Gender Gap in Performance Evaluation. Are Women Stricter?

*Veronika HEDIJA**

Abstract

This study examines the presence of gender bias in performance evaluations, with a particular focus on whether assessments differ based on the gender of the actors. Using the sample of Czech economics-oriented university students, the research investigates whether gender influences students' assessments of seminar paper presentations. A binary logistic regression model is employed to analyze the relationship between evaluator and student gender and the likelihood of receiving the highest grade. The results indicate that female evaluators are generally stricter than male, significantly reducing the probability of a high assessment. However, there is no evidence that male students receive more favorable evaluations than their female counterparts, nor that individuals favor those of the same gender.

Keywords: *gender, gender gap, evaluation, performance*

JEL Classification: J16, J71, M52, M54

DOI: https://doi.org/10.31577/ekoncas.2025.03-04.03

Article History: *Received:* July 2024 *Accepted:* July 2025

Introduction

Performance evaluation is a cornerstone of effective human resource management and plays a crucial role in determining career progression, remuneration, and opportunities for advancement. As such, fairness and objectivity in performance assessments are essential for both individual motivation and organizational success.

* Veronika HEDIJA, corresponding author, Masaryk University, Faculty of Social Studies, Department of Social Policy and Social Work, Joštova 218/10, 602 00 Brno, Czech Republic; e-mail: Veronika.Hedija@fss.muni.cz, ORCID: 0000-0002-5353-1505

However, numerous studies have shown that performance evaluations may be influenced by implicit biases, including those related to gender, which can undermine meritocratic principles and perpetuate inequalities in the workplace (Mengel et al., 2019; Hospido and Sanz, 2021; Diederich et al., 2023).

The findings of empirical studies that are devoted to the assessment of women's performance often conclude that women are evaluated worse than men (for example Balcar and Hedija, 2019; Mengel et al., 2019; Hospido and Sanz, 2021; Jansson and Tyrefors, 2022). Gender stereotypes and prejudices, which are the source of discrimination against women in the labor market, are often cited as the main reason for the poorer evaluation of women's performance. However, stereotypes, gender roles, and social norms are changing over time and are country-specific (Goldin, 2006; Ochsenfeld, 2014). One particularly important dimension of gender-related bias in performance assessment concerns the role of the evaluator. Research suggests that women and men may apply different assessment standards, and that the gender of the evaluator can significantly affect outcomes (Balcar and Hedija, 2019; Ahmed et al., 2021). These patterns are of particular concern in managerial roles, where evaluation and feedback are central to decision-making about hiring, promotions, and pay. If future managers carry implicit biases into their evaluations, it can systematically distort talent recognition and employee development.

This paper deals with the issue of gender bias in performance evaluations. The aim of the paper is to contribute to the understanding of gender bias in performance evaluation by examining whether assessments differ depending on the gender of both the evaluator and the evaluated in Czech environment. The study draws on a sample of male and female students from a Czech university with an economic focus, who evaluated the seminar presentations of their peers. A regression model is employed to identify potential differences in assessment outcomes based on the gender of both the evaluator and the evaluated. The findings indicate that performance assessment is influenced by the gender of the evaluator. The women were generally stricter in their assessments compared to men. However, it was not confirmed that men receive better evaluations than women or that individuals tend to favor those of the same gender.

This study brings new insights on several levels. Firstly, by focusing on university students of business economics and management, it provides information about gender differences in evaluative behavior among the upcoming generation of professionals, which will soon enter the labor market and what is their attitude at the beginning of their working career. Secondly, the focus on a Czech environment fills a gap in the literature, which has so far largely overlooked Central and Eastern European contexts, despite their distinct institutional and cultural characteristics.

Finally, provides information that the evaluation of individuals can be influenced by the gender of the evaluator. The study's findings showed that women were stricter than men when evaluating the work of their colleagues.

1. Literature Review

Relatively great attention is paid to the issue of evaluation of men and women. This is an issue on the border of economics, psychology, and sociology and it can be viewed from different perspectives. Firstly, from the point of view of equal opportunities for men and women and the equality in evaluation of their performance. Here, the research aims to find out whether the approach to men and women is the same and whether men and women are simply equally valued for the same performance. Secondly, the issue can also be viewed in terms of how women and men evaluate. Here, it is examined whether men and women rate on average in the same way or whether their ratings differ.

Individuals do not have to be evaluated the same for the same performance. Their assessment may depend on many factors, and gender may be one of them. The gender of the evaluator may also play a role.

