Gender Bias in Instructor Evaluations—What Can We Do About It?

Caitlin McGrew

Rachel Garshick Kleit, PhD

The Ohio State University

December 2019
Contents

Executive Summary .......................................................................................................................... ii
Introduction ..................................................................................................................................... 1
Gender bias’s influence in student evaluations of instruction ..................................................... 5
  Gendered expectations, perception, and exposure ............................................................................ 5
  Student gender and assessment ........................................................................................................... 9
The perception is what matters, not the instructor’s sex ................................................................. 13
Evaluation Scales .............................................................................................................................. 19
Open-ended comments ...................................................................................................................... 21
Mitigating Bias .................................................................................................................................. 21
  Institutional recommendations ............................................................................................................ 21
  Classroom level recommendations ..................................................................................................... 23
  Future Research ............................................................................................................................... 24
Conclusion ....................................................................................................................................... 25
References ....................................................................................................................................... 27
Appendix A. Teaching Students How to Give Feedback Slides......................................................... 32
Appendix B. Mechanical and Aerospace Engineering’s Departmental SEI Materials .... 34
Appendix C. Summary of Peterson et al. (2019) ............................................................................. 45
Executive Summary

After reviewing four decades of investigation, most research on the impact of gender bias shows that student gender, teacher gender, and student evaluations of instructors (SEIs) interact with each other. Predicting the relationship can be difficult, and some studies suggest that over time gendered expectations can influence not only male students to rate female instructors lower on SEIs but also can influence how they engage in the classroom. For example, traditionally women are expected to be nurturing, warm, and friendly. Women have to work harder than their male counterparts to fulfill students’ gendered expectations and the expectations of their position. Furthermore, students tend to rate same sex instructors higher. Students are biased towards one another and toward instructors. Representation is also an important factor in ratings and expectations—the higher proportion of female to male instructors in a discipline can reduce differential gender expectations.

The existing literature about recommendations to mitigate gender bias in SEIs is limited and suggests both systematic institutional remedies along with limited classroom-level remedies. One recommendation was to include language in the evaluation instrument explaining bias with the intent of raising student awareness about the influence of bias in evaluations. Another was to use a 6-point evaluation scale instead of a 10-point scale due to cultural associations with the number 10 and the ideal of brilliance and perfection. Within the classroom, the research suggests that female instructors make sure to mention their credentials, dress professionally, and explicitly explain the benefits of their grading practices. Lastly, it is important for evaluation questions to be objective and behaviorally based (rather than asking students to evaluate instructors’ content expertise, for example).
STEM specific sources were missing from the existing literature. Rather than examining a specific field, many of the studies included students from a variety of disciplines. Most studies conducted on this topic did not control for students’ disciplinary affiliation, and therefore the research is controversial concerning whether gender bias drastically differs between disciplines. Literature about classroom mitigations was also a rarity. The majority of the literature recommends working with the existing bias that is culturally embedded. However specific recommendations for classroom management techniques were largely missing from the literature.
Introduction

Research is inconsistent regarding the role gender bias plays in student rating of instruction. Generally, the student’s sex, instructor’s sex, student’s expectations of instructors, and the representation of the sexes in the discipline can interact to influence student evaluations of instruction (SEIs). When gender bias—the “unfair difference in the way women and men are treated” (gender bias, 2020)—it can have an impact on faculty salaries, respect, grant proposal success, promotion opportunities, and employment (Roper 2019). Additionally, gender bias is widespread in the sciences and SEIs have a special impact because these are often a key piece of the assessment of faculty when seeking promotion. When systematic differences exist in SEIs or in other aspects of faculty’s roles, these dynamics can create unequal outcomes for men and women. Gender bias is a form of implicit bias, meaning that it arises from “thoughts and feelings outside of conscious awareness and control.” (Project Implicit, 2011) Easterly and Ricard (2011) conclude that unconscious bias and gender schemas have an impact on the slower increase of women in STEM higher education disciplines. Addressing gender bias, therefore, is important in the continued growth of women as STEM faculty.

This concern about gender bias arises takes place in the context of a larger conversation that is occurring in the academic community about the impact of sexual harassment of women in STEM. While the representation of women in STEM has been increasing in the United States, continuing sexual harassment¹ of women in STEM is having an adverse impact of women’s

---

¹ “There are three categories of sexually harassing behavior: (1) gender harassment (verbal and nonverbal behaviors that convey hostility, objectification, exclusion, or second-class status about members of one gender), (2) unwanted sexual attention (verbal or physical unwelcome sexual advances, which can include assault), and (3) sexual coercion (when favorable professional or educational treatment is conditioned on sexual activity). Harassing behavior can be either direct (targeted at an individual) or ambient (a general level of sexual harassment in an environment).” (National Academies of Science, Engineering and Medicine, 2018:2)
career’s in the academy (National Academies of Sciences 2019). When female faculty experience sexual harassment in the academic workplace, they experience a decline in job satisfaction, withdrawal from the organization, increase in job stress, and decline in productivity/performance (National Academies of Sciences 2018). When students experience sexual harassment at universities their motivation to attend classes declines, they drop classes, pay less attention in class, and therefore earn lower grades. Systematic factors contribute to the conditions under which sexual harassment occurs in academia. These conditions include a perceived tolerance for sexual harassment (the most potent predictor for the occurrence of sexual harassment in academia), male dominated work settings, hierarchical power structures, symbolic compliance with legislature such as Title IX and Title VII, and uninformed leadership. Furthermore, when the leadership lacks intentions to aggressively reduce and eliminate sexual harassment and lack tools to achieve this objective, the culture that enables sexual harassment perpetuates. While gender bias is not harassment, is part of an environment that can perpetuate it and lead to continued differential treatment of men and women. Thus, this literature review and its accompanying tool box for mitigating gender bias in SEIs is a small step in the support of a culture where sexual harassment is not tolerated.

is not enough to address climate or organizational culture. The report’s recommendations include, in brief:

- Create diverse, inclusive, and respectful environments that include female representation at every level within the institution.

