presented. Subsequently, the role of gender stereotypes will be discussed. In particular, the section will highlight how descriptive and prescriptive stereotypes concerning men’s and women’s behavior may lead those making evaluations to hold different expectations of male and female leaders, thereby suggesting male and female leaders may be judged on different aspects.

Same Behavior – Different Evaluations

Research on gender bias in evaluations has deep roots, and a considerable body of empirical research confirms the significance of stereotypes in evaluations (Heilman 2001; Moss-Racusin et al. 2012; Rudman & Glick 2001; Steinpreis et al. 1999). Below, these findings will be presented in more detail.

Already, seminal work by Goldberg (1968) in the 1960s showed that identical texts were rated differently depending on whether readers were told they had been written by a man or by a woman. Texts assumed to be written by a man were rated more favorably, if they represented a male-typed field of expertise (Goldberg 1968; for later replications see also Paluda & Strayer 1985 and Knobloch-Westerwick et al. 2013). Similarly, a number of studies have confirmed that the CVs of fictive applicants are rated differently based on applicant gender, with females being rated lower on competence and therefore offered lower salaries (e.g., Moss-Racusin et al. 2012) or less likely to be hired (Steinpreis et al. 1999).

A recent meta-analysis of experimental simulations of employment decision-making confirmed that women receive less favorable ratings than men in hiring, promotion, and other job contexts (Koch et al. 2015). Student evaluations of their teachers appear to be no different: experiments at a North-American University showed that instructors teaching online courses were graded higher when using a male as opposed to a female name in communication with students (MacNell et al. 2015). Taking the exact same



amount of time to grade student work, professors assumed to be male got higher ratings for promptness. Similarly, Mitchell and Martin (2018) found that students evaluated non-instructor specific aspects (e.g., workload, technology, information available) of an online course more positively when the course was given by a male rather than a female professor, even when the above-mentioned aspects were identical and not dependent on the instructor.

While the effects of gender on performance evaluations are difficult to establish in individual real-life cases, archival/aggregate data can provide useful insights. Wennerås and Wold (1997) compared the research merits of men and women who had been awarded post-doctoral fellowships in Sweden, and analyzed these with respect to three generally established measures of research performance – number of articles, impact factor of journals published in, and citations. They found that female applicants had to be significantly more productive than the average male applicant to receive the same competence score, indicating higher performance standards for women. Similarly, in a follow-up study, the findings of Sandström and Hällsten (2008) show that women scoring in the top quintile in terms of citations and journal impact factors were rated more negatively than their male counterparts, and that there appeared to be weaker relationships between some objective performance criteria and ratings for women than for men.

Furthermore, Lyness and Heilman (2006), who studied the relationships between performance evaluations and if candidates were promoted during the subsequent two years following those performance appraisals, found that promoted women had received more positive performance appraisals than the promoted men, again suggesting that women were held to stricter standards for promotion. This is in line with earlier findings on higher performance standards for women (Biernat & Kobrynowicz 1997). In contrast, low performing women appear to receive more lenient evaluations than low performing men (Biernat & Kobrynowicz 1997; Sandström & Hällsten 2008). However, this may have fewer practical implications as low performing candidates, irrespective of their exact ratings, are unlikely to be selected, promoted, or awarded grants or other rewards.

As discussed earlier, women may be evaluated less favorably than men for and in leadership positions because of the perceived incongruity between the female gender role and leadership roles (Eagly & Karau 2002). This is particularly so in more male-dominated and male-typed settings. Meta-analyses provide support for the significance of perceived role incongruity: in particular, men appear to be rated more positively for male gender-typed jobs, whereas findings on ratings of women vs. men in female-dominated jobs are more mixed (Davidson & Burke 2000; Koch et al. 2015).

Furthermore, not only may men's and women's performance be rated differently, but research suggests success may also be subject to different attributions. While, more often, men's successes are ascribed to skill, women's are put down to luck or hard work (Swim & Sanna 1996). Also, when there is any ambiguity about the contribution of individual members to a specific outcome, stereotypical expectations are likely to influence inferences about who actually deserves credit (Heilman & Parks-Stamm 2007). Research has indicated, for example, when a woman works together with a man on a joint task, she is given less credit for a successful joint outcome, viewed as having made a more trivial contribution, and as less competent than her male teammate (Heilman & Haynes 2005). In line with this, recent research at the Clayman Institute suggests women's performance appraisals more often make reference to 'team accomplishments', whereas men's highlight individual achievements (Correll & Simard 2016; Silverman

2015). Similarly, Kulich et al. (2011) found that the compensation of female executives was less performance-sensitive than that of males, suggesting female executives were seen as less responsible for their company's performance.

The findings summarized above point to gender differences in performance ratings, with research suggesting female candidates are evaluated less favorably than male despite similar merits, particularly in male sex-typed jobs (see Koch et al. 2015). In the next section, the role of gender stereotypes and their effect on leader evaluations will be discussed in more detail.

Gender Stereotypes and Different Criteria

Gender stereotypes, both descriptive and prescriptive, may play an important role in how men and women and male and female leaders are evaluated. According to Heilman (2001), men are typically expected and encouraged to be agentic (i.e., assertive and independent), whereas women are expected and encouraged to be communal (i.e., nurturing and caring). Failure to live up to these expectations or exhibit gender-incongruent behavior may lead to negative reactions (Rudman & Glick 2001).

Recent research by Correll and Simard (2016) confirms that these expectations show up in real-life performance appraisals. In the performance reports they analyzed, women were described twice as often as men as supportive, collaborative and helpful, and their appraisal contained twice the number of references to team as opposed to individual accomplishments (Silverman 2015). In contrast, men's appraisal focused on assertiveness, independence, and self-confidence, and feedback was much more often linked to business outcomes or technical expertise. While development feedback for men was typically more specific and actionable, feedback given to women was typically vaguer and often focused on their 'communication style' (Correll & Simard 2016). This is also supported by the researchers' finding that men were in their performance appraisals praised more often than women for assertiveness, yet women received 2.5 times more comments on 'aggressive' communication styles (Silverman 2015), pointing to similar behavior possibly being labeled differently depending on whether or not it is seen as gender-congruent.