The issue of evaluation of individuals and the explanation of their approach to assessment can be viewed from the perspective of various theories. Discrimination theory explains inequalities in the treatment of individuals, differences in remuneration, and the evaluation of equal performance by the existence of discrimination. Becker's (1957) theory of taste for discrimination and the model of statistical discrimination, authored by Arrow (1973) belong to the most commonly used theories of discrimination. According to these theories, the evaluation of individuals is determined by the environment in which individuals are assessed, and is a reflection of the state of society. According to these theories, the reasons for the different evaluations of individuals are the personal preferences of individuals and stereotypes in society. From the point of view of evaluating the performance of men and women, worse score of women compared to men can be explained mainly by assessing female abilities based on stereotypes, when women are perceived as less efficient compared to men. For this reason, they are also a priori worse rated due to the expected lower performance. On the other hand, the assessment of individuals may depend on the gender of the evaluator. Taste for discrimination against women or assessing female abilities based on stereotypes could be lower in the case the evaluator is a woman.

Sociological theories offer arguments in the form of existing gender stereotypes and gender roles determining the position of men and women in society and influencing their behavior (de Beauvoir, 2001; Oakley, 1972; Connell, 1995).

Women are assigned the role of mother and caretaker of the family men are seen as the breadwinner. In the same way, women and men are attributed typical feminine and masculine characteristics and are viewed from this perspective. Existing gender stereotypes and societal expectations regarding the „typical" woman and man can significantly influence the evaluation of performance. Women are often perceived as less competent, less psychologically resilient, and less efficient than men. As a result, identical performance may be rated more negatively when exhibited by a woman. Over time, women may become accustomed to receiving harsher evaluations, which in turn may lead them to adopt a more critical stance when rating others, in contrast to men, whose performance is typically assessed with greater leniency.

From the point of view of socio-psychological theories like the theory of social identity and similarity – attraction theory, performance assessment may be influenced by the gender of the evaluator. According to the theory of social identity, individuals tend to better evaluate members of the group to which they belong against other groups' members (Tajfel, 1982; Tajfel and Turner, 1979). Thus, this theory postulates that men will tend to subjectively better evaluate the performance of men compared to women, and women, on the contrary, the performance of women. The similarity – attraction paradigm claims that similar individuals are attracted to one another (Byrne, 1971).

Hence, women could better remunerate women and vice versa. Contrary, the self-enhancement drive attitude postulates that individuals tend to identify themselves with positively viewed groups. Individuals from lower-status groups tend to identify themselves with members of a higher-status group to promote a positive identity (Burkley and Blanton, 2005). According to this theory, women may identify themselves with a higher-status group of men and may follow their behavior.

The role of stereotypes, gender roles, and social norms and their changes in time and the effect on gender inequalities are also discussed in the literature (Goldin, 2006; Ochsenfeld, 2014). A higher level of education for women and higher representation of women in important social positions may change traditional gender roles, norms, and stereotypes and eliminate the taste for discrimination or the effect of statistical discrimination.

Economically oriented studies dealing with this issue examine most commonly the existence of the pay gap between men and women and analyze their causes, differences in labor market positions, and the likelihood of success in competitions or, in general, performance evaluations of men and women (for example Hedija, 2017; Balcar and Hedija, 2019; Ahmed et al., 2021; Hospido and Sanz, 2021; Adamus and Ballova Mikuskova, 2024; Johnson Filipová et al., 2024).

Hedija (2017) estimated the unexplained part of the gender pay gap in European countries and identified the existence of unexplained pay differences between men and women and these vary within individual sectors. However, the study only works with rough characteristics of men and women (age, level of education, sickness, partnership, occupation, sector, company size, contract, and managerial position) and does not capture all knowledge and skills that could explain wage differences and capture possible differences in productivity. Balcar and Hedija (2019) try to overcome these limits. They estimate the wage differences between men and women in the Czech Republic using highly detailed data containing in addition to common characteristics such as age, occupation, education also the data on individuals' cognitive and non-cognitive skills and their personal preferences. Taking into account the detailed characteristics of the employees, the study shows that women and men are evaluated differently without objective reasons. They also examined the effect of manager gender on wages and concluded that the gender of the manager affects the level of wages. Employees under female managers face a −6.30 percent wage penalty. This may indicate women's tendency to evaluate more strictly than men. Diederich et al. (2023) examine to what extent the gender pay gap at the top executive level is linked to gender stereotypes using the data of large European companies. They identified the unexplained gender pay gap at the executive level and found out that women are paid better when they hold less masculine and therefore more role-corresponding functions. Hospido and Sanz (2021) take a different approach to examining gender differences in evaluation and identifying potential discrimination against women. They studied the assessment of women's and men's submissions to economics conferences and identified the statistically significant differences in acceptance rate. They concluded that all-female-authored papers are less likely to be accepted than all-male-authored ones and suggested that the gap in evaluation is driven by gender stereotypes. The results of a study by Ahmed et al. (2021) also indicate the existence of statistical discrimination by gender. They examined the degree of gender discrimination across occupations analyzing the responses of fictitious job applications in Sweden. They concluded that male applicants were less successful than female ones in receiving a positive employer response in female-dominated occupations. For male-dominated and mixed occupations no significant differences were found. Adamus and Ballová Mikušková (2024) highlight that existing stereotypes can influence the evaluation of both women and men. Using experimental design, they investigate gender biases in hiring and firing decisions in Slovakia. The results show clear evidence of wage discrimination against women, who were offered significantly lower salaries than men despite being rated equally in terms of competence and suitability. At the same time, a backlash effect was observed