- Evaluate staff members on criteria regarding greater cooperation, respect in the workplace, and professionalism.

- Offer and/or mandate population specific training focusing on changing behavior, as a means to provide skills for administration, faculty, and students to reduce sexual harassment, versus solely ensuring legal compliance, and

- Address the most common form of sexual harassment: gender harassment.

- Institutions need to move beyond legal compliance and address cultural and climatic issues by engaging with and listening to students.

- Improving transparency about processing incident reports

- Diffuse hierarchical relationships between trainees and faculty

- Provide support for targets of harassment

- Measure progress when studying harassment within one’s institution

- Incentivize change,

- Address failures to meaningfully

- Enforce Title VII,

- Conduct necessary research,
• and make the entire community responsible for reducing harassment (p. 5-12).

The report and the convening are the context in which this project began. Additionally, women in the College of Engineering at Ohio State have time to time discussed their experience of differential assessment by students compared to that received by male colleges and its impact on their careers. In one case, a female faculty member had a course with several sections. One section was aviation, while the other was city and regional planning (CRP). For the same course that met in the same classroom at the same time, SEIs had a 1.5-point differential, with much higher scores in the CRP section than for aviation. This sort of experiences motivated this investigation into the mitigation of bias in SEIs.

This review of the literature begins with an overview of the research regarding gender bias in student evaluation of instructors, including how gendered expectations, perception, and exposure can influence SEIs. It then turns the relationship between the sex of the student and the sex of the instructor, which also interact in SEIs. Some studies are inconclusive a concerning the impact of gender bias on SEIs, perhaps due to research design or adaptive behaviors on the part of instructors. Measurement within SEIs also can influence the degree to which bias gets magnified. Augmenting the literature review are a series of interviews with Ohio State faculty and an expert in the field to discuss recommendations or best practices for mitigating bias in SEIs. The paper concludes with a set of recommendations based on the findings addressing the mitigation at both the institution and classroom levels and an analysis of paths for future research.
Gender bias’s influence in student evaluations of instruction

Student sex, instructor sex, student expectations of instructors, and the representation of the sexes in the discipline interact to influence student evaluations of instruction—in most studies. In their review of the literature, D'Entremont and Gustafson (2017) summarize that some studies found no or small differences, while others found female professors to be rated lower than males. Measurement can contribute to the some of the differentials in student evaluations of instruction (SEIs); some are indeed the product of gender bias.

Gendered expectations, perception, and exposure

Gendered expectations of the rater can influence SEIs and women typically receive lower ratings (D'Entremont and Gustafson 2017). For example, students may be more likely to assume a male instructor has a PhD than does a female. Students are more likely to refer to male instructors as “Doctor” or by their last name. Notably, enthusiasm has been a significant factor in increasing SEI scores, and is consistently a male stereotype, therefore students discount female instructors’ enthusiasm. Furthermore, male faculty are rated as more skilled and able instructors compared to women, and these biases transfer into student assessments of course materials, i.e. rating textbooks and readings assigned by men higher than those assigned by women, and reviewing women as taking longer to return feedback even when in actuality feedback is provided at the same time. Overall, men are judged more so for their subject matter expertise, and women are judged for their interpersonal qualities. These findings demonstrate varying expectations for female and male instructors.

Further evidence supports the relationship between expectations and student evaluation. Students expect professors to behave according to certain “masculine” and “feminine” traits; consequently, instructors who adhere to these gendered expectations are viewed more favorably.
Generally speaking, “student perceptions and evaluations of female faculty are tied more closely to their gender expectations than for male faculty.” (Bachen, McLoughlin and Garcia 1999:196). Female faculty who do not meet the high expectations of embodying gendered interpersonal traits and the expectations of the role of a professor of maintaining high degree of professionalism and objectivity are viewed as less effective than men (Basow 1995, Boring 2017, MacNell, Driscoll and Hunt 2015).

Additionally, Miller and Chamberlin (2000) examined student perceptions of female and male instructors’ credentials, and found the perceptions to be “subjectively experienced, yet they reflect accurately the extant social order, characterized by gender stereotypes and stratification within colleges and universities.” (Miller and Chamberlin 2000:283). Students misattributed in male instructors’ attainment in an upward direction, and female attainment levels in a downward direction. The study collected data from sociology courses, and found students to attribute Ph.D. achievement to males, and “are less likely to attribute the educational credentials actually achieve by a woman to a woman.” (Miller and Chamberlin 2000:295). This reveals the “professor” status for many students is reserve solely for male instructors. The authors hypothesize student attributions are due to gender bias and stereotypes, because male instructors are more likely to communicate their credentials and achievements. This could also be because men are more likely to use their scholarly work in their teaching, and female instructors may use teaching strategies that reminds them of the high school experiences, and therefore that misperceive males to be “professors” and females to be “teachers.” (Miller and Chamberlin 2000:295) The disparity observed in student ratings may be a result of failure to meet gendered expectations, rather than differences in teaching.
In a study of student evaluations of over four hundred faculty made by over nine thousand students, Sidanius and Crane (1989) found male faculty were given higher ratings on global effectiveness and academic competence compared to their female counterparts. Furthermore, female faculty were not found to be more sensitive to student needs, and students placed more weight on academic competence for male faculty than female. In their analysis, the authors controlled for “confounding variables such as academic rank, student sex, student GPA, student’s expected grade, academic discipline, the number of students in the course, and the percent of women faculty in a given department.” (Sidanius and Crane 1989:191) The results of men being more rated as more competent and more adequate were found before and after controls were included. However, women were rated more sensitive to student needs before controlling for the above variables.

