

Students' Gender Bias in Teaching Evaluations

Narissra Punyanunt-Carter*
Department of Communication Studies
Texas Tech University
n.punyanunt@ttu.edu

Stacy L Carter
College of Education
Texas Tech University

Abstract

The goal of this study was to investigate if there is gender bias in student evaluations. Results revealed that there is certainly some gender bias at work when students evaluate their instructors. It was also found that gender bias does not significantly affect the evaluations. Most researchers studying the subject have found some sort of pattern regarding gender bias in evaluations, but still agree that it is inconsequential.

Keywords: teaching evaluations; teacher evaluation; faculty assessment; gender bias; student gender; teacher gender

Introduction

At the beginning of every semester, students at a typical university sit down in their first class and wait for their professor to step into the room. From the moment they begin the first lecture, professors are being watched and judged by their students. At the end of every semester, the professor faces the mass judgment of an entire class in the form of an evaluation.

This standard evaluation is handed out in every classroom on the college campus. It is meant to give the students a voice by allowing them to vocalize what they like or do not like about a specific professor. Almost every journal article we found stressed the importance of instructor evaluations regarding promotion, tenure, and salary. Because these evaluations are so important, what happens when they are filled out by students with a bias against an instructor? By bias, we mean some preconceived factor that would affect the evaluation of a teacher in an untrue way. An example of an obvious bias would be class size, where the student receives more attention in a smaller class, therefore affecting the instructor's evaluation.

The point and overall goal of this particular project was to see if there are any gender biases when it comes to student evaluations of professors. As such, the following paper works to answer the question: What is the impact of gender bias in student evaluations of teachers? We will discuss previous research and thoughts on the subject, our own research and results, as well as our own thoughts on the problem of gender bias in these evaluations. We hope to make the reader of this paper understand and recognize the problem, and prevent it from happening in our classrooms at other college campuses.

*Corresponding author

Literature Review

This particular study was based on the work published by Centra and Gaubatz (2000), “Is there gender bias in student evaluations of teaching?” The authors of that article took evaluations of instructors and analyzed them according to the gender of the student to find out if there was any sort of gender biases at work. Other studies have been conducted on the same subject, regarding gender biases in evaluations (see Andersen & Miller, 1997; Baker & Copp, 1997; Basow, 1995; Huston, 2006; Miller & Chamberlin, 2000). However, current research regarding this topic is not as abundant as the publications that can be found before the year 2000. Certainly, there is current research regarding different types of biases in academia, such as hiring practices (see Corrice, 2009), gender disparities in STEM (see Hill, Corbett, & St. Rose, 2010; National Academy of Sciences, National Academy of Engineering, & Institute of Medicine of the National Academies, 2006), and even citation gaps (see Maliniak, Powers, & Walter, 2013), among many other areas.

Regarding gender bias in student evaluations of teachers, most of the available studies were published before the year 2000. A study conducted in 1973 looked to find some sort of correlation between the ratings students give instructors and things such as the student or teacher’s demographics (Granzin & Painter, 1973). The interesting thing about this study is the fact that after finding such a small correlation between student gender, teacher gender, and the ratings given out, the researchers actually threw out sex as a variable altogether (1973).

A study conducted in 1975 found male students evaluated their female teachers less favorably than their male teachers (Ferber & Huber, 1975). Female students were found to rate all instructors higher than males did, also rating female instructors higher than they did male instructors. Therefore, the researcher found the students to show a bias toward their own sex. Additionally, the study found that positive past experiences with women instructors greatly reduced the preference students had for male instructors (1975).

Yet, another study had found that sex bias in student ratings do not generally occur (see Wilson & Doyle, 1976). However, in this study there were some students who participated, but did not reveal their gender. In this new category, the mean ratings for female instructors were slightly lower than the mean ratings for male instructors. Still, the researchers found this to be statistically insignificant due to the low number of people in the category (1976).

Another researcher used a modified questionnaire in a study to test the idea that teaching evaluations “are influenced by zones of acceptance based on sex stereotypes” (Martin, 1984, p. 488), meaning students are influenced on evaluations by stereotypical authoritative gender roles (i.e., women should act as women and men should act as men). The researcher believed that low instructor evaluations were partly as a result of a sort of revenge on teachers who upset students by rejecting the zones of acceptance (1984). This hypothesis only held up in the evaluation of female social science instructors. Male students rated these teachers higher only when they combined feminine traits with masculine traits in their teaching style (1984).