Similar patterns have been reported in student evaluations of university professors. A systematic review of 14 million online student evaluations on 'RateMyProfessor' revealed professor gender affected which qualities students paid attention to (Schmidt 2015a,b; Storage et al. 2016). For instance, not only the words 'brilliant', 'genius', but also 'idiot' were used significantly more often to describe (good vs. bad) male professors than female professors across disciplines. In contrast, 'friendly' and 'unfriendly', and 'helpful' and 'unhelpful' were used much more often in evaluating female professors, indicating a focus on the existence, or lack of, communal traits.

Descriptive stereotypes of women may also cause hesitation as to whether they are likely to occupy leadership roles successfully, roles which have traditionally been male gender-typed. Seminal work by Schein (1973; 1975) showed that descriptions of women in general were far less congruent with descriptions of successful managers than were descriptions of men. Research suggests that perceptions of 'think leader, think male' are surprisingly persistent, and still today stereotypes continue to portray successful leaders as more similar to men than women (Koenig et al. 2011), whereas follower traits are more strongly associated with female traits (Braun et al. 2017).



Another aspect of gender stereotypes is that women who exhibit agentic traits and succeed at tasks traditionally seen as male are often considered to be violating gender-stereotypic prescriptions (Heilman & Okimoto 2007). Research shows that successful women, such as strong and competent female leaders, are typically assumed to lack communal traits, be cold, and have low interpersonal skills, which also may result in lower ratings on likeability and hireability (e.g., Heilman et al. 2004). Similarly, students who were asked to write stories about fictive male and female CEOs and their behavior described female CEOs as highly competent, but lacking in interpersonal skills, whereas the male CEOs were described not only as successful but also as naturally competent leaders of people (Katila & Eriksson 2013). Good ratings on competence do not automatically result in a female candidate being seen as 'suitable'. For example, when performance evaluations of male and female junior attorneys at a Wall Street law firm were analyzed, Biernat et al. (2012) found that male supervisors offered narrative comments displaying either no sex effects or greater favorability toward women, but still judged male attorneys more favorably than female attorneys on the numerical ratings that mattered for promotion. Furthermore, they found that predictors of numerical ratings differed by sex: narrative ratings of technical competence were more significant for men than women, and those of interpersonal warmth mattered more for women than men. The data suggest subtle patterns of gender bias, where women were tarnished by not meeting gendered expectations of interpersonal warmth, and benefitted less than men by meeting masculine standards of high technical competence.

Research Question and Context

Aim of the study

While considerable research, as described above, points to the historical significance of gender in evaluations (e.g., Koch et al. 2015), other research suggests such biases may be diminishing over time as gender equality in society increases (Paustian-Underdahl et al. 2014). In response to this controversy, this study aims to examine the possible persistence of gender bias in the evaluation of leaders. More precisely, the study aims to examine the following research question:

RQ) How does gender affect evaluation of leaders?

This article reports on findings from two different studies, both designed to examine how gender affects leader evaluations. The two different studies approach the question from slightly different angles. First, the focus of Study 1 is to study if leader gender influences the evaluation of specific leadership behaviors. Second, the focus of Study 2 is whether employees pay attention to different qualities when evaluating male and female leaders.

Research context

Both studies were conducted in Finland, which typically ranks very high in international studies on gender equality. According to the World Economic Forum's (2018) Global gender gap report, Finland was the fourth most gender-equal country in the world. Also,

it has been ranked as the second best place to be a mother (Save the Children 2015). Finnish women have almost the same employment rate as men (68.5% vs 70.7% of those aged 15–64) and slightly higher educational attainments (Statistics Finland 2018).

However, the high international ranking easily creates the illusion that gender equality has already been achieved (Korvajärvi 2002) and that gender inequality and gender bias are a thing of the past or something that occurs only in other countries. Still, statistics point to a clear under-representation of women in high positions in business life in Finland (World Economic Forum 2018). As for listed companies, 27.2% of board members were women, 5% of chairs of the boards were women, and 8% of managing directors were women in 2017 (Statistics Finland 2018). Findings from Statistics Finland (2018) further show that women's monthly salaries amount to 84% of those of men. While men typically work in the private sector (84.8%), women are strongly overrepresented in the municipal sector. Gender differences can also be noted when it comes to fixed term work: while 13.2% of men have fixed term contract, the corresponding number for women is 18.8%. This gender difference is particularly pronounced in the younger age groups.

Finland's high scoring in international rankings of gender equality means it is often portrayed as an egalitarian country in societal discourses. However, as shown by the numbers above, this does not necessarily mean that gender equality has been achieved. Still, Finland's high ranking means that it offers a particularly interesting case in which to study the possible persistence of gender stereotypes. If relatively high levels of equality are associated with the disappearances of gender stereotypes, there are trends of interest in other countries worthy of study. On the other hand, the persistence of gender stereotypes in one of the world's most gender-equal countries would present particularly strong evidence of persistence further afield.

Method

The research question 'How does gender affect evaluation of leaders?' was examined with the help of two different empirical studies. First, respondents were asked to rate two different fictive leaders, an abusive/authoritarian leader and a participate leader, and the names of the leaders were manipulated to portray either a male or female leader. The results were analyzed statistically. Second, a different group of respondents were asked to describe both good and poor real-life leaders they had encountered. The stories were analyzed with respect to what kind of characteristics participants paid attention to when describing male and female (good and poor) leaders. Qualitative content analysis was used for this part of the study. The two empirical studies are described in more detail below.

Study I

Study 1 sought to examine whether there are differences in how the same leader behavior is described and evaluated, depending on whether respondents are evaluating a male or female candidate. More specifically, 358 business school students, potential future leaders, with work experience, were asked to read a teaching case where leader gender



had been manipulated. The case description comprised an account of both an abusive, highly authoritarian leader and of a participative leader, both of whom were successful in financial terms (see Rollinson 2005). The authoritarian/abusive leader was described as someone who set clear targets for their staff and kept them 'on their toes' by quickly and publicly reprimanding those who failed to perform in line with expectations. The participative leader was described as working closely with employees on problem solving and regularly praising them and showing appreciation, thus trying to use positive reinforcement rather than punishment to elicit the desired behavior.

