


Article

Challenging Assumptions: Gender, Peer Evaluations, and the Broken Rung in Leadership Trajectories

Saskia L. Shirley and Jennifer Feitosa * 

Department of Psychological Sciences, Claremont McKenna College, Claremont, CA 91711, USA

* Correspondence: jfeitosa@cmc.edu

Abstract: The concept of the ‘glass ceiling’ represents the significant barriers that women face in climbing the corporate hierarchy, but recently, the focus has shifted to the ‘broken bottom rung’, where women are bypassed for initial leadership roles. This paper investigates the impact of gender on performance evaluations, particularly female-to-female peer ratings, which are critical to career progression. Our study tested three hypotheses about the disparity in female allyship within professional contexts. Participants ($N = 160$) from psychology classes in 2018–2019 evaluated their peers in project teams using five ITPMetrics measures. Contrary to previous research suggesting that women receive more critical evaluations than men, this study found no evidence supporting such bias. However, it revealed that women scored higher in process-based skills rather than outcome-based skills, aligning with role congruity theory and the notion of gendered skills. These findings highlight the need for further research into female peer evaluations and their impact on career advancement. This study challenges assumptions about women’s roles in the workplace and advocates for organizations reconsidering the emphasis placed on performance appraisals, proposing alternative assessment methods to foster more equitable and inclusive professional environments.

Keywords: gender; peer ratings; career; leadership



Citation: Shirley, S.L.; Feitosa, J. Challenging Assumptions: Gender, Peer Evaluations, and the Broken Rung in Leadership Trajectories. *Merits* **2024**, *4*, 263–276. <https://doi.org/10.3390/merits4030019>

Academic Editor: Luis Miguel Ciravegna Fonseca

Received: 8 July 2024

Revised: 26 July 2024

Accepted: 31 July 2024

Published: 8 August 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Emerging from a history of suppression and exclusion, women have shown steady progress toward equal representation in traditionally male-dominated fields. Despite this promising trajectory, women are still underrepresented in occupational leadership roles [1]. This has been historically connected to gender beliefs, sex segregation, and overt gender discrimination, with the literature continuing to show evidence of unequal opportunities and exclusion in organizations [2–4]. In connection with the findings of these studies, it is evident that the gender beliefs and stereotypes held by society penetrate the workforce, acting as barriers preventing women from assuming leadership positions [5].

There are promotional discrepancies between men and women, with just 38% of first-level managerial positions being occupied by women, despite comparable numbers of men and women at entering the workforce [6]. These first-level leadership positions act as the bottom rung on the corporate ladder, providing the foundation for career advancement. Thus, this disparity greatly impacts women’s ability to move toward top-level C-suite positions. Without equal representation from the bottom up, women will continue to be underrepresented in high-ranking roles.

By repairing this bottom rung and focusing on greater inclusion efforts, we have the potential to instigate organizational change, generating momentum to propel women toward highly sought-after, top-ranking managerial positions. To achieve this, attention must be given to understanding the obstacles preventing women from assuming first-level leadership roles. Capable and competent female employees are often overlooked for promotions, with the few who are offered managerial roles facing harsher performance reviews and evaluations than their male counterparts [7]. Standing in the way of equal

representation are the risks posed by unconscious biases. The presence of these biases often remains hidden while their impact is simultaneously made visible. While there is substantial evidence to support this ideology, there may be more to the story. Attention is beginning to be paid to the role that women themselves may play in perpetuating inequalities cast by gender stereotypes and biases in the workplace.

The rise of feminism caused a paradigm shift, progressing society socially, culturally, and professionally. According to William O'Neill's 1989 book *Feminism in America: A History*, there was a rapid growth in the number of women employed in the workforce, from representing just one-sixth of workers in 1890 to representing one-third of all employees by 1950 [8]. Women have continued to make professional progress and are being recognized for their potential to have great authority and influence as a collective, especially to empower other women [9,10]. The social phenomenon of 'women supporting women' is now far-reaching and integral to the female identity. The media have described this as the 'Power of Pack', a term recently coined in *Forbes* magazine [11].

Though it is encouraging, this phenomenon does not appear to successfully translate to the workplace. Instead, it seems that the bond created by the shared female identity evaporates in professional environments and the script is flipped. The same women who are allies outside of the workplace suddenly become threatening within it. This can be seen in the distancing of female leaders from their female subordinates, known as the 'Queen Bee' phenomenon, which undermines the mentorship potential of female coworkers [12]. Though women have the potential to harness the power of allyship in a professional environment, creating opportunities for each other to move up the ladder toward positions of influence, it seems that the prospect of collaborative, mutually beneficial success is upset by outdated social ideology and cognitive biases. The disconnect between social and professional allyship is alarming and may cause opportunities for cooperation, collaboration, and mentorship to fade. If rivalry and competition continue within this historically marginalized group, the professional progress of women could be set back by decades.

While this is alarming in any context, its impact is particularly potent following the COVID-19 pandemic. As working environments evolve to suit a post-pandemic world, greater emphasis is being placed on flexibility, adaptivity, and creativity. To capture these specific skill sets at play, 360-degree peer ratings and performance reviews are crucial. 360-degree evaluations are designed to assess a ratee's performance against measures relevant to their role, obtaining the feedback from those in positions above, below, and in line with the ratee. Originally, these 360-degree feedback processes were used solely for the purpose of managerial development, but they have recently been adopted by organizations as performance management mechanisms [13]. Now, more than ever, these evaluation systems will determine promotion eligibility and consideration for leadership positions. Though their importance is evident, research suggests these evaluation systems capture cognitive biases, inhibiting an accurate evaluation process. Research summarized by the authors of *The Handbook of Strategic 360 Feedback* caution against using these tools for anything beyond development, including for promotions [14]. Despite this, organizations are still relying on 360-degree performance appraisals and reviews to determine promotion eligibility.