These gendered expectations have implications for the workload of female faculty. Sprague and Massoni (2005) studied how 288 college students at two campuses describe their best- and worst-ever teachers and found “indications that students hold teachers accountable to certain gendered expectations” (Sprague & Massoni, 2005, p.779). These expectations place “labor-intensive burdens” (Sprague and Massoni 2005:779), and lead to greater hostility towards women who do not meet gendered expectations of students. More students rated a woman as the best teacher they ever had, however, the authors focused on the “perceptual categories through which these students recalled their teachers” (Sprague and Massoni 2005:783) Six of the top eight words used to describe their best-ever teacher were the same for both gender: caring, understanding, intelligent, helpful, interesting, and fair. However, the words most used to describe best male teachers include caring, understanding, and funny, in comparison to women teachers who were most often described as caring, helpful, and kind. There are notable
distinctions behind the meaning of these words for each gender. For example, caring can refer to an attitude, (i.e. someone who “gives a darn”), or someone “who actively does the hands-on work of taking care of a student (i.e. a behavioral pattern).” (Sprague and Massoni 2005:783) In sum, students remember “their best ever teachers as caring, intelligent, and energetic, but they remember their best men teachers as funny, whereas their best women teachers were more likely to be described as caring and nurturing.” (Sprague and Massoni 2005:791) These findings reveal the burden of gendered expectations, and the time-consuming nature of the burden that fall on women. Female faculty may need to invest more “emotional labor” to work with individual students, listen to them, support them, and help them. These findings, although exploratory, are significant because they reveal “student ratings need to be interpreted by comparison with qualitative feedback from students...with a sensitivity to gender stereotyping.” (Sprague and Massoni 2005:792)

When instructors live up to expectations set by students, their evaluation rating is higher, “instructors who fit stereotypes received better evaluations than those who deviated.” (Andersen and Miller 1997:217). These expectations vary depending on gender and discipline. Females are expected to fulfill their gender role of being friendly and nurturing, however the expectations of their gender and professional role may be incompatible, as “anecdotal evidence certainly suggests that many female faculty feel that no matter how they act, their behavior is ‘not quite right.’” (Sandler 1991:7).

Buser, Hayter and Marshall (2019) concur with this dilemma. Their research demonstrates that over the course of a term, ratings of female instructors decline over the course of a semester, relative to male instructors, associated with returning grades. As students’ expectations are realized and feedback is received, their ratings of female instructors decline
Past experiences--specifically prior exposure--reduces the influence of stereotypes and gendered expectations on evaluations. Fandt and Stevens (1991) suggest that prior experience with female faculty influences student evaluations of teachers. At the same time, students gave male professors high ratings due to a professor’s style of organization, enthusiasm, credibility, effectiveness, and students’ willingness to take a course with the professor. The findings of this study are significant because “the findings imply that exposure to women in positions of responsibility may reduce stereotypical attitudes regarding women’s ability to function in gender-atypical roles.” (Fandt and Stevens 1991:469) Similarly, students who have fewer interactions with their instructor will rely more on stereotypes to inform their evaluation. Within engineering classrooms, there is a low percentage of women faculty and heightened gendered exceptions, and, as d’Entremont and Gustafson comment, “this can translate into being perceived as exceptions to gender stereotypes as well, reducing the dissonance of women faculty acting in traditionally masculine ways.” (D’Entremont and Gustafson 2017:2)

**Student gender and assessment**

The gender of the student can matter in the assessment, although some studies found no relationship (Basow and Silberg 1987, Sidanius and Crane 1989). In their review of the literature, D’Entremont and Gustafson (2017) found that female students were more likely to choose a female as their “best” professor, and male students were more likely to choose a male instructor as their “best” instructor. In engineering classrooms, the potential advantage of female student ratings is “overbalanced by a preponderance of lower ratings from male students.” (D’Entremont and Gustafson 2017) Similarly, in evaluating data from a French university, Boring (2017) concluded that male students hold a bias favoring male professors and the teaching dimensions students value coincide with gender stereotypes. Male students also rated
male professors higher. On the dimensions of preparation and organization of classes, quality of instructional materials, and usefulness of feedback, male students rate male professor significantly higher than female professors. Knowledge and delivery style were two dimensions in which women received significantly lower ratings. However, “on availability and quality of contact, female professors obtain their highest predicted excellent score.” (Boring 2017:33).

Sometimes specific instructor qualities influence student’s evaluations. Basow, Phelan and Capatosto (2006)’s examined gendered differences between “best” and “worst” professors. They conducted an analysis of 175 student descriptions of their best and worst professors. Student and professor gender were associated with variations in descriptions. Male students were more likely to identify a male professor as their best professor. Female students chose female professors as their best in direct proportion to their availability. “Best” teachers were described as: enthusiastic, stimulated interest, knowledge, dedicated, caring, available, fair, interactive, open, clear and organized. Students chose “twice as many male as female professors… as [their] worst (66.1% compared to 33.9%).” (Basow, Phelan and Capatosto 2006:30) Female students rated male and female professors the same, while male students were likely to describe their “worst” female professor as rigid and closed-minded. Male students also described their “worst” female professor as having poor communication. Female students were more likely to describe their “worst” professor as having poor teaching skills, as unenthusiastic, and boring.