Approximately half of the students received the original version, where the abusive leader was male (Charles) and the participative leader female (Helene). The other half received a manipulated version, where the text was identical except for the name of the leaders and corresponding pronouns. Charles was renamed Charlotte and Helene became Henry. Student groups were randomly assigned one of the two versions, and students completed their evaluation individually in class and handed it in. This was immediately followed by a discussion of the case in class, where the focus was on how different leadership styles and punishment and positive reinforcement affect follower behavior and attitudes.

Respondents were asked to rate the two leaders on several qualities, including perceived effectiveness, likeability, and hireability. All statements began with the phrase 'In my opinion Charles/Charlotte/Helene/Henry ...' *Perceived effectiveness* was measured with two items: (i) ... is effective as a leader, and (ii) ... leads his/her employees to excellent achievements. Cronbach's alpha for perceived effectiveness was 0.80. *Likeability* was also measured with two items: (i) ... is someone I would get along with well, and (ii) ... is someone I could see myself working for. Cronbach's alpha was 0.95. *Hireability* was measured with a single-item: ... is someone I would consider hiring for a managerial position. All questions were measured on a 7-point Likert scale from (1) strongly disagree to (7) strongly agree.

After rating each leader, respondents were given the opportunity to comment upon the leader in writing 'Your evaluation of/comments about Charles (alt. Charlotte/Henry/Helene) as a leader'. These replies were used to further increase understanding of how respondents made sense of the different leaders.

As for sample characteristics, the respondents had an average of 2.6 years of work experience. About 36.3% (130) were male and 62.0% (222) female, corresponding to the typical gender distribution in the subject of Management and Organization at the Business School in question. About 1.7% (6) did not report their gender. No significant difference in work experience was found between the male and female respondents.

Study 2

The aim of Study 2 was to examine what dimensions subordinates pay attention to when evaluating leaders, and to examine possible gender differences in those dimensions. Study 2 involved collecting subordinate descriptions of good and poor ('not so good') leaders, and analyzing whether gender affected the type of dimensions the subordinates chose to focus on. The study used a convenience sample; yet, seeking to target primarily young and highly educated employees from the capital area, that is, groups typically assumed to embrace high levels of gender equality. First, a link to an online

questionnaire was distributed via social media. Second, business school students attending a class on leadership were asked to share experiences of leaders they had worked for, using a printed version of the very same questions. This resulted in a total of written 119 stories, all including descriptions of both good and poor leaders.

The instructions to the participants were first to think of a good leader they had worked for, and then (a) describe the leader's most important characteristics, and (b) describe the leader's typical behavior. The participants were next instructed to think of a poor/not so good leader they had worked for, and reflect upon the very same questions. At the end of the stories, the participants were asked to provide background information about the leaders they had commented on (sector, gender, approximate age), in addition to background information about themselves (age, gender, education).

Of the participants, 74% (88) were women and 76% (90) were under 35 years of age. Fifty percent either had a university degree or were currently enrolled at a university. As for the good leaders, the participants chose to describe, 61 were female and 58 male. Of the poor leaders, 53 were female and 66 male.

The replies were analyzed using qualitative content analysis supported by the ATLAS.ti tool (a qualitative analysis software package). The first step in qualitative content analysis is to develop the coding frame (Schreier 2012). This was partly theory-driven (based on a literature review of relevant characteristics and behaviors highlighted in the leadership literature), and partly data-driven (adding categories from the material in question). A large set of different qualities found in the stories was first listed, to be clustered later into more specific groups based on similarity and links to the literature. After the author had performed a first round of coding to establish the different categories and the coding frame, all the data were systematically coded, by two coders trained to use the frame. Coder 2's results largely corroborated those of coder 1: 84% of coder 1's results were replicated by coder 2. After coding the data, the coders met to discuss the coding, comparing differences and similarities (cf. Schreier 2012). When in disagreement, the author compared the codings and made the final decision on which category best fit the interview response. Frequencies for each category were counted, and used to corroborate the overall impressions and ensure no individual, particularly articulate responses led us to make claims not supported by the rest of the material.

Results

Study I

The first step of Study 1 involved comparing the evaluations respondents had given of the male versus female authoritarian/abusive leader. This was analyzed separately for the male and female respondents (Table 1).

For perceived effectiveness, it appeared men rated the male authoritarian/abusive leader clearly higher than the female authoritarian/abusive leader ($p = 0.014$). Similarly, male respondents rated the authoritarian/abusive male leader higher on likeability than his female counterpart ($p = 0.035$). Moreover, the male respondents reported a marginally, albeit not significantly, higher preference to hire the male candidate ($p = 0.06$). For the female respondents, no significant differences were found in the ratings of male and female authoritarian/abusive leaders.

**Table 1** Evaluation of male and female authoritarian/abusive leader by male and female students

	Men (n = 130)			Women (n = 222)	
	Male leader	Female leader		Male leader	Female leader
Effectiveness	4.83	4.32	*	4.15	4.24
Likeability	2.76	2.28	*	2.19	2.13
Hireability	3.43	2.95	†	2.58	2.74

* $p < 0.05$, † $p < 0.10$.

It is worth noting that the female leader was evaluated very similarly by male and female respondents – the ratings are almost identical as can be seen in Table 1, and none of the differences between male and female respondents were statistically significant. Also, the female respondents' ratings of the male leader were very similar to how both men and women rated the female leader. In contrast, the male respondents' ratings of the male leader clearly stood out. In fact, the ratings given by male and female respondents for the male authoritarian/abusive leader were significantly different on all three criteria: effectiveness ($p < 0.001$), likeability ($p < 0.01$), and hireability ($p < 0.01$). It thus appears that a man rating a male abusive leader cuts him some extra 'slack' not given to women, and not given by women to either male or female abusive leaders.