There is substantial evidence to show that women receive harsher ratings than their male counterparts, despite exhibiting comparable competence, particularly in industries that are not aligned with the social expectations of the female identity [7,15]. Prejudice against women is especially pronounced when raters perceive incongruity between female gender roles and a woman's professional role, with women receiving lower evaluations as occupants of leadership positions, as supported by the findings of one study [16]. Further, results from the same study showed that behavior associated with these leadership roles is perceived less favorably when performed by a woman compared to a man. Considering the overarching findings of these studies, the discrepancy between male and female peer-evaluation systems is apparent. Thus, prejudice assessment tools emerge as one

obstacle preventing women from reaching top-ranking positions at the same rate as their male coworkers.

However, the explanation for this could go beyond traditional gender stereotypes and biases. It may be that women are each other's own worst critics. Instead of the allyship maintained in social settings, perceptions of threat, competition, and rivalry in professional settings may be reflected in intra-gender ratings. That is, it may be that women rate women more harshly than men rate women.

The differential ratings caused by gender-based biases are well supported in the current literature. One study showed that of the 46% of participants who claimed to have a preference for the gender of their boss, men were favored by more than a 2:1 ratio [17]. Interestingly, the same study found that male workers judged their female bosses more favorably, whereas female workers judged their male bosses more favorably. Though this research provides robust evidence to support partiality toward male leaders, there is considerably less evidence to suggest why women in particular show a preference for male bosses over fellow women. Few studies concentrate on relationship dynamics between women at the peer level or in a team context. It is apparent that more research is needed to explore whether intra-gender peer ratings reveal patterns that are consistent with role congruity theory, in which positive evaluations are ascribed to those who behave in alignment with their socially prescribed roles, and the Queen Bee phenomenon [16,18]. Assumptions that women will align themselves with other women may be prohibitive to observing the harmful reality that this misalignment is causing for the career progression of women as a collective. The results of this study have the potential to provide important insights into the barriers that women face that potentially prevent them from assuming highly sought-after, top-ranking managerial positions.

With the above in mind, the purpose of this study is to observe the relationship between gender and peer ratings, particularly those characterized by female-to-female evaluations. This paper aims to explore whether the peer assessment scores of male and female students significantly differ from each other, revealing how the rater's gender may act as a moderator in peer evaluations. The gendering of skills in society has led to the perception of specific traits being aligned with either male or female identities. Building from the theoretical foundation of role congruity theory, gender-assigned skills are rooted in behaviors aligned with each gender's socially prescribed roles. As previously mentioned, role congruity theory explains that positive perceptions and evaluations of people occur when their characteristics and behaviors are consistent with those of their socially prescribed roles [16,18]. In the context of women in the workplace, this theory manifests as less favorable perceptions of women compared to their male counterparts in terms of their performance, ability, and potential, due to inconsistencies with social beliefs about a woman's roles [16]. The effects of role congruity theory cause a misalignment between the identities held by women and the identities of those in the workplace, especially those in leadership positions. Women are socialized to be passive, whereas leaders often need to be active and dominate in an environment. This incongruity prevents women from being seen as having the potential to be successful in male-dominated fields or assume these positions of leadership. This study explores how a ratee's gender evokes high or low ratings in categories supported by, and opposed to, their gender orientation, that is, the observation of higher ratings for women in process-based skills, such as communication or organization, and the inflation of male ratings in outcome-based skills, including knowledge and problem solving. Each of these research questions connects back to the intention of this paper, which is to interrupt current assumptions about women in the workplace so that more appropriate measures may be considered to create equitable and inclusive professional environments.

2. Theoretical Background and Hypothesis Rationale

Socially, women are perceived as more passive, sensitive, and introverted, rewarded for being compliant and polite compared to their dominant, assertive male counterparts [19]. These socially ascribed personality dimensions influence women's behavior in professional

environments, explained by the internalization of gender roles, as reinforced by role congruity theory. This theoretical background leads to the first hypothesis:

Hypothesis 1. *Female team members receive harsher team ratings than their male team members, resulting in lower overall average scores for female participants compared to their male counterparts.*

If it is supported by data from this study and other research, then this hypothesis will cast light on the obstacles faced by women at work. Female employees may experience harsher ratings than their male counterparts, potentially due to the influence of other women. The ‘Queen Bee’ phenomenon describes efforts by women in leadership positions to distance themselves from female subordinates to integrate into male-dominated organizations [12]. This phenomenon, coined in 1973 [20], is a response to the social identity threat, where negative stereotypes about one’s group affect performance [21,22]. Women may give harsher appraisals to female peers, hindering their advancement. This behavior contradicts the call for workplace solidarity [23].

Further, this behavior disrupts the potential for women to be allies to each other. The journey toward a top-ranking position is steep and challenging for all women, but there is an apparent lack of allyship despite this [24]. Without this allyship, women may view each other as competitors rather than partners. This dynamic could contribute to harsher performance appraisals among women, satisfying a protection instinct against perceived threats. Therefore, the second hypothesis of this paper hones in on female-to-female peer ratings:

Hypothesis 2. *Female team members rate their fellow female peers more harshly than male team members rate their female peers.*

Here, the gender of the ratee serves as a moderator that changes the relationship between gender of the rater and peer evaluation. Building upon previous studies and the aforementioned hypothesis, Table 1 presents the expected results of peer ratings, highlighting the interaction between the gender of raters and ratees.

Table 1. Expected results for peer ratings relating to ratee/rater gender interaction.

| | | Ratee | |
|-------|--------|--------------------|---------------------|
| | | Female | Male |
| Rater | Female | Lowest Rating (<M) | Mean Rating (=M) |
| | Male | Mean Rating (=M) | Highest Rating (>M) |

However, the discrepancies between male and female performance appraisals are much more nuanced. Peer ratings reveal patterns of gender-based socialization at work, further manifesting inequities. Through institutional and cultural socialization, particular traits, qualities, and skills are deemed to be more closely aligned with either female or male identities [3]. These authors show how gender segregation and stereotypes illustrate the potential strength of cultural beliefs to determine the ‘gendering’ of particular skills. These patterns are evident in the workplace, particularly in male-dominated occupations, where men are expected to assume leadership roles, while women are often appointed to supportive roles. A recent study highlighted the mark left by these stereotypical views in corporate environments, reflected in compensation gaps and evaluations of female leaders [25].