against men. Male candidates with average or below-average performance were perceived as less likable than their female counterparts, and men with frequent absences were more likely to be selected for dismissal than equally absent women. These findings highlight the complex ways in which gender stereotypes operate in the labor market and suggest that women face direct pay discrimination, while men may be affected by a backlash effect, especially if they do not meet traditional expectations associated with the male role in the workplace.

Many studies have examined the evaluation of men's and women's performance in educational settings, which offer a valuable context for investigating potential biases in the assessment of performance across genders (e.g., Boring, 2017; Mengel et al., 2019). Within these environments, both students are assessed for their academic performance, and teachers are evaluated by their students, allowing for the analysis of gendered aspects in both roles.

Studies conducted at universities have identified that male students more often receive better grades for similar performance, even in the same subjects, and that female performance is undervalued. Mengel et al. (2019) examined the teaching evaluation by students in Netherlands. They found that the female instructors received systematically lower teaching evaluations than male colleagues and the gap is larger for mathematical courses and junior women. There was no evidence that these differences are driven by gender differences in teaching skills. The study also showed that not only male but also female students evaluate instructors less when they are female. The female students were stricter in the rating of both genders and gave on average worse ratings than male students. The role of gender bias is also confirmed in Boring (2017), which examines student assessment of teaching at a French university. Male professors were rated higher than female professors, and the different dimensions of teaching that students value in male and female professors tend to match gender stereotypes.

Also study by Özgümüs at al. (2020) suggests that gender bias can influence teacher evaluation even when all other variables, such as personal interaction or presentation style, are removed. Authors examined whether students evaluate identical instructional materials differently depending on the gender of the instructor. In experiment, students were presented with the same instructional materials presented under either a male or female name. The results showed that male respondents rated materials presented under a female name lower than those presented under a male name, even though the content was identical. The results also showed that experience with a university environment can play a role. When the research was conducted on a sample of respondents who did not have a university education, the differences in evaluations disappeared.

The existence of gender bias and the tendency to rate women less highly in areas that are perceived as masculine is highlighted in a study by Jansson and Tyrefors (2022). This study examines the effect of anonymous grading on female students' performance in introductory macroeconomics courses at Stockholm University. The authors found that female students achieved significantly better grades when graded anonymously compared to when their identities were known to the graders. Girard and Pinar (2009) focus on the presentation skills of the students. This study investigated whether gender similarity between presenters and evaluators affects presentation ratings. It found no significant impact of gender on objective evaluations. However, students rated female presenters as better overall, which suggests that women are perceived as better presenters. The findings also show that women tended to rate female presenters more highly.

This paper builds on the aforementioned studies, seeks to expand existing knowledge, and examines gender differences in evaluation using data from the Czech Republic, where this issue is currently not fully covered.

2. Data and Methods

To test the assessment of men and women and gender bias in evaluation, the data from a Czech economically oriented university was used. It can be expected that thanks to their study focus, these students could represent the approach of future generations of managers who will evaluate the performance of their subordinates, and decide on the filling of job positions and the level of their wages. The research is based on the evaluation of the quality of the students' seminar work. Within the course, Corporate Economics, students of the course prepare a seminar paper and present it at the seminars. Two seminar groups with a balanced gender composition were selected. About half of the students of each seminar were men and the other half were women. Within the course, students evaluated the quality of the seminar work presentation of their colleagues, both in grade and verbally. The evaluation provided valuable feedback to students from their colleagues and made it possible to explore gender differences in the assessment.

The research was conducted during the winter semester of the 2021/2022 academic year. Students participated in an anonymous peer evaluation process. Each student was provided with evaluation cards (yellow for female students and white for male students) and asked to assess their peers both qualitatively (through written comments) and quantitatively (using letter grades ranging from A to F). These evaluation forms were subsequently collected, and the data were recorded and attributed to the respective evaluated individuals.