The written open replies given by the students supported the findings reported above. While students did express concerns that both Charles and Charlotte were 'strict', 'hard', and 'harsh' and expressed concerns about the job satisfaction and well-being of employees, the replies suggested respondents were still somewhat more tolerant toward Charles as a leader. Typical comments about Charles were, for instance:

'Goal-oriented, but does not treat the employees well'

'Even though he's strict and employs a hard routine, the fact is that he still manages to get good results'

'He is a more old-fashioned leader, who to some extent controls by fear'

While the comments made about Charles often portrayed him as 'harsh, but successful', respondents seemed to react even more negatively toward Charlotte. Although some of the respondents acknowledged that Charlotte was getting good results, she was criticized even more than Charles, and the criticism appeared more personal, including words such as 'bully', 'unprofessional', and 'dictator', as shown in the quotations below.

'She acts like a dictator rather than as a team player'

'Unprofessional. She is leading a store, not an army, she needs to be less of a bully'

Although the class discussions held after students had returned their written evaluations were not systematically recorded, it is worth noting that also in these discussions

the male and female authoritarian/abusive leaders were described with very different words. Whereas the male abusive leader was described as a ‘strong, military type of leader’, albeit somewhat ‘old-fashioned’, highly derogatory terms were used to describe the female leader (including ‘bitch’). Although several students also questioned the effectiveness of the male abusive leader’s behavior, no similar personal statements were made about him. Thus, the same behavior elicited very different reactions and interpretations, depending on leader gender.

For the participative leader, no significant differences in ratings were found based on leader gender. The ratings on perceived effectiveness, likeability, and hireability were almost identical for the male and female versions of the story (see Table 2). In the open replies, the respondents described Henry/Helene as a ‘constructive’ and ‘supportive’ leader, for whom they would happily work themselves. Some respondents felt there was risk subordinates could possibly try to take advantage of the fact that their leader was ‘too soft’. However, as with the numerical ratings, comments about Henry and Helene were highly identical and no notable gender differences emerged.

Table 2 Evaluation of male and female participative leader by male and female students

	Men (n = 130)		Women (n = 222)	
	Male leader	Female leader	Male leader	Female leader
Effectiveness	5.80	5.75	5.70	5.70
Likeability	6.15	5.94	6.09	6.09
Hireability	5.82	5.81	5.89	5.74

None of the differences were significant on a $p < 0.05$ level.

Study 2

The second part of the study sought to analyze whether participants used the same or different dimensions when evaluating male and female leaders. A particular emphasis was placed on dimensions over-represented in descriptions of both good and poor leaders in one gender compared to leaders in the other. Over-representation in both types of story seems to indicate that particular emphasis was given to a specific dimension, more so than signaling actual differences between males and females in leader behaviors.

Overall, the dimensions participants mentioned most often when describing good leaders were related to demonstrating integrity, showing fairness, being available, communicating effectively, showing consideration, supporting employees, motivating/inspiring employees, organizing work, taking responsibility, and being emotionally stable (see Table 3). Similarly, *the lack* of the same qualities surfaced regularly in the descriptions of poor leaders. Although all of these dimensions were mentioned repeatedly for both male and female participants, some of them appeared to surface much more frequently in the descriptions of female leaders, regardless of whether the story focused on their existence or the lack thereof.

Table 3 Categories mentioned in descriptions of good and poor leaders, and percentage of participants mentioning examples from these categories

Characteristic or behavior mentioned (or lack thereof)	Explanation	Good female leaders (n = 61)	Good male leaders (n = 58)	Poor female leaders (n = 53)	Poor male leaders (n = 66)
Appreciation	Gives positive feedback; notices and praises employees for good performance	16%	13%	11%	16%
Determination	Takes charge, gets things done, not afraid to lead/intervene, firm, has authority, assertive	21%	23%	6%	14%
Availability	Easy to approach, is present, listens actively	43%	22%	28%	19%
Communication	Gives clear instructions, shares information, constructive feedback	34%	29%	10%	25%
Consideration	Friendly, caring, emphatic, understanding	40%	25%	28%	13%
Expertise	Knowledgeable, competent, experienced, expert in own field	13%	25%	5%	14%
Fairness	Treats all subordinates equally/ doesn't have favorites	39%	24%	38%	11%
Flexibility	Flexible, willing to negotiate	13%	12%	7%	10%
Humility	Does not think too much of him/herself, 'an equal', not arrogant	10%	18%	9%	19%
Humor	Fun to be around, not too serious, sense of humor	2%	8%	2%	0%
Integrity	Trustworthy, honest, keeps his/her word, ethical, consistent	38%	26%	23%	24%
Motivation/ inspiration	Inspires employees to good performance, sets challenging goals, charismatic	23%	25%	7%	6%
Openness	Open-minded, receptive to new ideas, suggestions, etc.	16%	10%	11%	15%
Extrovert	Sociable, outgoing, not shying away from social contact	10%	14%	0%	9%
Problem-solving/ organizing	Structured and systematic, good planning and decision-making skills	20%	28%	21%	13%
Responsibility	Takes responsibility, dependable, hard-working	19%	26%	25%	28%

Characteristic or behavior mentioned (or lack thereof)	Explanation	Good female leaders (n = 61)	Good male leaders (n = 58)	Poor female leaders (n = 53)	Poor male leaders (n = 66)
Emotional stability	Even-tempered, calm, can handle stress, not prone to nervousness, anger etc.	23%	21%	30%	20%
Support	Supportive, helps out	36%	25%	12%	8%
Trust	Trusts his/her employees, doesn't micromanage	16%	20%	13%	12%
Inappropriate treatment/bullying	Abusive, belittling, yelling, takes credit for subordinate's work, using subordinate as a scapegoat, leading by fear			23%	32%

Being (un)available, (in)considerate, and (un)fair were all themes that characterized the descriptions of female leaders to a clearly greater extent than those of male leaders. Availability involved being ‘easy to approach’, present, and listening actively. Consideration involved being friendly, caring, emphatic, and understanding. Fairness mostly centered around treating all subordinates equally rather than having favorites. Being (un)supportive was a related category, although gender differences were somewhat less pronounced. To a greater extent than their male counterparts, women were thus praised when they expressed communal traits, and demonstrated care and nurturing, and criticized if they failed to do so.