Gendered skill attribution can be traced back to social beliefs about the nurturing nature of women versus the analytic, quantitative abilities of men. Role congruity theory explains societal expectations for women to excel in nurturing communication and relational

skills, while research shows an aversion to aggression, competition, and risk-taking behaviors among women [26–28]. Conversely, social expectations for men champion aggressive, competitive, and risk-taking behaviors, aligning the male identity with impressions of dominance, ambition, and agency [29,30]. In the workplace, role congruity theory suggests that women are more strongly associated with process-based skills, such as relational management and communication, whereas men are linked with outcome-based skills, like leadership and analytic abilities. In connection with the previous hypotheses, the third hypothesis of this paper is as follows:

Hypothesis 3. *Female team members will be rated higher in process-based skills than in outcome-based skills.*

According to the measures used in this study, process-based skills correspond to ratings in commitment, communication, and emphasis of high standards, and outcome-based skill ratings are built on the foundation of knowledge and focus. To summarize, Figure 1 illustrates our theoretical model; specifically, the gender of the ratee may influence both process- and outcome-based skill ratings, and this relationship is further moderated by the gender of the rater.

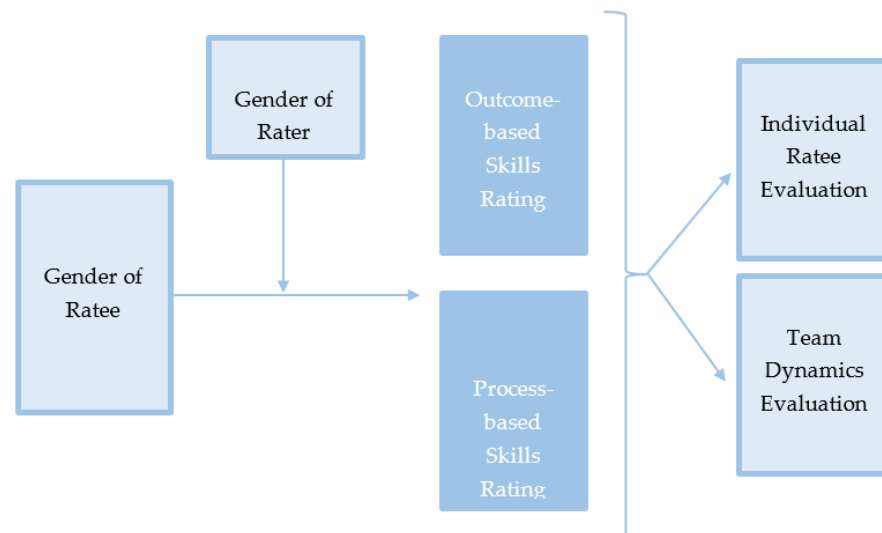


Figure 1. Theoretical Model of the Influence of Gender on Team Performance Evaluations.

3. Method

3.1. Procedure

Data from this study were collected from semester-long team projects at a university on the Eastern Coast of the United States (2018–2019). These projects required teams to lead discussions and present topics, with smaller deliverables due throughout the semester, culminating in a final presentation. After the completion of the team project, students participated in the Individual and Team Performance (ITP) Metrics Survey by Dr. Thomas O’Neill [31]. This online assessment (itpmetrics.com) provides instructions to each team to anonymously rate their peers in five key teamwork competencies (see Table 2 for details). The survey uses a round-robin rating system, where each student rates themselves and their team members multiple times, similar to a study on peer ratings and citizenship behavior [32]. To prevent grade-related bias, data were collected before releasing grades. Peer ratings impacted final grades, simulating 360-degree feedback in professional settings. Personal data, except for gender, were anonymized before analysis.

Table 2. ITP Peer Rating Measures.

| Construct | Items |
|---------------|---|
| Commitment | Takes on a fair share of the team's work. Demonstrates commitment to the team's work. Prepared for team meetings. Keeps deadlines and delivers complete, accurate work. |
| Communication | Communicates clearly and shares information. Exchanges information with teammates in a timely manner. Asks teammates for feedback and uses their suggestions. Seeks appropriate team input before taking action. |
| Capabilities | Acquires new skills or knowledge to improve the team's performance. Learns about other teammates' tasks and roles. Has sufficient knowledge, skills, and abilities to excel in the team's activities. |
| Standards | Encourages and motivates the team. Expresses a belief that the team can do excellent work. Believes that the team will achieve high standards. Cares about the quality of the team's work. |
| Focus | Monitors conditions affecting the team and notices problems. Giving teammates specific, timely, and constructive feedback. Helps the team plan and organize work and anticipates issues. |

3.2. Participants

Peer ratings were collected from nine classes (seven graduate and two undergraduate courses) over two years. Graduate classes included six first-semester mandatory courses and one elective. The total sample comprised 56 teams of three to four members each, averaging 6 teams per class. Students rated themselves and each peer, resulting in 647 individual rating sets, with an average of 3 ratings per ratee, corresponding to 198 raters. After excluding 78 ratings due to incomplete data, 160 raters remained: 95 female and 65 male raters. Exclusions included 18 female raters (28 sets), 10 male raters (14 sets), 4 raters identifying as *none*, and 6 raters who left their gender blank, each due to incomplete ratings or missing ratee gender information.

3.3. Measures

Team members rated each other in five dimensions on a Likert-type five-point scale. Responses were scaled from (1) 'to no extent' to (5) 'a great extent'. Ratees were also given the opportunity to opt out of giving a numerical rating and instead select 'not familiar with team member's behavior'. These peer ratings were conducted against five measures adapted from the key teamwork competencies of the ITP Metrics Survey: *commitment*, *communication*, *capabilities*, *standards*, and *focus* [33]. For each ratee, an average of their scores in each of these measures was calculated to be used in the analyses.