The resulting dataset comprises 479 observations. A total of 37 students (19 women and 18 men) were evaluated by their peers. In total, 218 evaluations were submitted by male students and 261 by female students.

The paper deals with the gender aspects of the performance evaluation phenomenon. The following research question is addressed in this work.

- **RQ**: Is the evaluation of the individual dependent on the gender of the evaluated and the evaluator?
- According to selected economic and socio-psychologic theories stated above in the „Literature review" the three hypotheses are formulated:
- **H1**: Female students receive worse evaluation, compared to male students.
- **H2**: Women are stricter in their evaluation. If a student is evaluated by a woman, they will receive a worse rating.
- **H3**: Individuals will tend to evaluate better the same-sex individuals. Male evaluators will rate better the male students and the female students will be rated better by female evaluators.

To test the hypothesis, the binary logistic regression model is used. The model is captured by Equation (1) and has the following form:

$$\begin{aligned} logit\left(evaluation_{ij}\right) &= \beta_0 + \beta_1.student\ gender_i + \beta_2.evaluator\ gender_j + \\ &+\beta_3.evaluator\ gender_j.student\ gender_i + \beta_4.seminar\ group_i + \\ &+\beta_5.final\ exam\ score_i + \beta_6.teacher\ evaluation_i + \beta_7.attractiveness_i \end{aligned} \quad (1)$$

where i denotes the evaluated student and j the evaluator. $Evaluation_{ij}$ is a binary (or ordinal) variable indicating how evaluator j evaluated student i.

The dependent variable is the evaluation of the seminar work. It is a dummy variable that takes on the value 1 if the student received the mark A and value 0 if the student received a B-F rating. Only a two-point scale is chosen here because the students were moderate in their assessment, and an A rating represented 64 percent and reflected an „excellent" result without comments. The remaining B-F ratings represented 36 percent of the evaluations and indicated some reservations about the work.

The explanatory variables are the gender of the student, the gender of the evaluator and their interactions, the final exam score, the seminar group, teacher evaluation, and attractiveness. The gender of students is a dummy variable which is 1 for female students and 0 for male students. The gender of the evaluator represents the sex, it equals 1 if the evaluator is a woman and 0 if the evaluator is a man. Seminar group is a categorical variable that indicates the seminar group that the student attends. This control variable is included to filter out unobserved differences between individual seminar groups, where students may get better grades

due to, for example, a better atmosphere there. The final exam score is a proxy for the cognitive skills of the evaluated student. It is a quantitative variable that expresses the percentage point gain from the final exam (it can take on a value from 0 to 100). Teacher evaluation is another explanatory variable that is intended to objectively assess student performance in terms of the quality of the seminar paper. The student received an A – E grade from the teacher for the seminar paper (no student submitted a seminar paper of such poor quality that it received an F grade). All students were assessed by one teacher (female).

The attractiveness of the assessed person can also be one of the factors that can influence the rating that the individual receives. Many professional studies confirm, that more attractive people are more likely to be employed and are paid better compared to less attractive individuals (Gehrsitz, 2014; Fletcher, 2009; Aný̌žová and Matějů, 2018). On the other hand, others show that attractiveness may not play a significant role in some professions (for example Balcar, 2021). The attractiveness of individual students was assessed by four independent persons (two men and two women) based on submitted photographs of the students' faces (obtained from the information system of the university). Students were evaluated using a 5-point Likert scale (1 – lowest, 2 – low, 3 – medium, 4 – high, 5 – highest). In the regression model, it appears as a categorical variable that includes three categories according to the average rating obtained: 1 – 1.5 low, 1.51 – 4.49 middle, and 4.5 – 5 high to distinguish between significantly very attractive and less attractive individuals. Because the attractiveness of individuals can be perceived differently depending on the gender of the evaluator (the same person may appear differently attractive to men and women), the attractiveness of each individual was determined from the perspective of women and the perspective of men. For women, as the average of the women's ratings, for men, as the average of the men's ratings. In the model, attractiveness was assigned to each individual depending on the gender of the evaluator. Accordingly, when the evaluator was a woman, the assessment of attractiveness was based on female perceptions; conversely, when the evaluator was a man, attractiveness was considered from the male perspective.