‘My supervisor is understanding and she takes her subordinates and their wishes into consideration’ (good female leader)

‘I didn’t see her much, she was always late for meetings, she didn’t support me in difficult situations’ (poor female leader)

Taking a closer look at the consideration category, some more specific differences emerged. One noteworthy pattern was the use of the word ‘cold’. While this term was employed to describe some poor female leaders, it was never used to describe a poor male leader. Also, some female leaders, but not male leaders, were criticized for ‘not smiling’.

Given that the same dimensions recurred in descriptions of leaders good and bad, it seems likely they reflect differences in follower expectations. The other top categories did not follow clear gender patterns. Either no differences could be found or gender differences varied depending on whether good or poor leaders were being described, making it impossible to tell to what extent these differences reflected expectations or leader behavior.

For the poor leaders, a separate category termed ‘inappropriate treatment/bullying’ was created to cover behaviors that related to power abuse, harassment, and other directly inappropriate behavior. In contrast to the category termed lack of consideration, this implied not only a lack of nurturing behaviors but also the presence of active,



negative behaviors. These included belittling, yelling, withholding information, taking credit for subordinates' work, and leading by fear. While being inconsiderate was mentioned for more women than men, this more active category of bullying was used more often in descriptions of male than female leaders.

'Had a very arrogant attitude and never listened to anyone's ideas or advice. Very loud and yelled a lot – created fear throughout the office'. (poor male leader)

Based on previous findings, it could have been assumed that men would be more likely than women to be rated on agentic traits, such as showing determination, taking responsibility, and engaging in organizing/problem-solving, but this material did not provide clear support in that respect. However, expertise (or lack thereof) was mentioned clearly more often for the male than female leaders.

'He was confident enough, but this confidence was also based on knowledge and competence'. (good male leader)

'Doesn't develop his own expertise'. (poor male leader)

Interestingly, humility was also something men were evaluated on more often than their female counterparts. For instance, one male leader was praised for 'not highlighting his position: an outsider would easily have mistaken him for a rank-and-file employee'. But more men than women were criticized for being arrogant or thinking too much of themselves. This included comments such 'self-centered, thinks his opinion is the only right one' and 'treats subordinates as if they are worthless'.

Although the sample was too small for an in-depth analysis of how leader gender and participant gender interacted, analyses were conducted to examine how participant gender affected the dimensions that were mentioned. Overall, female participants mentioned a higher number of characteristics in their stories. While communal traits were mentioned by a higher percentage of women, the same was also true for more gender-neutral categories, such as communication, integrity, motivation, and some agentic traits, such as problem-solving/organizing. Thus, the fact that women somewhat more often rated female leaders and men somewhat more often male leaders does not appear to explain the differences found.

Discussion

This study investigated the possible persistence of gender bias in a sample of young, highly educated participants in a country reported to be one of the most gender-equal countries in the world. The findings support many of the gendered biases reported in the previous research and suggest gendering of leadership is still ongoing.

First of all, the results provide support for the earlier assertion that women are to a higher extent than men expected (and required) to exhibit communal traits (e.g., Eagly & Karau 2002; Heilmann 2001). This was illustrated by the fact that when respondents were asked to describe good and poor leaders, respondents paid more attention to communal aspects when describing female than male leaders. While being available,

considerate and fair were important both when describing male and female good leaders, respondents paid particular attention to these qualities when discussing female leaders. Similarly, the lack of these qualities seemed to play a more important role when labelling a female leader as poor, compared to a male leader. This study thus provides support for the argument that gender stereotypes is one of mechanisms that explain different ratings for male and female leaders.

The findings from the other study, where respondents were asked to rate fictive leaders, provide additional support for this. When respondents were asked to rate a participative leader, who according to the story was supportive, considerate, and listened to their subordinates, the ratings for male and female leaders were more or less identical. However, when respondents were asked to rate an abusive/authoritarian leader, who lacked the communal traits typically expected from women, clear gender differences emerged. Now the female leader (who lacked communal qualities) was rated significantly lower than her male counterpart. This is in line with the previous research, which has shown that female leaders who deviate from prescriptive norms for women to be communal and nonaggressive are rated more negatively (Eagly et al. 1992), and that the use of intimidation has more negative effects for female leaders than male (Bolino & Turnley 2003). This suggests that women have a narrower set of leadership styles at their disposal.

Overall, the results suggest that respondents have lower tolerance for negative behavior from female than from male leaders. As discussed above, the female abusive/authoritarian leader got lower evaluations than her male counterpart. Also, the written stories about real-life leaders suggested that female leaders were often criticized and labeled as ‘poor leaders’ simply because they failed to show the level of consideration and support expected. In contrast, a clearly higher amount of the men labeled as poor leaders were reported to exhibit active negative behavior, such as power abuse or harassment. That fewer poor female leaders than male leaders were reported to engage in direct forms of inappropriate treatment and abuse could also reflect the findings from Study 1, that is, that female showing such traits are given lower ratings and therefore not considered for and selected to leadership positions to the same extent as males exhibiting similar behavioral patterns. Again, showing a lack of communal traits and even exhibiting negative interpersonal behavior is thus likely to be seen as more gender-incongruent behavior for women, and as such may lead to stronger negative reactions (cf. Rudman & Glick 2001).

Moreover, the results also point to the fact that men were more likely than women to be affected by leader gender in their ratings. This was clearly shown in Study 1. Again, this is in line with the previous meta-analytical evidence, suggesting male raters exhibit greater gender-role congruity bias than do female raters for male-dominated jobs (Koch et al. 2015).

In particular, the results seem to suggest that men have a higher acceptance of authoritarian/abusive male leaders. While men gave low ratings to abusive female leaders and women gave low ratings to both male and female abusive leaders, men seemed to tolerate abusive behaviors from a male leader to a somewhat higher extent. It is possible this may be partly explained by cultural notions of masculinity or masculinity contest cultures (Matos et al. 2018), prescribing socially dominant and even abusive behavior as acceptable or even desirable among men.