In a 1985 study, it was found that generally for male students, gender plays a very small role in the evaluations of teachers (Tieman & Rankin-Ullock, 1985). Departments at a southern university were split into two separate fields: traditional and nontraditional. The traditional field included male and female faculty members who were teaching in stereotypical fields such as males teaching math, or females teaching English. The nontraditional fields included male and female instructors who were teaching in fields not generally taught by their particular gender, such as women teaching biology or men teaching nursing. Males were actually discovered to

show a slight favoritism toward female faculty in both traditional and nontraditional fields; the same study found female students to show favoritism toward the underdog in the field (1985). Female students actually rated men teaching in stereotypical female dominated fields higher than the women instructors. The same was the case for male stereotypical fields: women were given higher ratings on evaluations than the male instructors (1985).

Arubayi (1987) also used previous research to form his own conclusions. He found that there were previous studies conducted that concluded there was no relationship between the sex of the evaluator and faculty rating. The personality of the evaluator was found to play a bigger role in the way a student evaluates (Arubayi, 1987, p. 270).

In a 1989 study, Dukes and Victoria hypothesized that the professor's gender and the student's gender would interact statistically on the results of an instructor evaluation. They also predicted that students evaluate instructors of the opposite sex more highly, an effect that is heightened by effective teaching. The findings of the study reported that female instructors were not evaluated lower than male instructors for the same performance (1989). Enthusiasm was found to be the thing that causes higher cross-sex evaluation ratings from students (1989).

Two articles used data mining to infer the researcher's conclusions. A 1997 article found that students evaluate male and female instructors differently because they have different expectations for the way male and female instructors should behave, specifically, in areas such as likeability and competence (Anderson & Miller, 1997). Male and female professors were overall rated as equals, but when the instructor adhered "to the gender appropriate model" (Anderson & Miller, 1997, p. 218), they were found to receive higher ratings on the evaluations. The second article that utilized data mining discusses evaluations in general, not just those completed in universities. The researcher found that women are evaluated less favorably than men when they are highly qualified. Interestingly, women were evaluated more favorably than males when both the male and female were not well qualified. "This implies a different reward system for males and females—one that rewards success and competence in males, and failure and incompetence for females," (Nieva & Gutek, 1980)

Other research focused not on the actual paper evaluations of the instructor, but rather on the student's perception of professional status and education credentials of the instructor based around gender. It was shown that students are more likely to perceive a male instructor as higher in status and credentials than females (Miller & Chamberlin, 2000). Another study involving gender and student-teacher perception found that females with male instructors reported a significantly less favorable overall impression of their instructors (Cromie, Pyke, Silverthorn, Jones, & Piccinin, 2003). This was in comparison to female students with female professors and males with either female or male professors (2003).

A study published in 2000 compared evaluations for both male and female instructors by male and female students (Centra & Gaubatz). The researchers used the *Student Instructional Report II*, an evaluation form that is used in universities, and had courses fill them out three times over three semesters. Data was collected from 741 courses, all of which had at least 10 females and 10 males in each. Multivariate analysis of variance (MANOVA) was used to analyze the data (Centra & Gaubatz, 2000).

Results from the study showed female instructors received higher evaluation ratings from female students six out of eight times, while male instructors received equal ratings from males and females (Centra & Gaubatz, 2000). Female instructors were also viewed as better

organized, better communicators, more interactive, and better at giving quality exams and feedback (2000). Female instructors were found to teach in a different style than males, encouraging discussions and lecturing less (2000).

Centra and Gaubatz viewed this difference in teaching style as potentially causing the higher ratings among female instructors by female students. Likewise, males were found to view male instructors as better organized and more systematic (2000). Cross-gender biases were found to be very small. The authors concluded that, although there might be some gender biases, particularly with females evaluating female instructors, the effects were extremely minimal and most likely caused by differences in teaching style (Centra & Gaubatz, 2000).

The research for this article was based around Centra and Gaubatz since it involves an evaluation that every student has filled out multiple times in a way that is different from the normal usage. The authors of this article also recognized the flexibility this provided in order to tweak the survey in order to target and answer the proposed research questions as well. Also, it must be acknowledged most of the articles focused on this topic are from the 1970s and 1980s, while Centra and Gaubatz is relatively more recent, providing for better results validation.

Research Questions

The researchers of this paper were most interested in seeing just how much a professor's gender influences the student's perceptions of the teacher based on the student's own gender. As such, the authors came up with three basic questions to be answered by the research and analyses conducted. The questions are listed below:

RQ1: Is there a gender bias when students evaluate teachers?

Based on the literature review, the researchers hypothesize that females are more likely to judge their female professors kindly; the same going for male students and male professors. The researchers think there is a sort of gender solidarity that makes it easier for a female to give a female and a male to give a male a good evaluation overall.