The first dimension focuses on a team member's *commitment* to the team's work. The *communication* measure assesses how well a student connected and conversed with their peers. *Capabilities* refer to the strength of team members' foundation of knowledge, skills, and abilities. The *standards* measure addresses a team member's emphasis on high standards, targeting the level of motivation, execution, and quality of contribution exhibited throughout the project. Finally, the *focus* measure assesses how well a student was able to keep the team on track toward meeting set goals. Each category was broken down into three or four specific behaviors, according to the ITP Metrics tool, which were given to students to use to rate their team members. Table 2 describes the rating sets provided to each participant in the survey.

Though ratees were scored against each of the behaviors within the five dimensions, an average score was given per measure. That is, a score for *communication* was calculated from the average ratings given to each of the behaviors within this category. Thus, individual

reliability ratings could not be calculated for this study. However, support for the reliability of these measures was found in previous studies using the ratings of 30,486 raters, many of which included student teams in higher education, as well as some industry teams [34].

In connection with Hypothesis 3, the five measures were grouped into outcome-based or process-based skills. Outcome-based skills refer to those orientated toward the product of a task, whereas process-based skills refer to behaviors essential to the progress toward or procedure of completing a task. The behaviors outlined in the standards and capabilities measures are connected to outcome-based skills. *Capabilities* and *standards* indicate greater focus on the result of a project and thus were grouped together. *Communication*, *commitment*, and *focus* are more closely aligned with process-based skills, demonstrating behaviors that are orientated toward making progress toward an outcome rather than generating the outcome itself. For the items assigned to each measure according to the ITP Metrics survey, Table 3 depicts the skill groupings.

Table 3. Measures Categorized into Outcome-Based and Process-Based Skills.

| | Outcome-Based | Process-Based |
|----------|---------------------------|--------------------------------------|
| Measures | Capabilities Standards | Communication Commitment Focus |

3.4. Data Analysis

Peer rating scores were sorted into three datasets prior to analysis. The first set included the ratee's gender and corresponding average calculated from their scores against all five measures. This dataset was used during analyses to test Hypothesis 1 in this paper. To test this hypothesis, an independent samples *t*-test was performed to compare the overall mean score for female participants to that of male participants. An average was taken from scores in each of the five measures per student to give an overall performance score, which was included in the analysis. All 160 participants were included in this analysis ($n = 160$), with 95 being female, and 65 male. The second dataset included a similar average taken from each measure; however, it included only the scores for the female participants in the study, as well as the corresponding gender of the rater responsible for assigning these scores. The outcomes for these female participants were used to test Hypothesis 2. To test this, another independent samples *t*-test was conducted. The analysis compared the average rating score that female participants received from their male peers to the average rating score that men received from their female peers. The ratings of all 95 female participants were included in this analysis. In total, 55 male participants and 40 female participants rated the 95 female team members, with female self-ratings excluded from this analysis. The final dataset used to conduct analyses connected to Hypothesis 3 included the gender of the ratee, as well as individual scores for each of the five measures for all 160 participants. To test Hypothesis 3, a two-way mixed ANOVA test was performed. From these data, an average score was calculated for each participant relating to the process-based and outcome-based skill groupings.

4. Results

Hypothesis 1 predicted that female team members would receive harsher team ratings than male team members. That is, female participants would garner lower overall average scores compared to the overall average scores of male participants. Evidence was not found to support this hypothesis. Female team members were not rated more harshly ($M = 4.46$, $SD = 0.48$) compared to their male counterparts ($M = 4.38$, $SD = 0.63$), $t(158) = 0.82$, $p = 0.21$, Cohen's $d = 0.55$. Therefore, it cannot be concluded that any difference exists between the ratings of male and female participants in this study.

Hypothesis 2 theorized that female team members would rate their fellow female peers more harshly than male team members would rate their female peers. Similarly,

evidence was not found to support this hypothesis. Female team members did not receive significantly lower ratings from their female peers ($M = 4.37, SD = 0.46$) than from their male peers ($M = 4.41, SD = 0.72$), $t(93) = 0.34, p = 0.37$, Cohen's $d = 0.62$. Thus, there is insufficient evidence to conclude that female team members rate other female team members differently than their male counterparts (see Table 4).

Table 4. Descriptive Statistics of Peer Ratings in Male and Female Team Members.

| Rater Gender | Gender of Ratee | | | |
|---------------------|-----------------|-----------|-------------|-------------|
| | Female | | | |
| | <i>M</i> | <i>SD</i> | <i>Min.</i> | <i>Max.</i> |
| Male ($n = 55$) | 4.41 | 0.72 | 1.8 | 5 |
| Female ($n = 40$) | 4.37 | 0.46 | 2.8 | 5 |
| Total | 4.39 | 0.59 | 1.8 | 5 |

Finally, Hypothesis 3 expected female team members to be rated highly in process-based skills than in outcome-based skills. According to the measures used in this study, this corresponds to high ratings in commitment, communication, and emphasis of high standards, and low ratings in the foundation of knowledge and focus. The results indicated there was no significant main effect of skill on overall scores ($F(1,158) = 2.99, p = 0.086, \eta_p^2 = 0.19$). However, there was a significant interaction between skill and gender in terms of average scores ($F(1,158) = 8.41, p = 0.004, \eta_p^2 = 0.51$), such that male team members scored higher in skills that were outcome-based and female team members scored higher in skills that were process-based (see Figure 2). For this test, the assumption of homogeneity of variances was not met. This is further explored in the discussion of the limitations of the study.