In sum, this study reveals male students to be sensitive to language and communication style regardless of their instructor’s gender. The authors suspect this is due to it conveying something about their own status, “with a male professor, they may have been particularly sensitive to being slighted; with a female professor, they may have been particularly sensitive to not having their voices heard and not liking the professor’s voice (“poor communication”
included comments about vocal qualities, such as having an accent or talking in a monotone).” (Basow, Phelan and Capatosto 2006:33) Based on this study’s findings, authors state it is imperative for student evaluations to be evaluated critically to eliminate any gender bias. Furthermore, teachers also need to be aware that certain teaching styles and qualities may be differentially effective based on their own, as well as their students’ gender. The current results suggest that if professors want to receive positive student ratings, it may be particularly important for female professors to appear strong in Instructor–Individual Student Interaction, especially to be viewed as accessible, and for male professors to appear knowledgeable, dynamic, clear, and organized. Thus, although best professors combine both active-instrumental and expressive-nurturant traits, the latter may be more important for female professors (especially with male students), and the former for male professors. Overall, it is important that student ratings and recommendations of teachers are used with greater awareness of the subtle role gender variables play.

Additionally, the sex of the student interacts with the sex of the instructor in enhancing differential expectations for female instructors, as Basow (1995) comments in her summary of her own work. Ratings of female college instructors are lower than male students, and male students evaluate them on a heavier set of expectations. This could be due to the norm of male professors in college teaching. Women who violate expectations of students are rated more negatively than those who behave as expected; women are expecting to fulfill their gender role of being nurturing, warm, competent, and, also, knowledgeable. Additionally, male students are more influenced by gender stereotypes than females. Male students in business majors, economics, or engineering rate female faculty most negatively and hold more traditional attitudes
toward women. Ratings of instructors reflect gender differences in their teaching style preferences as well.

Discipline matters as well. Basow (1995) studied four years of student evaluations at a private liberal arts school. She found a significant multivariate interaction between teacher gender and student gender. Male professor ratings did not experience an impact of student gender, but female professors received the highest rating from female students, and the lowest from male students. Basow (1995) also points out that student value teaching styles differently; for example, female students like warmth and availability. Teaching evaluations are especially influenced when in a gender atypical field, as their “gender is particularly salient… her teaching evaluations may be especially affected by her student-perceived personality characteristics and by the gender of her students.” (Basow 1995:656) Professor ranking and representation is also important, as men are overrepresented in higher ranks, in science departments, and women are overrepresented in introductory courses. Throughout the study, there was a remarkable interaction between student gender and teacher gender, and teacher gender and division interaction. Male faculty are rated similarly by male and female students, while female faculty are rated differently by male and female students. Females rate female students higher than do their male peers. Dr. Basow notes it is important for research to examine teacher gender in interaction with both student gender and divisional affiliation, prior to concluding the impact of gender on student evaluations. She also suggests that gender typed personality traits could be more important in evaluations than teacher/student gender.

43 professors were included in the study, 23 were male and 20 females. 27% of faculty included were full professors, 22% were associate professors, 51% assistant, and 49% were tenure-track. Professors rated themselves using the same 26 questions, plus an additional 9 exploratory prompts. In student ratings, female professors were rated higher on 2 interpersonal factors and rated lower on most factors in natural science courses. The study showed a lack of correspondence between student ratings of faculty and faculty self-ratings, significant divisional differences in faculty and study ratings, and student ratings of faculty were affected by professor gender, and divisional affiliation, and the questions on the evaluation.

Templeton (2016) discusses this dynamic in her editorial on findings about the impact of medical school faculty gender on teaching evaluation results. Male faculty earned higher ratings in all rotations, even female dominated departments. There was no remarkable difference in ratings between male and female students. Interactions between faculty and medical students are anticipated to suffer from unconscious bias in ambiguous situations. As in MacNell, Driscoll and Hunt (2015), men are expected to be authoritative and competent and women are expected to be nurturing and sympathetic; if these characteristics are not exhibited than they will be subconsciously considered different. Templeton (2016) concluded that discussion about unconscious bias with medical students is necessary, and student evaluations of faculty should not be the sole source for evaluation of faculty teaching.

*The perception is what matters, not the instructor’s sex*

While most studies suggest the relationship between gender bias and student ratings and that gender plays a complex role in rating of instructors, the research does vary in demonstrating the relationship. For example, Basow and Howe (1987) found contradictory results. Results
show the professors’ sex-type influences their evaluation, but not their sex—rather it is the reported personal qualities of the faculty member that matter.

MacNell, Driscoll and Hunt (2015) suggest this perception is what matters, not the actual sex of the professor. In their experiment, 72 students were randomly assigned to one of six discussion groups, four of which were taught by an assistant instructor. Each assistant instructor taught two courses, one of which was taught under their own identity, the second course was taught under the other assistant’s identity. Therefore, one of the two groups thought they had a female instructor, when they actually had a male instructor. At the end of the course, instructors were evaluated through an online survey, which consisted of 15 closed-ended questions, rating instructors on a 5-point Likert scale. When analyzing the results, authors compared the actual gender of the assistant instructor, and the perceived gender of the assistant instructor. Upon comparison of perceived gender identities, males received higher scores on professionalism, promptness, fairness, respectfulness, enthusiasm, giving praise, and the student ratings index.

The results of the experiment reveal that the male identity was rated higher than the female identity, and results support their hypotheses that gender bias in ratings exists. Female instructors are expected to exhibit particular interpersonal traits and are not rewarded in their ratings when said traits are exhibited. This study is significant because it reveals that “gender bias is an important deficiency of student ratings of teaching” and “the use of student ratings of teaching as a primary means of assessing the quality of an instructor’s teaching systematically disadvantages women in academia.” (MacNell, Driscoll and Hunt 2015:301) This study’s findings align with other literature regarding the interaction of gender, student evaluations, and students’ gendered expectations.
Gender alone is not a consistent predictor of evaluation ratings (Hopefl and Frey-Hartel 1995). While female faculty are evaluated lower than males due to their gender; in some situations they are rated higher. The circumstances under which female earned a higher evaluation rating was typically a result of the interaction between gender and other variables, such as grade level. Additionally, when female instructors are perceived as warm, they earn high ratings.