RQ2: Do students view professors of their same gender as more effective, clear, knowledgeable, likeable, etc.?

The researchers also hypothesize students will find professors of their same gender as overall better teachers than professors of their opposite gender. They might also find the professor's teaching style to be more effective and find the teacher to be overall respectful and encouraging. It is the opinion of the researchers that students find teachers of their same gender to be easier to learn from and easier to like.

RQ3: How likely is a student to choose a teacher of their same gender from the very beginning, during class selection?

If there is gender bias in these teacher evaluations, where does the bias actually begin? To answer this, the researchers question the possibility that the bias might actually begin in the registration process, when students are choosing the classes they will take. The researchers of this study think that if it is possible for there to be any sort of gender bias, it must begin when they choose between taking a female professor or a male professor's class. If this inference is correct, then the bias problem might be much bigger than just evaluations.

Methods

The data for this project was gathered using a very straight-forward survey method. The researchers used a slightly modified version of the teacher evaluation forms that students fill out at the end of each semester. The survey was filled out by students in a basic introductory communications class. Half the class was instructed to fill out the survey about a male professor, and the other half a female professor. All together, 30 males and 28 females completed the evaluation. Of those, 17 males and 15 females evaluated a female professor, and 13 males and 13 females evaluated a male professor. All students who took the survey remained anonymous to the research group.

The survey had questions regarding the professor's overall effectiveness, clarity, and level of knowledge. The survey also asked about the ways in which the instructor conducts the class and treats the students. All of these questions were taken straight from the Texas Tech teacher evaluation form.

A few modifications were made to the survey to better fit the researchers' needs and answer the research questions. The biggest change made to the survey was adding a place for the student to enter the professor's gender as well as the student's own gender. This was added to allow the researchers to know the gender of the student and the gender of the teacher so that they could compare the two.

The researchers also added questions at the end of the survey regarding the professor's likeability and how much the student felt he or she learns from the teacher. The researchers added these questions because they were curious to know: if an instructor is well liked, but not necessarily someone you learn a lot from, will he or she still receive a good evaluation? In other words, the researchers wanted to find out if evaluations are based off of good teaching, or just how much the instructor is liked.

Lastly, a question was asked regarding whether or not a student is likely to choose an instructor of his or her same sex when choosing courses. This question was added to see if some students solely choose classes based on some sort of gender bias.

Students chose from five different responses: strongly agree, agree, neutral, disagree, and strongly disagree. All results were given numerical value, entered into a spreadsheet in excel, and analyzed by the group.

Analysis and Results

For an easier presentation of the analysis and results of this study, the researchers will simply break the evaluation down, question by question, and review the findings for each one. This gives a detailed accounting of exactly what was found.

The first question asked about the instructor's effectiveness in the classroom. For the female professor, 82% of males and 67% of females agreed or strongly agreed that the instructor was effective. Sixty-nine percent of females and 69% of males agreed or strongly agreed that the instructor was effective.

Question 2 asked if the instructor stimulated student learning. Eighty-two percent of males and 60% of females agreed or strongly agreed that the female professor stimulated

learning. Seventy-seven percent of males and 69% of females agreed or strongly agreed that the male professor stimulated learning.

Questions 3 and 4 regarded the fairness and respectfulness of the instructor. Both of these questions garnered very similar results from the students. Overwhelmingly, females and males alike viewed both the female and the male instructors as being fair and respectful. According to the findings in this study, 82% of males and 87% of females surveyed either agreed or strongly agreed that the female instructor treated students fairly. Eighty-two percent of males and 85% of females either agreed or strongly agreed that the male instructor treated students fairly. Likewise, 88% of males and 93% of females agreed or strongly agreed that the instructor treated students with respect. Eighty-two percent of males and 100% of females agreed or strongly agreed that the male instructor had respect for the students.

Question 5 of the evaluation asked if the instructor welcomed and encouraged questions and comments in the lecture. Eighty-four percent of males and 87% of females either agreed or strongly agreed that the female instructor encouraged student input in the classroom. As for the male instructor, 100% of males and 77% of females found him to welcome student questions and comments.

Question 6 asked about the clarity of the instructor. Seventy-six percent of males and 69% of females agreed or strongly agreed that the female instructor was easy to understand. Eighty-five percent of males found the male instructor to be clear when teaching. However, only 31% of females agreed the male instructor was clear. Additionally, only 2 of the 13 females surveyed strongly agreed that the male instructor was easy to understand.

Question 7 discusses the instructor's knowledge on the subject being taught. 94% of males and 87% of females found the female instructor to be knowledgeable on the subject. Overwhelmingly, 100% of males found the male instructor to be knowledgeable, while 85% of females agreed.