For testing the stated hypotheses H1 – H3, it is important to estimate the coefficients β_1, β_2 and β_3 of the regression model. The coefficient β_1 captures the effect of the student's gender on the probability of receiving a better evaluation, conditional on the evaluator being male. A positive and significant β_1 would indicate that, among male evaluators, female students are more likely to receive favorable evaluations compared to their male peers. The factor β_2 reflects the effect of the evaluator's gender, conditional on the student being male. It shows whether male students are evaluated differently depending on whether the evaluator is male

or female. Finally, β_3 represents the interaction between the student's and the evaluator's gender. This term tests whether the effect of the student's gender on evaluations differs depending on the evaluator's gender. A statistically significant β_3 would suggest that female evaluators rate female students differently than male evaluators do. The regression coefficients listed above and their combinations are used to confirm or refute the established hypotheses and take into account all combinations of the evaluated person and the evaluator by gender. **H1**: „Female students receive worse evaluation, compared to male students" is confirmed if: $\beta_1 < 0$ and $\beta_1 + \beta_3 < 0$. **H2**: „Women are stricter in their evaluation. If a student is evaluated by a woman, they will receive a worse rating" is confirmed if: $\beta_2 < 0$ and $\beta_2 + \beta_3 < 0$. Finally, **H3**: „Individuals will tend to evaluate better the same-sex individuals. Male evaluators will rate better the male students and the female students will be rated better by female evaluators" is confirmed if: $\beta_1 < 0$ and $\beta_3 > 0$. Descriptive statistics for individual variables are shown in Table 1.

T a b l e 1

Descriptive Statistics

Variable	Category	Number	Statistics	Distribution (%)
Evaluation (grade)	A B-F	307 172		64 36
Evaluation for (student)	Men Women	218 261		45.5 54.5
Evaluation from (evaluator)	Men Women	213 266		44.5 55.5
Seminar group	First Second	290 189		60.5 39.5
Final exam score (student)	(quantitative variable)	479	Max (100) min (57.5) mean (74.1)	
Teacher evaluation	E D C B A	94 114 67 106 98		19.6 23.8 14 22.1 20.5
Attractiveness	Low Middle High	31 418 30		6.5 87.3 6.3
N		479		

Source: Own calculation.

3. Results and Discussion

To examine the gender bias in evaluation and answer the research question, the coefficients of the binary logistics model were estimated (Equation 1). The regression model was built step by step by adding explanatory variables in order to test the robustness of the results. The results are presented in Table 2.

The findings indicate that male students are not evaluated more favorably than their female counterparts (Model 3). The coefficient $\beta_1 > 0$, which means that female students (if evaluated by men) score better than male students. The odds ratio is 1.003, which indicates that if the woman is evaluated by man, the chance of a better rating increases by approximately 0.3 percent, i.e. only very slightly. Moreover, the coefficient of the regression function is not statistically significant in this case. On the other hand, female students tend to rate women better. If a female student is rated by a woman, her rating improves (the coefficient $\beta_3 > 0$). In this case, the odds ratio is 2.16, which means that women's chances of getting a better rating increase by 116 percent if the evaluator is female. The sum of coefficients β_1 and β_3 is higher than one, which means that women receive better ratings even from female evaluators (compared to the situation when women rate men). These findings lead to the rejection of hypothesis H1, which stated that female students receive worse rating, compared to male students.

These findings are surprising to some extent since many professional studies conclude that women are rated worse than men (Mengel et al., 2019; Balcar and Hedija, 2019; Ahmed et al., 2021; Hospido and Sanz, 2021; Jansson and Tyrefors, 2022). The worse evaluation of women is explained mainly from the position of discrimination theories, where the source of inequalities are existing stereotypes and perceived and assigned gender roles in society. Women are assigned the role of caretaker of the family, they are perceived as physically weaker, less psychologically resistant, and less efficient. For this reason, they are also a priori rated worse due to the expected lower performance.

In our case, it was not clearly proven that women were rated worse than men. Several factors may account for this outcome. One possible explanation is that the evaluation focused not only on the content of the seminar papers but, more importantly, on the students' presentation skills. In this domain, women may be perceived as equally competent (or even superior) to men, regardless of traditional gender roles. This interpretation is supported by the findings of Girard and Pinar (2009), who examined students' perceptions of presentation skills and found that female presenters were rated more favorably overall. It is also possible that evaluated output (quality of seminar paper presentation) is well compatible with the idea of work suitable for women and that it is well compatible with the female role. In this case, women's performance may be evaluated better than men's performance. Diederich et al. (2023) conclude that women are evaluated better when they hold less masculine and therefore more role-corresponding functions.