Contrary to other sources, Basow, Codos and Martin (2013) found that SEIs did not vary by teacher gender and or with student gender. However, performance on a quiz did. In their experiment students watched a 3-minute engineering lecture presented by a computer animated professor. After the lecture, students completed an instructor evaluation form and a 10-question true/false quiz to assess teacher effectiveness. The study found that male students gave higher ratings than female students on most teaching factors, while students who had a male professor earned high quiz scores, possibly due to students paying more attention to the “normative professor.” (Basow, Codos and Martin 2013:352). Similar to previous sources mentioned, authors discuss the implications of student expectations of instructor’s gendered behavior. Women have to work harder to be perceived as equally competent as their male counterparts. Thus, student evaluations may not be a good indicator of teaching effectiveness. Unique compared to earlier sources, they suggest that gender bias is more difficult to discern than in the past, “but it may still operate at the level of student attention and interest and show up in student performance.” (Basow, Codos and Martin 2013:361).

In an effort to identify attitudes students hold toward professors and teaching assistants, Miles and Helsel (1989) conducted a survey of 30 questions, collecting 356 usable responses. The survey found students discount instructor age, dress, and gender when evaluating their
experience, and cosmetic differences do not strongly influence student evaluations of teaching. There was no clear indication in the study regarding student preference of male or female instructors. However, when respondents were divided between male and female, 60% of women disagreed with the statement “men make better professors than women”, and 30% of men disagreed with the same statement (Miles and Helsel 1989:82). One common complaint the study revealed was that instructors were not accessible to students, and teaching assistants were more accessible than professors. This survey is significant as it reveals that “professionalism, enthusiasm, and course content are very important to students and that variables such as dress have negligible impact on their learning.” (Miles and Helsel 1989:83) This survey is unique as its findings are contrary to previously reviewed sources regarding gender preference of instructor.

Focusing on the relationship between teacher expressiveness and teacher gender, Basow and Distenfeld (1985) conducted an experiment in which 121 students watched a video of an actor giving a lecture using expressive and non-expressive communication. Students completed an evaluation of the instructor (or “trait questionnaires”) and a content test following the lecture. The expressive instructor earned the highest student evaluations rating. Students who watched the lecture presented by a nonexpressive female teacher earned the highest achievement on the 15-short-answer-content test. Instructors who smile, gesticulate, and express enthusiasm are viewed as student-oriented, organized, stimulating, and knowledgeable (p. 50). Notably, teacher sex did not have a main effect on student evaluations of teachers.

Arbuckle and Williams (2003) also studied the relationship between student perception of teacher expressiveness and gender stereotypes. In this laboratory experiment “three hundred male and female students watched slides of an age-and gender-neutral stick figure and listened to a neutral voice presenting a lecture, and then evaluated it on teacher evaluation forms indicated 1
of 4 different age and gender conditions”; students rated the “young” male higher than they did the “young” female (Arbuckle and Williams 2003:507). In this experiment, student ratings varied based on who was exhibiting enthusiasm, conscientiousness, acceptance, interest, desirable voice tone, and confidence. They found no interaction between teacher gender and student gender. These findings are suggestive that “students may expect their professors to express themselves in the classroom according to socio-cultural scripts that limit full human development and influence people to appraise men and women differently even when they are displaying identical behaviors.” (Arbuckle and Williams 2003:514)

Basow and Howe (1987) found contradictory results. In this study, approximately 700 students evaluated written profiles of professors. Professors’ sex-type significantly corresponded with students’ sex and instructors’ sex on questions relating to interactions with students outside of the classroom. The study was conducted at a small liberal arts college, 45% of the sample were women, and 34% of the participants were freshmen, 30% sophomores, 20% juniors, and 16% seniors. Four of the professors’ profiles used adjectives such as “third year of teaching, Ph.D., from prestigious university, three papers, two published articles, working on book manuscript, high evaluations and enrollments”, this was the “control” paragraph (p. 672). The affective descriptions used the control paragraph and the sentences “Dr. Jones is a concerned individual who is available for students outside of class. She/he is a likeable person and has a good rapport with her/his students.”; instrumental, which was the control paragraph plus “Dr. Jones is a powerful individual, active in the department and on faculty committees. She/he is effective in her/his work and in communicating to students”; and the androgynous profile- the control paragraph plus the affective and instrumental descriptions.
Results show the professors’ sex-type impacts their evaluation, the androgynous professor earning the highest rating, and the control professor the lowest. The primary finding from this study is “students’ sex primarily reflects female students rating themselves as more willing to take another course with the professor than are male students.” (Basow and Howe 1987:677) Furthermore, as men and women possess different and stereotyped qualities, “the effect of professors’ sex does not appear. The personal qualities of the professor, here extremely positive ones, appear to outweigh the importance of the professors’ sex.” (Basow and Howe 1987:677) In conclusion, these results provide context to the ratings of instructors and the impact of stereotyped qualities on students’ likelihood of taking another course with the instructor.

Feldman (1993) acknowledges the insignificant association between gender and overall evaluation, in practical terms, and on some dimensions female teachers receive very slightly higher ratings on their sensitivity to and concern with class level than men. He also reiterated students rate same-gendered teachers higher than opposite gendered expectations and that evaluation ratings can be enhanced by gender-typical and sometimes gender-atypical behaviors and attributes. In his review, Feldman (1993) concludes that students’ expectations of the two genders could underlie similar ratings, and may underlie differences in ratings, as students possess varying expectations between genders. However, it should be noted that any differences in students’ views in their teachers, they may have helped to encourage or to “produce” by their own expectations and demands. For example,

Female faculty repeatedly expressed the sentiment that “there is no middle ground for a woman,” and resentment at the stark choice they felt they faced. As one put it, “If you open that door but the least crack, the flood rushes in. So, you have to make it clear from
the start that you’re one of the boys in this respect—and that means that you’re too busy
to be every student’s mother.” Yet, the same woman also spoke for many others when
she continued, “but the feminist in me rebels at having to be [one of the boys]. I feel
badly about that, that I’m cheating my students and myself denying part of my nature and
responsibility to care.” (Feldman 1993:179).