Contrary to expectations, no clear gender differences could be found with respect to how much attention participants paid to most agentic traits, such as showing



determination, being responsible, and engaging in organizing/problem-solving. This seems to suggest participants do not have different expectations with respect to these, and may suggest the traits per se are no longer seen as male gender-typed in the Finnish context. However, expertise was still an aspect that participants paid more attention to when evaluating male than female leaders. This is in line with the previous findings that for male candidates, there appears to be a stronger link between technical competence and overall ratings than is the case for women (Biernat et al. 2012; Sandström & Hällsten 2008).

The research designs replicate previous studies in the field, and provide support for the persistence of gender bias and gender stereotypes in leader evaluations. The results of this study clearly show gender stereotypes and gender bias still affect evaluations. On a more general level, the study also provides support for the ongoing gendering of leadership (cf. Holgersson 2013). Rather than being a thing of the past, gender continues to permeate the way we think about leadership and successful leaders, often to the detriment of female leaders.

Finding support for gendered ratings of leaders in Finland is of particular importance, as global gender equality reports typically rate the country one of the most equal societies. Also, the fact that the sample is young and highly educated is pertinent, since both are often cited in the public discourse as factors believed to automatically reduce (or even eliminate) gender stereotypes and gender discrimination.

Implications

That gender stereotypes and gender bias continue to affect leader evaluations has implications for organizations and candidates (primarily female candidates) in many respects. Above all, it highlights the risk of discrimination in recruitment and selection decisions, as well as in performance appraisal, which is often strongly linked to promotion and compensation decisions. This points to the importance of taking active measures to combat such tendencies, rather than assuming the mere intention to select the most qualified candidate or provide a correct assessment will result in a fair and nondiscriminatory outcome.

Several methods to reduce gender bias have been put forward in the previous research. Although educating decision-makers about unconscious bias is one important step, Correll (2017) reminds us of the importance of reviewing and revising organizational processes. This may involve clarifying criteria and reducing ambiguity, ensuring criteria do not have built-in gender biases, avoiding definitions of success that are unduly narrow or irrelevant, and ensuring (and monitoring) that criteria are applied equally to men and women. Also, continuously monitoring whether decisions are bias-free and encouraging leaders to examine each other's decisions from that perspective, thus creating a culture of peer accountability, is considered important (Correll 2017). The present study points to the continued importance of actively taking these kinds of measure to reduce the risk of bias.

It is also worth highlighting that Study 1 served an important pedagogical function for its participants. The author's experience is that business school students in Finland often meet gender issues with some skepticism, frequently arguing that gender discrimination has been consigned to history, albeit still a factor in some parts of the world, and

is in particular of no relevance to them as young professionals in the contemporary, gender-equal Finnish business context. Being able to present findings from this study at the end of their introductory course on leadership when gender issues were raised, gave rise to a very different climate of discussion, where students actively sought to understand why they as a group had responded in the way they did. This experience points to the importance of regularly collecting empirical evidence of gender bias to use as a tool for discussion not only with students but also with managers, politicians, and policy makers (cf. Correll 2017).

Limitations and Suggestions for Further Research

While this study provided support for the persistence of gender bias in leader evaluation, there are a number of limitations to consider. As Study 1 was originally designed to provide a basis for discussion on gender bias in class, the volume of data collected was rather limited, both in terms of the number of observations and variables included. For instance, data collected on gender role beliefs (e.g., Buchanan 2014) could have been added, to see if it was a specific group of male respondents who were particularly negative towards authoritarian women. Also, more systematic recording of the class discussions that followed after the respondents had read the case, could have provided insights into how students perceived the male vs. female leaders. Student comments about ‘old fashioned’, ‘strong military type of leader’ vs. ‘bitch’ suggest respondents constructed the authoritarian male and female leaders quite differently. Further research should examine these images in more detail.

Furthermore, this study provided only one description of an authoritarian/abusive leader and one of a participative leader. A broader range of different leadership styles could have been provided. For instance, it could be of interest to see if leader gender affects ratings of laissez-faire leaders who score low on both task-related and relationship-related aspects, and as a result score low on both communal and agentic traits. This should, thus, be addressed in further research.

Study 2 provided insights into the dimensions participants paid attention to when thinking of good and poor leaders, and how leader gender influenced this. While the study provided insights into the characteristics participants spontaneously listed, and thereby hopefully those most salient to the them, participants could alternatively have been asked first to think of a specific good or poor leader, and then asked to rate the particular leader on a predefined list of characteristics. Further research could, thus, examine whether similar differences would emerge, if participants were to rate candidates in that way. Furthermore, given the limited sample size, it was not possible to control for industry sector or other contextual factors that may have affected employee expectations of ‘good leadership’. Such additional analyses should be added to increase the robustness of the findings.

Conclusion

The results of this study suggest gender bias in evaluation continues to affect leader ratings. Female and male leaders were rated differently for identical behavior, particularly



when the rater was male and the leader exhibited gender-incongruent behavior. In addition, raters paid attention to different aspects as a function of leader gender, suggesting leader expectations are gendered. The results point to the importance of taking active measures to offset the effects of bias in recruitment and performance appraisal processes, and to ensure the same evaluation criteria and performance standards are applied irrespective of the gender of the candidate being assessed.

Acknowledgments

This study was partly funded by the Academy of Finland (grant number 308843). The author would like to thank Kiira Louhivuori and Cecilia Grönroos for their help with data collection and coding.