The estimated coefficient β_2 from the logistic regression model (Model 3) and the corresponding odds ratio indicate that when the evaluator is a woman, the likelihood of assigning a higher rating decreases ($\beta_2 < 0$ and odds ratio is 0.37). If the

evaluator is a woman, the odds of giving a better rating is about 63 percent lower compared to the situation where the evaluator is a man. Thus, women are generally stricter in the evaluations than men. The rating given by a woman tends to be worse (than that given by a man) even when a female student is rated by woman. In this case, the probability of a better rating increases, but the coefficient β_3 is statistically significant only at the 10 percent significance level. The sum of the regression coefficients β_2 and β_3 is less than zero, indicating that when a woman is rating, the rating received is worse than that received by a man. The fact that women show a stricter assessment is also confirmed in Model 2, where the coefficient $\beta_2 < 0$ at the 5 percent significance level. These results lead to the confirmation of hypothesis H2, which states that women are stricter in their evaluation. If a student is evaluated by a woman, they will receive a worse rating.

These conclusions align with the findings of studies by Mengel et al. (2019) and Balcar and Hedija (2019). Both studies confirm that women tend to be stricter in their evaluations than men, both when providing feedback and when awarding rewards. Gender stereotypes, particularly beliefs about women's lower performance, may contribute to this increased strictness. One possible explanation is the pressure women face to demonstrate that they are just as capable as men and to succeed on par with them. As a result, women may evaluate others through the lens of these heightened expectations, leading to a stricter assessment.

The interaction of evaluation of a woman by a woman (female student x female evaluator) shows that when a female student is assessed by a woman, the rating is more favorable compared to the situation when a woman evaluates a man. The coefficient β_3 is higher than one (Model 3). If a woman is rated by a woman, the chance of a better rating increases by 116 percent. However, the coefficient of the logistic model is statistically significant only at the 10 percent significance level. When it comes to the evaluation when the evaluator is a man, the coefficient $\beta_1 > 0$. The odds ratio for female students is higher than one, indicating that women have a greater chance of a better evaluation, even though the increase in chance is only small (by 0.2 percent) (the reference category was a male evaluator). However, the regression coefficient is not statistically significant, so we cannot clearly say that men tend to evaluate better women.

Hence, it cannot be unequivocally concluded that men and women evaluate individuals of the same sex better. We reject hypothesis H3, which states that individuals will tend to evaluate better same-sex individuals.

These conclusions do not follow the social identity and similarity – attraction theory, based on the idea that identification with the same gender could lead to a better evaluation of the performance of the same individuals. On the contrary, the students' ratings turned out to be independent of the evaluated gender. This indicates

that individuals could identify rather with a broad group of students as such, regardless of gender. These findings are also somewhat consistent with the earlier finding that the student sample did not show a tendency to discriminate against women.

T a b l e 2

Binary Logistic Regression – Evaluation of Seminar Work

Variable	Model (1) Coefficient	Model (2) Coefficient	Model (3) Coefficient	Model (3) Odds ratio
Female student (β_1)	0.4585* (0.2561)	0.4342* (0.2578)	0.0030 (0.3645)	1.0030 (0.3656)
Female evaluator (β_2)	–	–0.6231** (0.2451)	–0.9988*** (0.3359)	0.3683*** (0.1237)
Female student x female evaluator (β_3)	–	–	0.7704* (0.4645)	2.1606* (1.0036)
Second seminar group	0.9587*** (0.2982)	0.8057*** (0.3061)	0.8087*** (0.3083)	2.2449*** (0.6921)
Final exam score	–0.0019 (0.0126)	–0.0029 (0.0128)	–0.0027 (0.0128)	0.9973 (0.0127)
Teacher evaluation				
D	1.5293*** (0.3407)	1.5744*** (0.3453)	1.6295*** (0.3497)	5.1015*** (1.7839)
C	2.2451*** (0.4014)	2.3431*** (0.4071)	2.3780*** (0.4115)	10.7835*** (4.4377)
B	3.1314*** (0.4615)	3.2100*** (0.4685)	3.2200*** (0.4706)	25.0292*** (11.7793)
A	4.0390*** (0.5240)	4.0981*** (0.5311)	4.1360*** (0.5336)	62.5483*** (33.3748)
Attractiveness				
Middle	0.1061 (0.4918)	0.1161 (0.5011)	0.1506 (0.5122)	1.1626 (0.5955)
High	0.8754 (0.7711)	0.8038 (0.7708)	0.9020 (0.7862)	2.4646 (1.9376)
Constant	–2.0275* (1.1331)	–1.5863 (1.1600)	–1.4640 (1.1633)	0.2313 (0.2691)
$\beta_1 + \beta_3$	–	–	0.7734** (0.3292)	–
$\beta_2 + \beta_3$	–	–	–0.2284 (0.3377)	–
N	479			
Pseudo R^2	0.2475			

Note: Standard errors in parentheses, *** p < 0.01, ** p < 0.05, * p < 0.1, Chi quadrat is statistically significant, Hosmer-Lemeshow goodness of fit test is not statistically significant (p > 0.05), the model describes reality well.