Question 8 asked how much the student likes the professor. Seventy-one percent of males and 73% of females were found to like the female instructor either a lot or an average amount. Seventy-seven percent of males and 77% of females liked the male instructor they evaluated either a lot or an average amount.

Question 9 dealt with how much the student felt he or she learned from the evaluated instructor. Seventy-six percent of males and 53% of females felt they learned either a lot or an average amount from the female instructor evaluated. Fifty-four percent of males and 54% of females indicated that they had learned a lot or an average amount from the male instructor they evaluated. Additionally, 38% of males and 38% of females felt they had learned none or only some from the male instructor, a number that differs a bit from the rest of the survey findings.

The last question asked about the likelihood of a student specifically taking a course based on the teacher's gender. Seventy-seven percent of students, male and female combined, chose neutral, stating that it would not matter the gender of the teacher when choosing classes. Seventeen percent of males said they were not likely or not at all likely to choose a male instructor, and only 7% of females said they were not likely or not at all likely to choose a female instructor. No one surveyed said they were always likely to choose an instructor of their same sex when choosing classes.

Discussion

Implications

The study conducted shows a number of things about teacher evaluation practices. The researchers believe, based on the results shown here, there is very little gender bias in the student evaluations of teachers. However, there are a few patterns that emerged in the research results. Many times over, the male student evaluated the female instructor higher than the female student, or the female student evaluated the male instructor higher than the male student. Male students actually rated the female instructor higher than female students did 6 of 10 times. This result shows a slight, possible cross-sex bias.

When rating the male instructor, males and females tended to agree more often than when rating the female instructor. For the results regarding the male professor, male and female students were within 10 percentage points 7 of 10 times. However, when it came to clarity, although 85% of female students found the male professor to be knowledgeable of the subject matter, only 31% found him to be clear. Additionally, only 54% of females felt they had learned something from the male instructor. This suggests that female students are less likely to rate an instructor highly on an evaluation than a male student when it comes to teaching style and ability.

For the most part, likeability did not play a part in the student's evaluation of the teacher. Generally, even when a student indicated they did not like the professor, they were evaluating, they still gave him/her a good score in other areas. We take this to mean that students do not have any sort of likeability bias towards an instructor. Even if the student does not like the professor, he or she can still give that professor an objective evaluation.

Because no student surveyed said they always choose an instructor of their same sex, this shows that if there is any sort of gender bias in the evaluations it is entirely unintentional. Otherwise, the bias would begin from the very beginning when classes are chosen. If anything, the 17% of males that indicated they are more likely to choose a female instructor shows a bias toward females as stated before.

The original study, "Is there gender bias in the student evaluations of teachers?", had slightly different results than this study. While the original found female students to give female professors better evaluations, the present study found males to typically evaluate female professors higher. The original study also saw cross-gender biases to be few and far between (Centra & Gaubatz, 2000), while the researchers in this study found them to be prevalent. Although there are definite differences, both studies do conclude that the gender biases that do exist in student evaluations of instructors are not significant enough to actually affect the purpose of the evaluation itself.

Limitations

As with any research project, there were a few limitations to this study which, in turn, could have influenced our results. First of all, the research project was conducted during one semester, which necessarily implies time restraints. The survey was definitely not passed out to enough students, especially when the students who did have access to the survey were split into two parts, one group evaluating a male instructor and the other a female instructor.

Secondly, the survey distributed, as mentioned in the Methods section, featured a five-point scale. This scale included choices for strongly agrees, agree, neutral, disagree, and strongly disagree, from left to right. However, on the last three questions, the order of the choices was reversed, with the negative choices appearing on the left and the positive choices on the right. This might have led to some confusion when students got to the last questions, possibly skewing some of their responses.

Another problem exists while the student is taking the survey. There is no actual way to determine if the student was simply rushing through the evaluation or actually taking the time to fill it out correctly. Moreover, this is a problem with any sort of survey research.

Lastly, there were quite a few neutral responses. By answering neutral, no implications can be made about what the student thinks of his or her instructor. However, by simply taking the neutral choice off of the evaluations, you force a student to choose a response that they might not otherwise choose, again skewing the data. The solution to this issue is, obviously, to pass the survey out to more people. This will increase data amounts and therefore cut neutral responses down.

Future Research

Future research on the topic of gender bias in teacher evaluation could definitely include a much broader sample of students filling out the evaluations. It must be noted that, given the limited amount of recent research addressing this topic compared to the number of studies published before the year 2000, as well as the limited size of the study sample used in this case, the research conducted here can serve as a pilot or preliminary study. Further research with larger samples can yield more accurate the results, analysis, and findings. The evaluations could also be done in multiple departments of a single university to better gauge the thoughts of the entire campus, not just one group. Additionally, a multiple university study could be conducted to compare different schools as well as broaden the sample size to an even larger scale.