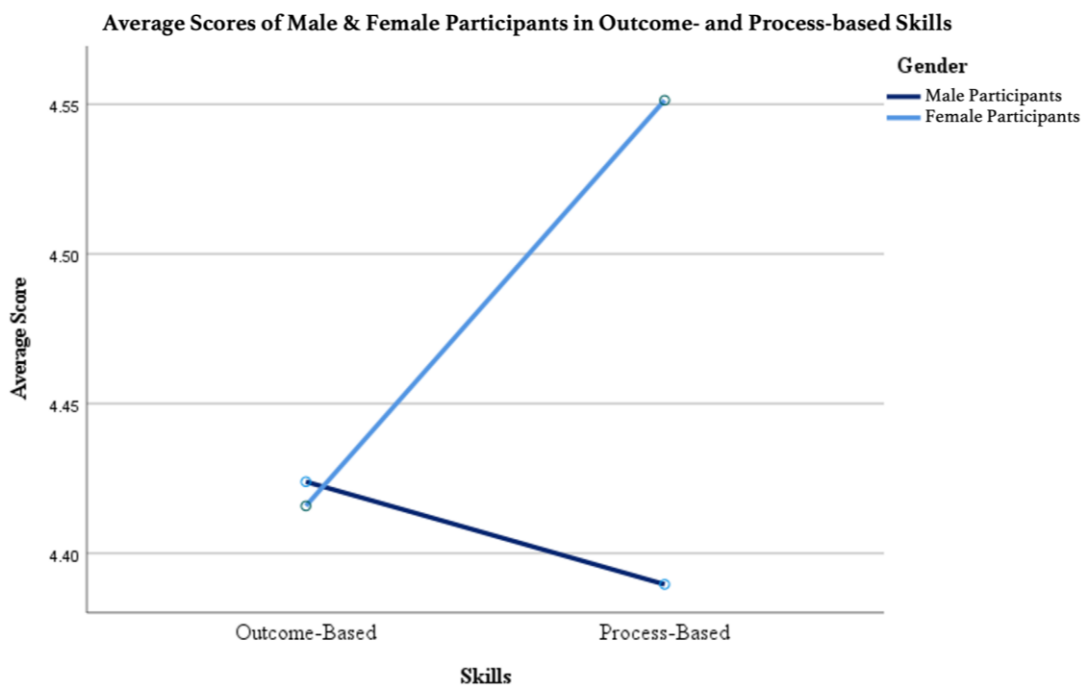


Figure 2. Average Scores of Male and Female Participants in Outcome- and Process-Based Skills.

On average, participants scored higher in process-based skills ($M = 4.49, SD = 0.45$) compared to outcome-based skills ($M = 4.42, SD = 0.44$). A further breakdown of these categories showed that for outcome-based skills, the average scores of male participants

($M = 4.42, SD = 0.56$) and female participants ($M = 4.42, SD = 0.34$) were extremely close. However, for process-based skills, female participants scored slightly higher ($M = 4.55, SD = 0.33$) than male participants ($M = 4.39, SD = 0.57$). While male participants scored higher in outcome-based skills ($M = 4.42$) than in process-based skills ($M = 4.39$), female participants showed the opposite pattern, scoring higher in process-based skills ($M = 4.55$) compared to outcome-based skills ($M = 4.42$), showing support for Hypothesis 3 and role congruity theory (see Table 5).

Table 5. Descriptive Statistics for Skill-Based Ratings in Male and Female Participants.

| | | Skills | | | | | | | |
|--------|--------------|---------------|------|------|------|---------------|------|------|------|
| | | Outcome-Based | | | | Process-Based | | | |
| Gender | | M | SD | Min. | Max. | M | SD | Min. | Max. |
| | | Male | 4.42 | 0.56 | 2.63 | 5 | 4.39 | 0.57 | 2.5 |
| | Female | 4.42 | 0.34 | 3.34 | 5 | 4.55 | 0.33 | 3.66 | 5 |
| | Total | 4.42 | 0.44 | 2.63 | 5 | 4.49 | 0.45 | 2.5 | 5 |

5. Discussion

This study explored the nature of performance evaluations characterized by rater and ratee gender in student teams to observe whether ratee gender identity acts as a moderator of the relationship between rater gender and assessment of peers. The hypotheses particularly directed attention to the dynamics of female-to-female peer ratings, providing further insights into the relationship between women in teams. Finally, continuing the research into gendered-skills, this study investigated how a ratee’s gender interacts with perceptions of their performance in areas aligned with and opposing their gender identity.

More specifically, the results of this study align with previous findings on gender-based socialization and role congruity theory, revealing nuanced patterns in peer ratings that reflect institutional and cultural influences. The data show that male participants score more highly in outcome-based skills compared to process-based skills. Conversely, female participants scored more highly in process-based skills than in outcome-based skills. This trend supports the notion that traits and skills are culturally gendered, as articulated by Donley and Baird [3].

Interestingly, the first two hypotheses were not supported. Despite sufficient evidence to show gender prejudice in performance appraisals from previous studies [16,18], we did not have enough evidence to support the idea that female team members would receive harsher team ratings than their male counterparts or that female team members would rate their female peers more harshly than male team members. It is possible that a pattern of lower female ratings was not detected in this study, despite being found in others, for a few reasons. Firstly, such close overall average scores between male and female participants suggest that the study’s participant sample may have influenced this particular pattern of results. Over three quarters of the student data collected were sourced from graduate classes; therefore, it is important to consider that the participants all had a great incentive to work hard and perform well to achieve a desired grade. Beyond this, though the majority of the selected courses were mandatory classes, it is possible that students who elected to study psychology at the graduate level were already primed to understand rating biases. Alternatively, it could also be that the strength of gender’s influence is lower among students compared to other populations, though there is ample evidence in the literature to suggest that gender biases are still present in college students [35,36]. Considering that gender biases are expected to exist among students and influence academic peer ratings, it is likely that these biases were not effectively captured by the assessment measures; thus, they were not detected among the study’s participants. However, additional research would need to be conducted to confirm, support, or negate either of these possible explanations. Further interpretation is discussed in the limitations of this study.