Source: Own calculation.

Conclusion

The paper aimed to investigate gender bias in evaluation and to assess whether the rating differs by gender. Evaluations of men and women and differences in evaluation by gender were investigated on a sample of students of an economically oriented university in the Czech Republic. Students rated the performance of their colleagues in the presentation of an assigned seminar paper.

The following research question was set: „Is the evaluation of the individual dependent on the gender of the evaluated and the evaluator?“. And subsequently, three hypotheses were tested: H1: Female students receive a worse evaluation, compared to male students, and H2: Women are stricter in their evaluation. If a student is evaluated by a woman, they will receive a worse rating, and H3: Individuals will tend to evaluate better the same-sex individuals. Male evaluators will rate better the male students and the female students will be rated better by female evaluators.

With the use of a regression model, it was proven that the evaluation was dependent on the gender of the evaluator. The results of the research showed that female students were stricter than male students when evaluating their peers. On the other hand, there was no proven tendency for men to be rated better than women, nor was there any tendency for individuals to rate individuals of the same sex better. The results thus led to the rejection of hypothesis 1 and hypothesis 3. On the contrary, hypothesis 2 was confirmed.

However, the findings have limited informative value and it is not possible to draw general conclusions from them that are valid for the entire upcoming generation that will enter the labor market and will most likely hold leadership positions in the future and therefore evaluate the performance of their subordinates. The sample studied included only students from one university and the sample was relatively small. It may not have fully captured the weaker effects. The criteria used for objectively evaluating the performance of individuals in the sample may not have completely accurately reflected their actual performance. The study assessed the quality of seminar work and presentation skills. Also, the activity that was evaluated may have been generally perceived as gender-neutral or corresponding to the female role, which is why there was no tendency for men to be evaluated better.

Despite these limitations, this study brings new insights into gender bias in evaluation processes within the Czech Republic, presenting the perspectives of a selected group of young individuals as representatives of the emerging generation of managers. It presented the attitude of a selected young people group as representatives of the new generation who will form the behavior not only of companies but all of society in the future and could affect stereotypes and gender roles in Czech society.

Knowledge of behavior in the field of evaluation of men and women is interesting not only as a means of testing existing theories or forming new ones. It also brings usable knowledge for practice. The results provide valuable insights for human resource management and gender equality policy, as they reflect the evaluative attitudes of the emerging generation entering the labor market and indicate

the extent to which their assessments may still be influenced by gender bias. From the firms and their shareholder's and stakeholders' point of view the information about the expected behavior of new employees and their attitude to evaluation. From the point of view of economic policymakers, it provides information on the behavior of men and women and the patterns of behavior used in the area of equal treatment for men and women.

References

ADAMUS, M. – BALLOVA MIKUSKOVA, E. (2024): Gender Discrimination and the Backlash Effect in Recruitment and Dismissal Processes. Gender in Management, *39*, No. 1, pp. 107 – 129. DOI: 10.1108/GM-01-2022-0002.

AHMED A. – GRANBERG M. – KHANNA S. (2021): Gender Discrimination in Hiring: An Experimental Reexamination of the Swedish Case. PLoS ONE, *16*, No. 1, e0245513. DOI: 10.1371/journal.pone.0245513.

ANÝŽOVÁ, P. – MATĚJŮ, P. (2018): Beauty Still Matters: The Role of Attractiveness in Labour Market Outcomes. International Sociology, *33*, No. 3, pp. 269 – 291. DOI: 10.1177/0268580918760431.

ARROW, K. J. (1973): The Theory of Discrimination. In: ASHENFELTER, O. and REES, A. (eds): Discrimination in Labor Markets. Princeton, NJ: Princeton University Press, pp. 3 – 33. ISBN 9780691645711.

BALCAR, J. (2021): Non-Cognitive Skills Matter, Beauty Not That Much: Evidence on Hiring Technicians. Journal for East European Management Studies, *26*, No. 1, pp. 44 – 72. DOI: 10.5771/0949-6181-2021-1-44.

BALCAR, J. – HEDIJA, V. (2019): Influence of Female Managers on the Gender Wage Gap and Returns to Cognitive and Non-Cognitive Skills. Gender, Work and Organization, *26*, pp. 722 – 737. DOI: 10.1111/gwao.12261.

BECKER, G. S. (1957): The Economics of Discrimination. Chicago: University of Chicago Press.

BORING, A. (2017): Gender Biases in Student Evaluations of Teaching. Journal of Public Economics, *145*, pp. 27 – 41. DOI: 10.1016/j.jpubeco.2016.11.006.