Other options for future research could include in-depth interviews with students. Most of the research we have come across has all been qualitative. Doing interviews gives the researcher a chance to find out not only if there is gender bias, but also why it is there and how to get rid of it. By hearing the thoughts of actual students, the researcher can give insight that qualitative research sometimes lacks.

Conclusion

While there is certainly some gender bias at work when students evaluate their instructors, the researchers have found that it might not be enough to truly affect the evaluations in a strong way. Additionally, most researchers studying the subject have found some sort of pattern regarding gender bias in evaluations, but still agree that it is inconsequential.

In conclusion, it is important to remember that evaluations are not meant to judge the instructor as a person, gender member, and race member, among other. They are meant to evaluate the instructor's effectiveness in the classroom. As Tieman and Rankin-Ullock (1985) stated, "It is impossible to say whether student evaluations reflect actual performance differences by faculty or only the perceptions of students," (p. 189). This seems to be the recurring theme in the studies on gender bias in evaluations. It is hard to know when a student is judging the teachings of the instructor and when he/she is simply judging the teacher as a

human being. By keeping biases, gender biases included, out of evaluations, the student and instructor will reap the benefits. Students will have better instructors and the instructors will have earned their position fairly and without bias.

References

- Anderson, K., & Miller, E.D. (1997). Gender and student evaluations of teaching. *PS: Political Science and Politics*, 30, 216-219.
- Arubayi, Eric A. (1987). Improvement of instruction and teacher effectiveness: Are student ratings reliable and valid? *Higher Education*, 16, 267-278.
- Baker, P., & Copp, M. (1997). Gender matters most: The interaction of gendered expectations, feminist course content, and pregnancy in student course evaluations. *Teaching Sociology*, 25(1), 29-43.
- Basow, S. A. (1995). Student evaluations of college professors: When gender matters. *Journal of Educational Psychology*, 87(4), 656-665.
- Centra, J.A., & Gaubatz, N.B. (2000). Is there gender bias in student evaluations of teaching? *The Journal of Higher Education*, 70, 17-33.
- Corrice A. (2009). Unconscious bias in faculty and leadership recruitment: A literature review. *AAMC Analysis in Brief*, 9(2), 1-2.
- Crombie, G., Pyke, S.W., Silverthorn, N., Jones, A., & Piccinin, S. (2003). Students' perceptions of their classroom participation and instructor as a function of gender and context. *The Journal of Higher Education*, 74, 51-76.
- Dukes, R.L., & Victoria, G. (1989). The effects of gender, status, and effective teaching on the evaluation of college instruction. *Teaching Sociology*, 17, 447-457)
- Ferber, M.A., & Huber, J.A. (1975). Sex of student and instructor: A study of student bias. *The American Journal of Sociology*, 80, 949-963.
- Granzin, K.L., & Painter, J.J. (1973). A new explanation for students' course evaluation tendencies. *American Educational Research Journal*, 10, 115-124.
- Hill C., Corbett C., St. Rose, A. (2010). *Why so few? Women in science, technology, engineering, and mathematics*. Washington, DC: American Association of University Women.
- Huston, T. A. (2006). Race and gender bias in higher education: Could faculty course evaluations impede further progress toward parity? *Seattle Journal for Social Justice*, 4(2), 591-611.
- Maliniak, D., Powers, R., & Walter, B. F. (2013). The gender citation gap in international relations. *International Organization*, 67(4), 889-922.

- Martin, Elaine. (1984). Power and authority in the classroom: Sexist stereotypes in teaching evaluations. *Signs*, 9, 482-492.
- Miller, J., & Chamberlin, M. (2000). Women are teachers, men are professors: A study of student perceptions. *Teaching Sociology*, 28, 283-298.
- National Academy of Sciences, National Academy of Engineering, Institute of Medicine of the National Academies. (2006). *Beyond biases and barriers: Fulfilling the potential of women in academic science and engineering*. Washington, DC: National Academies Press.
- Nieva, V.F., & Gutek, B.A. (1980). Sex effects on evaluation. *Academy of Management Review*, 5, 267-276.
- Tieman, C. R., & Rankin-Ullock, B. (1985). Student evaluations of teachers: An examination of the effect of sex and the field of study. *Teaching Sociology*, 12, 177-191.
- Wilson, D., & Doyle, K.O. (1976). Student ratings of instruction: Student and instructor sex interactions. *The Journal of Higher Education*, 47, 465-470.