Furthermore, it is possible that this finding did not support the Queen Bee phenomenon because of the team structure. Participants were not prescribed hierarchical positions within their teams; thus, there is not the same power dynamic that exists in settings where the Queen Bee phenomenon is observed. As main effects for these outcome-based skills, male and female participants had, on average, virtually the same scores. Interestingly, this was the category predicted to show the most difference. It may be that the participant sample was biased toward high-achieving students and therefore reflected high levels of capability and standards. Supported by role congruity theory and stereotype threat, an alternative explanation is that female participants are indeed bringing in more knowledge and focus but their ratings are lower than their true performance because they are being compared against traditional gender roles for outcome-based skills that tend to favor male versus female participants. In terms of process-based skills, female participants scored slightly higher than male participants, with a small difference between the two averages. Considering the participants of the study, the path toward graduate school requires a high level of self-efficacy and sophistication in process-based skills, particularly as psychology majors, translating to elevated scores in this category [37,38]. An alternative explanation could be that men are showing higher levels of process-based skills but women are being favored due to their nurturing and communal stereotypes.

The findings of this study show an overall pattern that indicates that the gender of both the rater and ratee is complex in determining peer evaluations, as it relates to a given skill or category. These null results may provide further evidence for how performance evaluations are often biased, particularly in ambiguous situations, with women needing more evidence of competence to be rated equally [39,40]. This suggests that gender acts as a moderator during the rating process, as shown previously in Figure 1.

5.1. Limitations

Though efforts were made to mitigate the weaknesses and shortcomings of this study, there are limitations of its design and analyses. In terms of participants, there was an unequal number of male and female participants in the study. This equates to a gender breakdown of 59.4% female and 40.6% male. With gender playing such an important role in the study, a more even breakdown of participants would have been optimal, especially when determining overall performance ratings in response to Hypothesis 1. Thus, it is possible that the study's gender profile could help to explain why effects of previous studies showing lower ratings in female participants than male participants were not supported by the results of this study.

As referenced by the results of the Hypothesis 3 analysis, the assumption of homogeneity of variances was not met. This indicates that the spreads of the two populations, i.e., male and female ratees, were not considered equal. As a result, the populations of this study must be considered heterogeneous. However, ANOVA tests are, by nature, robust and are not overly influenced by small violations of assumptions. Nevertheless, this is a limitation of the study.

In addition, regarding limitations in analyses, it is important to note that average scores were used as the primary data rather than raw scores. Data were collected by collapsing raw scores for each measure into an average. Considering that these data were drawn from a series of classes spanning years, it was not possible to locate original raw scores. Because of this, reliabilities for each of the five measures could not be conducted. While this is a limitation, the ITP Metrics Assessment tool has been used numerous times in other studies, with support for the reliability of these measures coming from the scores of 30,486 raters, as mentioned previously [34].

Continuing to explore the study's participant sample, a possible limitation arises in the nature of the teams observed. With the use of student teams rather than employee teams, it is important to consider the limited extent to which results can be generalized to a professional setting. Though student teams function similarly to project teams in the

workplace, their outcomes differ in a few important ways, posing potential obstacles in extrapolating the findings to this study from an academic environment to a professional one.

The first difference is seen in how the rating system itself is used in conjunction with the outcomes and rewards of a team. For students, the desired outcome is a certain grade which many, if not all, students can achieve if they earn it. In a professional environment, promotions and upward mobility are the desired outcomes of performance appraisals, in which only a handful of employees will be elevated to a select number of managerial positions. Therefore, the same competition that exists in a corporate setting is not at play in a classroom environment. It is plausible that the effects found in this study would be more pronounced in an environment in which competition is also more pronounced.

The second difference between student and employee teams is related to the nature of their assembly and timeline. Student teams are often temporary with short lifecycles; the depth of relationship, trust, and collective efficacy between team members are much shallower than in an employee team that has worked together over many years or across many projects. There is also a relational aspect to this limitation. For students, the short timeline of project teams may influence the amount of attention paid to resolving problems. For most students, the cost/benefit of addressing an underperforming team member is not worth the effort, especially without the incentive of future teamwork with this peer [41]. However, for an employee team, resolving conflict is a much higher priority in the interest of protecting future working relationships.

5.2. Future Research

In response to the limited sample in this study, future research should focus on whether the results of this paper are replicated in professional working teams. Drawing from the role congruity theory and Queen Bee phenomenon, both highly applicable to corporate contexts, a similar analysis of peer ratings with consideration of gender would be important in advancing our theoretical understandings of gender prejudice.

Alternatively, research focusing on the breakdown of outcome-based and process-based skills would give further insight into how gendered skills influence peer ratings. The interaction found between gender and skills suggests that there is a relationship between a rater's gender and their scores in particular skills. However, with some overlap between the measures and, therefore, skill groupings, as identified previously, further research is necessary to distinguish between behaviors considered to be more 'masculine' or 'feminine'. A behavioral anchored rating scale (BARS) could be used to observe these behaviors and compare them to the subjective ratings. The extent to which perception and real performance converge is important to explore.

Even if the findings of this study showed support for female team members receiving harsher performance evaluations from their female peers compared to their male peers, the reasoning behind this phenomenon remains uncertain. It may be that role congruity theory offers an explanation, specifically, the subsequent incongruity between the female identity in academic and in professional roles. Alternatively, it may be that women perceive each other as threats, therefore rating each other more harshly to account for their competition. As discussed in connection to social identity threat theory, it could also be that women underperform in response to negative stereotypes about women in the workforce. Beyond this, it may be that female coworkers who observe this underperformance ascribe lower peer ratings as a means to distance themselves from women who fulfill these negative stereotypes. Yet another perspective could argue that this effect is entirely unconscious and is instead an indication of the cycle of socialization, in which women are socialized to be more passive, sensitive, and introverted. As an extension, lower female-to-female ratings could be a result of misalignment between these social female identities and the measures included in the rating tool. There could be many other alternative explanations for the results of the study; thus, it is apparent that further research is necessary.