Women face a challenge, then: if she resolves the conflict in a stereotypical “masculine”
teaching style, she could be resented but if she addresses it in a “feminine” teaching style, she
may be deemed incompetent.

Elmore and LaPointe (1974) found similar results, revealing no difference in mean
ratings between male and female faculty. The one exception is male instructors were rated higher
on “spoke understandably” and female instructors received higher ratings on “promptly returned
homework and tests” (Elmore and LaPointe 1974:368). These conclusions were drawn from
course evaluations, from 1,474 courses during 1971.

Evaluation Scales

Student expectations, stereotypes, and exposure are not the only factors that contribute to
bias having an impact on SEIs. Scales of evaluation also have an influence on bias (Rivera and
Tilcsik 2019). In other words, a scale of 6 versus a scale of 10 influences the size of the gender
gap in student evaluations of teachers because the rating system can shape the distribution of
ratings. In their experiment, Rivera and Tilcsik (2019) find that a change in rating scale from 10
to 6 eliminated the gender gap in evaluations present in male-dominated fields. The smaller scale
of 6 reduced the opportunities for finer quality differentiation and masked gender differences in
teaching performance. Rivera and Tilcsik (2019) argue the number of categories is consequential
for the distribution of rewards, due to the hierarchical status system; through changing the
number of categories in an evaluative scheme, boundaries of status competitions are redesigned. Interpretations of the scale influence how students map ratings because the “meaning attached to specific numbers serve as cognitive anchors/implicit standards of quality, shape raters’ likelihood of assigning specific scores to particular groups.” (Rivera and Tilcsik 2019:11). Meaning some numbers, such as 10, have cultural implications (10 signifying perfection, exceptional, or brilliant). Research reveals that people are less likely to assign the highest rating to a person on a 10-point scale. Therefore, a 6/6 rating is “less exclusive and could encompass not only flawless performance but also what raters perceive as (merely) very good performance.” (Rivera and Tilcsik 2019:12).

In two experiments, Rivera and Tilcsik (2019) investigate gender bias in SEIs by examining the potential impact of reducing the number of points on the rating scale. In the survey experiment, participants received identical transcripts of a lecture and randomly assigned instructor gender and the size of the rating scale. The results revealed the likelihood of receiving a top score of 10/10 was 5 percentage points lower for women than for men, and female instructors are more likely to receive the top score under the new lower-ranging rating scale. Women benefitted more from a 6-point scale; their likelihood of earning a 6/6 increased by 10 percentage points. Lastly, the ten-point scale produced a substantial gender gap based on stereotypes of brilliance that can harm women’s opportunities for hiring, promotion, and compensation. Due to the connection between “perfect 10” and idealized standard of beauty, a 10-point scale is prone to gender stereotypes. In sum, stereotypes not only impact students’ expectations and ratings, but how they use the numerical assignments that represent their assessments of teachers.
Open-ended comments

In addition to the scale of evaluations, open-ended comments also reveal gender bias. (Terkik et al. 2016) use sentiment analysis—that is, identifying positive and negative tone expressed in natural language text—to analyze eight years of student comments from college biology, chemistry, physics, and math courses and compared them for male and female faculty. They categorized the differences by: instructor competence, organization/presentation, personality/helpfulness and overall satisfaction. They also ranked words based on the strength of their co-occurrence with each gender, finding “students are significantly more likely to address their male instructors with more professional or formal titles such as Professor or refer to them by their last names. Female teachers are more likely to be addressed as teacher or instructor and more likely to be regarded as “wonderful” “organized” and “helpful” (Terkik et al. 2016:5). Men are more likely to be perceived as “funny” “knowledgeable” and likely to go on tangential topics (Terkik et al. 2016:6).

Mitigating Bias

Existing literature regarding mitigating gender bias in evaluations is limited. Recommendations include adapting rating scales, adding disclaimer language to evaluations, knowing learning style preferences, and acknowledging and working with the existing biases. The literature discusses both classroom level mitigations and systemic or institutional mitigations (Laube et al. 2007, Stark and Freishtat 2014).

Institutional recommendations

1. Raise awareness of bias, whether through training about bias and actions to mitigate it (Basow 2019, D'Entremont and Gustafson 2017, Easterly and Ricard 2011, Laube et al. 2007, Roper 2019), through pointing it out in the introduction to the evaluation
itself (Peterson et al. 2019), or introducing the idea through Harvard University’s Project Implicit’s implicit bias test (Project Implicit 2011, Roper 2019). In an effort to reduce reactance to a diversity training, use nonconfrontational language in presentation, use inclusive language explicitly acknowledge that everyone holds biases. Optional diversity training reduces reactance and increases personal buy-in (Easterly and Ricard 2011). Lastly, workshops can promote consciousness of the influence of gender behaviors and interaction inside and outside of the classroom (Laube et al. 2007).

2. Evaluations of teaching should measure specific behaviors (Andersen and Miller 1997, Basow 2019, D’Entremont and Gustafson 2017). Evaluation forms themselves should include objective questions based on traits or facts that are easily observable and behavioral based. For example, if a student is asked how knowledgeable their instructor is, this is not a fair question as students are not experts on this. Students also might be asked first to assess their own behavior, like how hard they worked in the class, and then asked about the teaching behavior they are qualified to assess. Students are not equipped nor are they credible judges prepared to evaluate pedagogy (Stark and Freishtat 2014). Others suggest focusing on goals and outcomes of the course, which would not be as vulnerable to student gendered expectations for the teacher (Sprague and Massoni 2005).