References

- Acker, J. (1991). Hierarchies, jobs, bodies: a theory of gendered organizations. In J. L. Lorber & S. A. Farrell (Eds), *The Social Construction of Gender* (pp. 162–179), Newbury Park, CA: Sage.
- Biernat, M., & Kobrynowicz, D. (1997). Gender- and race-based standards of competence: lower minimum standards but higher ability standards for devalued groups, *Journal of Personality and Social Psychology* 72(3): 544–557. doi: <https://doi.org/10.1037//0022-3514.72.3.544>.
- Biernat, M., Tocci, M. J., & Williams, J. C. (2012). The language of performance evaluations: Gender-based shifts in content and consistency of judgment, *Social Psychological and Personality Science* 3(2): 186–192. doi: <https://doi.org/10.1177/1948550611415693>.
- Bolino, M. C., & Turnley, W. H. (2003). Counternormative impression management, likeability, and performance ratings: the use of intimidation in an organizational setting, *Journal of Organizational Behavior* 24(2): 237–250. doi: <https://doi.org/10.1002/job.185>.
- Braun, S., Stegmann, S., Hernandez Bark, A. S., Junker, N. M., & van Dick, R. (2017). Think manager—think male, think follower—think female: Gender bias in implicit followership theories, *Journal of Applied Social Psychology* 47(7): 377–388. doi: <https://doi.org/10.1111/jasp.12445>.
- Buchanan, T. (2014). The influence of gender role attitudes on perceptions of women's work performance and the importance of fair pay, *Sociological Spectrum* 34(3): 203–221. doi: <https://doi.org/10.1080/02732173.2014.895637>.
- Carli, L. L., & Eagly, A. H. (2016). Women face a labyrinth: an examination of metaphors for women leaders, *Gender in Management: An International Journal* 31(8): 514–527. doi: <https://doi.org/10.1108/GM-02-2015-0007>.
- Collinson, D. & Hearn, H. (1994). Naming men as men: Implications for work, organization and management, *Gender Work and Organization* 1(1): 2–22. doi: <https://doi.org/10.1111/j.1468-0432.1994.tb00002.x>.
- Correll, S. J. (2017). Reducing gender biases in modern workplaces: a small wins approach to organizational change, *Gender & Society* 31(6): 725–750. doi: <https://doi.org/10.1177/0891243217738518>.
- Correll, S., & Simard, C. (2016). Vague feedback is holding women back. Harvard Business Review on-line article. Available at <https://hbr.org/2016/04/research-vague-feedback-is-holding-women-back> Accessed 06 April 2018.

- Davidson, H. K., & Burke, M. J. (2000). Sex discrimination in simulated employment contexts: a meta-analytic investigation, *Journal of Vocational Behavior* 56(2): 225–248. doi: <https://doi.org/10.1006/jvbe.1999.1711>.
- Eagly, A. H., & Carli, L. L. (2007). *Through the labyrinth: The truth about how women become leaders*, Cambridge, MA: Harvard Business School Press.
- Eagly, A. H., Karau, S. J. (2002). Role congruity theory of prejudice toward female leaders, *Psychological Review* 109(3): 573–598. doi: <https://doi.org/10.1037/0033-295X.109.3.573>.
- Eagly, A. H., Makhijani, M. G., & Klonsky, B. G. (1992). Gender and the evaluation of leaders: A meta-analysis, *Psychological Bulletin* 111(1): 3–22. doi: <https://doi.org/10.1037//0033-2909.111.1.3>.
- Goldberg, P. (1968). Are women prejudiced against women? *Trans-action* 5(5): 28–30. doi: <https://doi.org/10.1007/bf03180445>.
- Heilman, M. (2001). Description and prescription: how gender stereotypes prevent women's ascent up the organizational ladder, *Journal of Social Issues* 57(4): 657–674. doi: <https://doi.org/10.1111/0022-4537.00234>.
- Heilman, M., & Haynes, M. (2005). No credit where credit is due: attributional rationalization of women's success in male-female teams, *Journal of Applied Psychology* 90(5): 905–916. doi: <https://doi.org/10.1037/0021-9010.90.5.905>.
- Heilman, M., & Okimoto, T. (2007). Why are women penalized for success at male tasks?: The implied communality deficit, *Journal of Applied Psychology* 92(1): 81–92. doi: <https://doi.org/10.1037/0021-9010.92.1.81>.
- Heilman, M. E., & Parks-Stamm, E. J. (2007). Gender stereotypes in the workplace: Obstacles to women's career progress. In S. J. Correll (Ed.), *Advances in Group Processes* (pp. 47–77), Amsterdam: Elsevier JAI.
- Heilman, M. E., Wallen, M. M., Fuchs, D., & Tamkins, A. S. (2004). Penalties for success: reactions to women who succeed at male gender-typed tasks, *Journal of Applied Psychology* 89(3): 416–427. doi: <https://doi.org/10.1037/0021-9010.89.3.416>.
- Holgersson, C. (2013). Recruiting managing directors: doing homosociality, *Gender, Work & Organization*, 20(4): 454–466. doi: <https://doi.org/10.1111/j.1468-0432.2012.00595.x>.
- Kanter, R. M. (1977). *Men and women of the corporation*, New York: Basic Books.
- Katila, S., & Eriksson, P. (2013). He is a firm, strong-minded and empowering leader, but is she? Gendered positioning of female and male CEOs, *Gender, Work and Organization* 20(1) 71–84. doi: <https://doi.org/10.1111/j.1468-0432.2011.00570.x>.
- Knobloch-Westerwick, S., Glynn, C.J., & Huge, M. (2013). The Matilda Effect in science communication: an experiment on gender bias in publication quality perceptions and collaboration interest, *Science Communication* 35(5): 603–625. doi: <https://doi.org/10.1177/1075547012472684>.
- Koch, A. J., D'Mello, S. D., & Sackett, P. R. (2015). A meta-analysis of gender stereotypes and bias in experimental simulations of employment decision making, *Journal of Applied Psychology* 100(1): 128–161. doi: <https://doi.org/10.1037/a0036734>.
- Koenig, A. M., Eagly, A. H., Mitchell, A. A., & Ristikari, T. (2011). Are leader stereotypes masculine? A meta-analysis of three research paradigms, *Psychological Bulletin* 137(4): 616–642. doi: <https://doi.org/10.1037/a0023557>.
- Korvajärvi, P. (2002). Gender-neutral gender and denial of difference. In B. Czarniawska & H. Höpfl (Eds.), *Casting the other. The production and maintenance of inequalities in work organizations* (pp. 119–137), London: Routledge.
- Kulich, C., Trojanowski, G., Ryan, M. K., Haslam, S. A., & Renneboog, L. R. (2011). Who gets the carrot and who gets the stick? Evidence of gender disparities in executive remuneration, *Strategic Management Journal* 32(2): 301–321. doi: <https://doi.org/10.1002/smj.878>.