BURKLEY, M. – BLANTON, H. (2005): When Am I My Group? Self-Enhancement versus Self-Justification Accounts of Perceived Prototypicality. Social Justice Research, *18*, No. 4, pp. 445 – 463. DOI: 10.1007/s11211-005-8570-x.

BYRNE, D. (1971): The Attraction Paradigm. New York: Academic Press. ISBN 9780121486501.

CONNELL, R. W. (1995): Masculinities. Berkeley: University of California Press. ISBN 978-0-520-24698-0.

DE BEAUVOIR, S. (2001): Druhé pohlaví (J. Černý, Překl.). Praha: Karolinum. (Původní dílo publikováno 1949).

DIEDERICH, S. – ISEKE, A. – PULL, K. – SCHNEIDER, M. (2023): Role (in-)Congruity and the Catch 22 for Female Executives: How Stereotyping Contributes to the Gender Pay Gap at the Top Executive Level. The International Journal of Human Resource Management, *35*, No. 7, pp. 1283 – 1311. DOI: 10.1080/09585192.2023.2273331.

FLETCHER, J. M. (2009): Beauty vs. Brains: Early Labor Market Outcomes of High School Graduates. Economics Letters, *105*, No. 3, pp. 321 – 325. DOI: 10.1016/j.econlet.2009.09.006.

GEHRSITZ, M. (2014): Looks and Labor: Do Attractive People Work More? Labour, *28*, No. 3, pp. 269 – 287. DOI: 10.1111/labr.12035.

GIRARD, T. – PINAR, M. (2009): An Exploratory Study of Gender Effect on Student Presentation Evaluations: Does Gender Similarity Make a Difference? International Journal of Educational Management, *23*, No. 3, pp. 237 – 251. DOI: 10.1108/09513540910941748.

GOLDIN, C. (2006): The Quiet Revolution That Transformed Women's Employment, Education, and Family. American Economic Review, *96*, No. 2, pp. 1 – 21. DOI: 10.1257/000282806777212350.

HEDIJA, V. (2017): Sector-Specific Gender Pay Gap: Evidence from the European Union Countries. Economic Research-Ekonomska Istraživanja, *30*, No. 1, pp. 1804 – 1819. DOI: 10.1080/1331677X.2017.1392886.

HOSPIDO, L. – SANZ, C. (2021): Gender Gaps in the Evaluation of Research: Evidence from Submissions to Economics Conferences. Oxford Bulletin of Economics and Statistics, *83*, No. 3, pp. 590 – 618. DOI: 10.1111/obes.12409.

JANSSON, J. – TYREFORS, B. (2022): Grading Bias and The Leaky Pipeline in Economics: Evidence from Stockholm University. Labour Economics, *78*, 102212. DOI: 10.1016/j.labeco.2022.102212.

JOHNSON FILIPOVÁ, L. – BALCAR, J. – PLASOVÁ, B. – BRIXIOVA SCHWIDROWSKI, Z. (2024): Explaining the Male-Female Wage Gap: Do Gender Role Attitudes Matter? Sociológia – Slovak Sociological Review, *56,* No. 2, pp. 120 – 153. DOI: 10.31577/sociologia.2024.56.2.5.

MENGEL, F. – SAUERMANN, J. – ZÖLITZ, U. (2019): Gender Bias in Teaching Evaluations. Journal of the European Economic Association, *17*, No. 2, pp. 535 – 566. DOI: 10.1093/jeea/jvx057.

OAKLEY, A. (1972): Sex, Gender and Society. London: Temple Smith.

OCHSENFELD, F. (2014): Why Do Women's Fields of Study Pay Less? A Test of Devaluation, Human Capital, and Gender Role Theory. European Sociological Review, *30,* No. 4, pp. 536 – 548. DOI: 10.1093/esr/jcu060.

ÖZGÜMÜS, A. – RAU, H. A. – TRAUTMANN, S. T. – KÖNIG-KERSTING, C. (2020): Gender Bias in the Evaluation of Teaching Materials. Frontiers Psychology, *11*, pp. 1074. DOI: 10.3389/fpsyg.2020.01074.

TAJFEL, H. (1982): Social Psychology of Intergroup Relations. Annual Review of Psychology, *33*, No. 1, pp. 1 – 39. DOI: 10.1146/annurev.ps.33.020182.000245.

TAJFEL, H. – TURNER, J. C. (1979): An Integrative Theory of Intergroup Conflict. In: AUSTIN, W. G. and WORCHEL, S. (eds): The Social Psychology of Intergroup Relations. Monterey: Brooks-Cole, pp. 33 – 47. DOI: DOI: 10.13140/RG.2.2.30820.60809.