6. Conclusions

This study contributes to the understanding of gender dynamics at play in 360-degree feedback assessments, specifically in relation to their influence on the perception of gendered skills. The results showed that male participants tend to score more highly in outcome-based skills and female participants tend to score more highly in process-based skills, aligned with the behaviors socially ascribed to the female identity. These patterns reflect broader cultural and institutional influences consistent with role congruity theory and gender-based socialization. However, contrary to previous research, we did not find evidence that female team members receive harsher team ratings or rate their female peers more harshly. This discrepancy could be due to the unique characteristics of our sample, such as their academic background and motivation, or the specific context of the study. Additionally, the absence of hierarchical team structures might have mitigated the Queen Bee phenomenon. The interaction shown between skill type and gender, specifically process-based skills, provides a foundation for future research to address the disparity between female allyship in social and in professional environments.

On a practical level, this paper calls for organizations and educational institutions to be aware of potential biases in performance evaluations and to develop training programs that address gender stereotypes. Implementing more structured and objective assessment criteria could help mitigate these biases. Theoretically, this study supports the notion that gender-based socialization and role congruity theory play significant roles in shaping peer evaluations, but it also suggests that these influences may vary depending on the context and population. Future research is needed to explore these dynamics in different contexts and to develop more robust ways to capture subtle gender biases in performance evaluations. Overall, these results underscore the complexity of gender as a factor in peer evaluations and suggests that biases may not be uniformly present across all settings.

Author Contributions: Conceptualization, S.L.S.; methodology, S.L.S.; software, J.F.; validation, J.F.; formal analysis, S.L.S.; investigation, S.L.S.; resources, J.F.; data curation, S.L.S. and J.F.; writing—original draft preparation, S.L.S.; writing—review and editing, J.F.; visualization, S.L.S.; supervision, J.F.; project administration, J.F.; funding acquisition, J.F. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki, and approved by the Conjoint Faculties Research Ethics Board (CFREB), University of Calgary (protocol code: REB19-0821 and 22-August-2019).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Data is unavailable due to privacy considerations.

Acknowledgments: The authors would like to thank Tom O'Neill for the data collection tool ITPMetrics and availability of the data of this research.

Conflicts of Interest: The author declares no conflict of interest.

References

1. Bonner, R.L. Organizational and environmental determinants of female representation in top management teams. *ProQuest Inf. Learn. Diss. Abstr. Int. Sect. A Humanit. Soc. Sci.* **2019**, *80*, 1-A(E).
2. Özbilgin, M.F. Gender and Jobs: Sex Segregation of Occupations in the World—By Richard Anker. *Gen. Work. Organ.* **2007**, *14*, 502–504. [[CrossRef](#)]
3. Donley, S.; Baird, C.L. The Overtaking of Undertaking?: Gender Beliefs in a Feminizing Occupation. *Sex Roles* **2017**, *77*, 97–112. [[CrossRef](#)]
4. Georgeac, O.; Rattan, A. Progress in women's representation in top leadership weakens people's disturbance with gender inequality in other domains. *J. Exp. Psychol. Gen.* **2019**, *148*, 1435–1453. [[CrossRef](#)] [[PubMed](#)]
5. Schmitt, M.T.; Ellemers, N.; Branscombe, N.R. Perceiving and responding to gender discrimination in organization. In *Social Identity at Work: Developing Theory for Organization Practice*; Haslam, S.A., van Knippenberg, D., Platow, M.J., Ellemers, N., Eds.; Psychology Press: London, UK, 2003; pp. 227–292.