- Lyness, K. S., & Heilman, M. E. (2006). When fit is fundamental: Performance evaluation and promotions of upper-level female and male managers. *Journal of Applied Psychology* 91(4): 777–785. doi: <https://doi.org/10.1037/0021-9010.91.4.777>.
- MacNell, L., Driscoll, A., & Hunt, A. N. (2015). What's in a name: Exposing gender bias in student ratings of teaching. *Innovative Higher Education* 40(4): 291–303. doi: <https://doi.org/10.1007/s10755-014-9313-4>.
- Matos, K., O' Neill, O., & Lei, X. (2018). Toxic leadership and the masculinity contest culture: How “win or die” cultures breed abusive leadership, *Journal of Social Issues* 74(3): 500–528. doi: <https://doi.org/10.1111/josi.12284>.
- Mitchell, K., & Martin, J. (2018). Gender bias in student evaluations, *PS: Political Science & Politics* 51(3): 648–652. doi: <https://doi.org/10.1017/S104909651800001X>.
- Moss-Racusin, C. A., Dovidio, J. F., Brescoll, V. L., Graham, M. J., & Handelsman, J. (2012). Science faculty's subtle gender biases favor male students. *Proceedings of the National Academy of Sciences* 109(41): 16474–16479. doi: <https://doi.org/10.1073/pnas.1211286109>.
- Paluda, A., & Strayer, L. A. (1985). What's in an author's name? Differential evaluations of performance as a function of author's name, *Sex Roles* 12(3/4): 353–361. doi: <https://doi.org/10.1007/bf00287601>.
- Paustian-Underdahl, S. C., Walker, L. S., & Woehr, D. J. (2014). Gender and perceptions of leadership effectiveness: a meta-analysis of contextual moderators, *Journal of Applied Psychology* 99(6): 1129–1145. doi: <https://doi.org/10.1037/a0036751>.
- Robertson, L., Galvin, B., & Charles, A. (2007). When group identities matter: bias in performance appraisal, *Academy of Management Annals* 1(1): 617–650. doi: <https://doi.org/10.5465/078559818>.
- Rollinson, D. (2005). *Organisational behavior and analysis: An integrated approach*. Instructor's manual, 3rd edition, London: Prentice Hall.
- Rudman, L. A., & Glick, P. (2001). Prescriptive gender stereotypes and backlash toward agentic women, *Journal of Social Issues* 57(4): 743–762. doi: <https://doi.org/10.1111/0022-4537.00239>.
- Sandström, U., & Hällsten, M. (2008). Persistent nepotism in peer-review, *Scientometrics* 74(2): 175–189. doi: <https://doi.org/10.1007/s11192-008-0211-3>.
- Save the Children (2015). *The urban disadvantage: state of the world's mothers 2015*. Fairfield, CT: Save The Children Federation, Inc. Available at <https://www.savethechildren.org/content/dam/usa/reports/advocacy/sowm/sowm-2015.pdf>. Accessed 06 April 2018.
- Schein, V. E. (1973). The relationship between sex role stereotypes and requisite management characteristics, *Journal of Applied Psychology* 57(2): 95–100. doi: <https://doi.org/10.1037/h0037128>.
- Schein, V. E. (1975). Relationships between sex role stereotypes and requisite management characteristics among female managers, *Journal of Applied Psychology* 60(3): 340–344. doi: <https://doi.org/10.1037/h0076637>.
- Schreier, M. (2012). *Qualitative content analysis in practice*, Thousand Oaks, CA: Sage Publications.
- Schmidt, B. (2015a). Gendered language in teacher reviews. Available at <http://benschmidt.org/profGender/#> Accessed 22 January 2019.
- Schmidt, B. (2015b). Gender bias exists in professor evaluations. *New York Times* on-line article. Available at <https://www.nytimes.com/roomfordebate/2015/12/16/is-it-fair-to-rate-professors-online/gender-bias-exists-in-professor-evaluations> Accessed 22 January 2019.
- Silverman, R. E. (2015). Gender bias at work turns up in feedback. *The Wall Street Journal*, Sept. 30, 2015. Available at <https://www.wsj.com/articles/gender-bias-at-work-turns-up-in-feedback-1443600759>. Accessed 06 April 2018.

- Steinpreis, R. E., Anders, K. A., & Ritzke, D (1999). The impact of gender on the review of the curricula vitae of job applicants and tenure candidates: A national empirical study, *Sex Roles* 41(7/8): 509–528. doi: <https://doi.org/10.1023/A:1018839203698>.
- Statistics Finland (2018). Gender Equality in Finland 2018, Helsinki: Statistics Finland. Available at http://www.stat.fi/tup/julkaisut/tiedostot/julkaisuluettelo/yyti_gef_201800_2018_19723_net.pdf Accessed 17 August 2019.
- Storage, D., Horne, Z., Cimpian, A., & Leslie S-J. (2016). The frequency of “brilliant” and “genius” in teaching evaluations predicts the representation of women and African Americans across fields, *PLoS ONE* 11(3): e0150194. doi: <https://doi.org/10.1371/journal.pone.0150194>.
- Swim, J., & Sanna, L. (1996). He’s skilled, she’s lucky: A meta-analysis of observers’ attributions for women’s and men’s successes and failures, *Personality and Social Psychology Bulletin* 22(5): 507–519. doi: <https://doi.org/10.1177/0146167296225008>.
- Wahl, A. (2010). The impact of gender equality on management and leadership: reflections on change and resistance. In L. Husu, J. Hearn, A.-M. Lämsä & S. Vanhala (eds.), *Leadership through the Gender Lens* (pp. 1–20), Helsinki: Hanken School of Economics research reports.
- Wennerås C., & Wold A. (1997). Nepotism and sexism in peer review, *Nature* 387: 341–343. Doi: <https://doi.org/10.1038/387341a0>.
- World Economic Forum (2018). *Global Gender Gap Report*, Geneva: World Economic Forum. Available at http://www3.weforum.org/docs/WEF_GGGR_2018.pdf Accessed 23 January 2019.