3. Evaluations of teaching should omit questions regarding expressiveness such as “enthusiasm, warmth, confidence, and voice tone” because of given gendered expectations of these qualities (D’Entremont and Gustafson 2017). Additionally eliminate questions on evaluations about “overall teaching effectiveness” and “value
of the course” from teaching evaluations, as they are misleading (Stark and Freishtat 2014) and “invite and obscure perceptual bias.” (Laube et al. 2007:96)

4. Use a 6-point evaluation scale, rather than a 10-point scale, to avoid the cultural affiliations with the number 10 and its implications of brilliance and perfection (Rivera and Tilesik 2019).

5. Universities should also offer training in interpretation of teaching evaluations (Laube et al. 2007). Evaluators of faculty need to remember that students evaluate likability and competence for men and women differently (Andersen and Miller 1997) and the effect of returning grades on evaluation—the impact is significantly negative for female instructors (Buser, Hayter and Marshall 2019). When examining SEIs, do not average of compare averages of SEIs; report instead the distribution, number of responders, and the response rate (Sprague and Massoni 2005, Stark and Freishtat 2014). Avoid comparing teaching in courses of different types, levels, sizes, functions, or disciplines.


Classroom level recommendations

8. Women faculty can work with existing bias. That is, dress professionally and communicate credentials. Some suggest starting each term in an authoritative manner and communicate your credentials (Kajfez 2019, Miller and Chamberlin 2000, Ulstad 2019) while others suggest trying harder to build rapport, to appear caring and nurturing (Abrams 2019, Basow 2019).

9. Women should be confident and explain their grading practices. Tell students that grading policies are to their benefit, for example, “I’m trying to help you by being strict in grading.” (Basow 2019)

10. Train students on how to give feedback (Kajfez 2019, Siston 2019).

11. Inside of classrooms, individual instructors can raise concerns about gender and evaluation in classroom, “encouraging students to think critically about these practices and to consider how they may be evaluating faculty unfairly due to preconceived ideas about members of certain social categories.” (Laube et al. 2007:96) Over the course of a term, instructors can build their own case for teaching effectiveness.

**Future Research**

Overall, few studies included specific recommendations to mitigate gender bias in evaluations, and there is no research examining the effectiveness of strategies to mitigate gender bias in evaluations. STEM specific sources were also limited, as many studies include students from a variety of disciplines. The research was not framed through the lens of gender bias in evaluations in STEM as a specific discipline. Furthermore, in the majority of studies there was no control for divisional affiliation, therefore it is often unknown students’ previous exposure to
instructors in a given field or their present exposure. There is also limited literature that allows us to conclude if gender bias varies by discipline.

Classroom management recommendations were also missing as a means to address the bias. Recommendations found largely pertained to systemic changes. Much research conducted about implicit bias was centered on mitigating instructors’ implicit biases, not that of the students. Additionally, replication of (Petersen et al., 2019)’s use a disclaimer to mitigate bias with a larger sample is necessary to see if such a simple solution can yield impressive results at a larger scale.

**Conclusion**

In sum, the findings are controversial about whether gender bias exists in student evaluations of teachers; however, the majority of findings agree bias influences student ratings. As a result of student expectations, representation, student prior experience, stereotypes, and student gender, the biases influence student ratings of instructors and the language they use in comments in evaluations. All individuals possess implicit biases, which can be mitigated through diversity trainings, role models, and through increasing awareness of the presence of implicit biases. According to the literature, Ohio State College of Engineering faculty, and Dr. Basow possible mitigations of gender bias in student evaluations include authoritarian demeanor of female instructors, disclaimer text in evaluations to raise awareness of presence of evaluation, departmental midterm and semester end evaluations, 6-point evaluation scale, and teaching students how to give feedback. Moving forward, replication of the Peterson et al. study in Iowa is necessary (Petersen et al., 2019). I also recommend further research about role models as means to increase female representation in STEM (see Appendix G for preliminary findings),
and the impact of increased representation of gender bias on SETs. Increased classroom level mitigations are also needed, as well as domestic and STEM specific research.
References


Hopefl, Jean Garrity and Cheri Frey-Hartel 1995. "He's Tough but She’s Mean; Gender Differences in Student Evaluations " Paper presented at the University of Wisconsin System Conference on Women's Studies, Madison, WI.


*ScienceOpen Research.*


Note: more citations on the topic can be found at

Appendix A. Teaching Students How to Give Feedback Slides

Feedback Slides, Engineering Education Practicum 1, provided by Dr. Kajfez

What feedback is useful?

Effective Feedback

- State the assessment and support it.
  - Positive
    - Rachel is great!
    - Rachel is great because she comes to class prepared and answers questions very effectively!
  - Negative
    - This class sucks!
    - This class could be improved. For example, the content in this class is so important, we could meet for two hours instead of one each week!
Effective Feedback

• Focus on what is important and relevant.
  – Positive
    • I love Rachel's COE jacket!
    • The class is structured in such a way that we are encouraged to share our perspective and support our peers.
  – Negative
    • I hated the blue paper my syllabus was printed on.
    • The Lab Chart Assignment did not accomplish what was intended. Next time, I'd suggest that a specific topic was given for the comparison to scope the project.

Rev: 09302019
Appendix B. Mechanical and Aerospace Engineering’s Departmental SEI Materials

Instructor,

As explained at the March 22 faculty and staff meeting, the department will be using an online survey in place of the paper-based evaluation of instruction this semester. The Qualtrics surveys are now active for your courses and will remain open until the end of Reading Day Tuesday, April 23rd at 11:55 PM. A unique URL for the evaluation for each course can be found below, which we encourage you to post to your Carmen page. 

WHATEVER THE COURSE IS

The following text below can be emailed to your students as written or modified as you see fit.