6. McKinsey & Company. Women in the Workplace 2021. Available online: <https://www.mckinsey.com/featured-insights/diversity-and-inclusion/women-in-the-workplace> (accessed on 2 November 2021).
7. Cecchi-Dimeglio, P. How Gender Bias Corrupts Performance Reviews, and What to Do about It. *Harv. Bus. Rev.* **2017**, *12*, 2017. Available online: <https://hbr.org/2017/04/how-gender-bias-corrupts-performance-reviews-and-what-to-do-about-it> (accessed on 12 April 2017).
8. O'Neill, W.L. *Feminism in America: A History*, 2nd ed.; Routledge: London, UK, 1989.
9. Richardson, R.A. Measuring Women's Empowerment: A Critical Review of Current Practices and Recommendations for Researchers. *Soc. Indic. Res.* **2018**, *137*, 539–557. [[CrossRef](#)]
10. Warrell, M. *Women Rising: Internal Facilitators to Lead from the Top*; ProQuest Dissertations & Theses: Ann Arbor, MI, USA, 2021.
11. Zalias, S. Power of the Pack: Women Who Support Women. *Forbes Magazine*, 6 March 2019. Available online: <https://www.forbes.com/sites/shelleyzalis/2019/03/06/power-of-the-pack-women-who-support-women-are-more-successful/> (accessed on 2 November 2021).
12. Derks, B.; Van Laar, C.; Ellemers, N. The queen bee phenomenon: Why women leaders distance themselves from junior women. *Leadersh. Q.* **2016**, *27*, 456–469. [[CrossRef](#)]
13. Champion, E.D.; Champion, M.C.; Champion, M.A. Best practices when using 360 feedback for performance appraisal. In *The Handbook of Strategic 360 Feedback*; Church, A.H., Bracken, D.W., Fleenor, J.H., Rose, D.S., Eds.; Oxford University Press: Oxford, UK, 2019; pp. 19–59.
14. Bracken, D.W.; Dalton, M.A.; Jako, R.A.; McCauley, C.D.; Pollman, V.A.; Hollenbeck, G.P. Should Degree Feedback Be Used Only for Developmental Purposes? Center for Creative Leadership: Greensboro, NC, USA, 1997.
15. Garcia-Retamero, R.; López-Zafra, E. Prejudice against women in male-congenial environments: Perceptions of gender role congruity in leadership. *Sex Roles A J. Res.* **2006**, *55*, 51–61. [[CrossRef](#)]
16. Eagly, A.H.; Karau, S.J. Role congruity theory of prejudice toward female leaders. *Psychol. Rev.* **2002**, *109*, 573–598. [[CrossRef](#)] [[PubMed](#)]
17. Elsesser, K.M.; Lever, J. Does gender bias against female leaders persist? Quantitative and qualitative data from a large-scale survey. *Hum. Relat.* **2011**, *64*, 1555–1578. [[CrossRef](#)]
18. Gervais, S.J.; Hillard, A.L. A Role Congruity Perspective on Prejudice Toward Hillary Clinton and Sarah Palin. *Anal. Soc. Issues Public Policy* **2011**, *11*, 221–240. [[CrossRef](#)]
19. Mirkin, H. The passive female the theory of patriarchy. *Am. Stud.* **1984**, *25*, 39–57.
20. Staines, G.; Travis, C.; Jayerante, T.E. The Queen Bee Syndrome. *Psychol. Today* **1973**, *7*, 55–60.
21. Martiny, S.; Nikitin, J. Social Identity Threat in Interpersonal Relationships: Activating Negative Stereotypes Decreases Social Approach Motivation. *J. Exp. Psychol. Appl.* **2019**, *25*, 117–128. [[CrossRef](#)] [[PubMed](#)]
22. Steele, C.M.; Spencer, S.J.; Aronson, J. Contending with group image: The psychology of stereotype and social identity threat. *Adv. Exp. Soc. Psychol.* **2002**, *34*, 379–440. [[CrossRef](#)]
23. Mavin, S. Queen Bees, Wannabees and Afraid to Bees: No More 'Best Enemies' for Women in Management? *Br. J. Manag.* **2008**, *19*, S75–S84. [[CrossRef](#)]
24. Platell, A. Why are women so awful to each other? In *Daily Mail*; DMG Media: London, UK, 2004; p. 15.
25. Wang, J.C.; Markóczy, L.; Sun, S.L.; Peng, M.W. She'-E-O Compensation Gap: A Role Congruity View. *J. Bus. Ethics* **2019**, *159*, 745–760. [[CrossRef](#)]
26. Chen, G.L.; Crossland, C.; Huang, S. Female board representation and corporate acquisition intensity. *Strateg. Manag. J.* **2016**, *37*, 303–313. [[CrossRef](#)]
27. Chen, S.; Ni, X.; Tong, J.Y. Gender diversity in the boardroom and risk management: A case of R&D investment. *J. Bus. Ethics* **2016**, *136*, 599–621.
28. Ho, S.S.M.; Li, A.Y.; Tam, K.; Zhang, F. CEO gender, ethical leadership, and accounting conservatism. *J. Bus. Ethics* **2015**, *127*, 351–370. [[CrossRef](#)]
29. Oakley, J.G. Gender-based barriers to senior management positions: Understanding the scarcity of female CEOs. *J. Bus. Ethics* **2000**, *27*, 321–334. [[CrossRef](#)]
30. Nekhili, M.; Chakroun, H.; Chtioui, T. Women's leadership and firm performance: Family versus nonfamily firms. *J. Bus. Ethics* **2016**, *153*, 291–316. [[CrossRef](#)]
31. O'Neill, T.A.; Deacon, A.; Gibbard, K.; Larson, N.; Hoffart, G.; Smith, J.; Donia, B.L.M. Team Dynamics Feedback for Post Secondary Student Learning Teams. *Assess. Eval. High. Educ.* **2018**, *43*, 571–585. [[CrossRef](#)]
32. Schmidt, J.A.; O'Neill, T.A.; Dunlop, P.D. The Effects of Team Context on Peer Ratings of Task and Citizenship Performance. *J. Bus. Psychol.* **2021**, *36*, 573–588. [[CrossRef](#)]
33. O'Neill. ITP Metrics Assessment Tool [Review of ITP Metrics Assessment Tool]. 2011. Available online: <https://www.itpmetrics.com/assessment.info> (accessed on 24 October 2021).
34. O'Neill, T.A.; Larson, N.; Smith, J.; Deng, C.; Donia, M.; Rosehart, W.; Brennan, R. Introducing a scalable peer feedback system for learning teams. *Assess. Eval. High. Educ.* **2019**, *44*, 848–862. [[CrossRef](#)]
35. Beyer, S. Gender differences in causal attributions by college students of performance on course examinations. *Curr. Psychol. A J. Divers. Perspect. Divers. Psychol. Issues* **1998**, *17*, 346–358. [[CrossRef](#)]

36. Gutierrez, A.P.; Price, A.F. Calibration between undergraduate students' prediction of and actual performance: The role of gender and performance attributions. *J. Exp. Educ.* **2017**, *85*, 486–500. [[CrossRef](#)]
37. Strohmetz, D.B.; Dolinsky, B.; Jhangiani, R.S.; Posey, D.C.; Hardin, E.E.; Shyu, V.; Klein, E. The skillful major: Psychology curricula in the 21st century. *Scholarsh. Teach. Learn. Psychol.* **2015**, *1*, 200–207. [[CrossRef](#)]
38. Chenneville, T.; Gay, K. Promoting writing as a core competency for psychology majors: Challenges and opportunities. *Scholarsh. Teach. Learn. Psychol.* **2024**, *10*, 191–206. [[CrossRef](#)]
39. Lyness, K.S.; Heilman, M.E. When fit is fundamental: Performance evaluations and promotions of upper-level female and male managers. *J. Appl. Psychol.* **2006**, *91*, 777–785. [[CrossRef](#)] [[PubMed](#)]
40. Smith, D.G.; Rosenstein, J.E.; Nikolov, M.C.; Chaney, D.A. The power of language: Gender, status, and agency in performance evaluations. *J. Bus. Psychol.* **2019**, *34*, 469–485. [[CrossRef](#)]
41. Abbasi, N.; Mills, A.; Tucker, R. Conflict Resolution in Student Teams: An Exploration in the Context of Design Education. In *Collaboration and Student Engagement in Design Education*; IGI Global: Pennsylvania, PA, USA, 2017; pp. 105–124. [[CrossRef](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.