Lastly, even though the survey will be open for 2 weeks, we encourage you to find ~5 minutes at the end of class to allow students to complete it on their phones/tablets/laptops/etc. We have heard from some students that they don’t complete the University SEIs because they have to make extra time in their schedule to do so, while they have dedicated time to complete the paper forms in class. Our experience with the pilot version of the online course evaluation this semester suggests that we might be able to have the best of both worlds if you provide them with some time in class. In fact, the response rates to the online surveys this semester were quite high, suggesting than a combination of time in and out of class, and reminders from you, can lead to a high submission rate.

Please let me or the advisors know if you have any questions.

Thanks again,

Rob

Dear students,

As you may know, the department conducts evaluations of instruction at the end of every semester. Historically, you have been given a few minutes at the end of a class to complete those evaluations on paper. However, this semester, the department piloted a new online evaluation of instruction for the department. The Undergraduate Studies Committee believes that this new tool could have several potential benefits:

- Instructors can receive better feedback on how to improve their teaching.
- The feedback can be shared with the instructors faster than what is done with the paper evaluations.
- The department can be environmentally responsible by not making literally thousands of copies for evaluations each semester.
Based on the success of that pilot, you will complete your evaluation of instruction for the department via an online survey. The online evaluation for this class will be available for 2 weeks, until the end of Reading Day on Tuesday, April 23rd at 11:55 PM, and can be found at this link [link inserted].

Please know that your instructor will receive only summary data and anonymous comments from the survey. You may go back and change earlier parts of the evaluation after you start it and do not have to complete it in one sitting. You will be asked to login with your name., which allows the system to save your responses for this sort of editing.

You will also be receiving a link to the University’s own evaluation of instruction. We request that you complete both. For additional information as to why the department created its own survey, how it differs from the University version, and some tips on how to effectively and professional provide feedback to your instructor and the department, please refer to this ~9 minute video.

Thanks again for your participation. If you have any questions, please email maeadvisor@osu.edu.
Usage Note

The department evaluation is built in Qualtrics and is formatted in modules. Every course has common “blocks” or modules; one “block” is specific to each course and is based on objectives for each course as provided by ABET (Accreditation Board for Engineering and Technology). Through Qualtrics, responses rates are in real time, and Dr. Siston is able to see who is advertising and not advertising the survey and can remind faculty to remind students to complete the evaluation. At the end of the data collection, instructors are able to see tables, charts, and free responses with student feedback. Students must log in using their university log in credentials, and therefore if any student writes a questionable response, survey administrators can see who submitted the questionable response.
End of Term Evaluation - Sample

End of Term Course Evaluation
Thank you for taking the time to complete this online survey to evaluate the quality of instruction you have received this semester.

Course: ME 3870
Instructor: Vishnu Sundaresan
Time: MWF 8:00AM - 8:55AM
Class Number: 10513

This evaluation is divided into 5 sections:
- The design of the course, separate from the practices of the instructor
- Practices of the instructor, separate from the design of the course
- Your engagement with the course
- Your assessment of how well you achieved the course objectives
- An overall evaluation, which is a combination of all aspects of the course

Please know that your instructor will receive only summary data and anonymous comments from the survey. You may go back and change earlier parts of the evaluation after you start it and do not have to complete it in one sitting.

The following information is being collected for informational purposes only. Please know that your instructor will receive only summary data and anonymous comments from the survey.

- First Name (1) ________________________________
- Last Name (2) ________________________________
- OSU Email (3) ________________________________
**Course Design** This section concerns the structure of the course. We encourage you to think about the course itself, independent of the teaching practices of the instructor.

<table>
<thead>
<tr>
<th>Strongly disagree (1)</th>
<th>Disagree (2)</th>
<th>Neither agree nor disagree (3)</th>
<th>Agree (4)</th>
<th>Strongly agree (5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>The course objectives were well explained.</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>The course assignments were related to the course objectives.</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>The number and length of course assignments were reasonable.</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>The amount of material presented in this course was reasonable.</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>The pace at which the material was covered was reasonable.</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>I understood what was expected of me in this course.</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
</tbody>
</table>
**Instructional Practices**  This section concerns the actions and practices of the instructor. We encourage you to think about the teaching practices of the instructor, independent of the design of the course.

<table>
<thead>
<tr>
<th></th>
<th>Strongly disagree (1)</th>
<th>Disagree (2)</th>
<th>Neither agree nor disagree (3)</th>
<th>Agree (4)</th>
<th>Strongly agree (5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>The instructor carefully explained difficult concepts, methods, and subject matter.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The instructor encouraged and was receptive to questions and discussion from the students during class.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The instructor encouraged and was receptive to questions and discussion from the students during office hours.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The instructor encouraged me to do my best work.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The instructor included materials, examples, etc. for a variety of applications.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Assessment

<table>
<thead>
<tr>
<th>Strongly disagree (1)</th>
<th>Disagree (2)</th>
<th>Neither agree nor disagree (3)</th>
<th>Agree (4)</th>
<th>Strongly agree (5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>The homework assignments reflected what was covered in the course.</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>The quiz/exam questions reflected what was covered in the course.</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>The grades that I have received reflected the quality of my performance in the course.</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>The criteria for good performance on assignments/quizzes/exams was clearly communicated.</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>The instructor’s evaluations of my performance were timely, constructive, and facilitated my learning.</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
</tbody>
</table>

**What practices used by the instructor, if any, most facilitated your learning?**

________________________________________

________________________________________

________________________________________

**What practices used by the instructor, if any, most hindered your learning?**

________________________________________

________________________________________

________________________________________
**Student Engagement**  This section considers your participation in the course, as separate from how the course was designed or taught.

**Approximately how many hours did you spend doing homework or otherwise studying for this course?**

- [ ] 1-3 hours per week
- [ ] 4-7 hours per week
- [ ] More than 7 hours per week

**Approximately how many times did you interact with the instructor outside of class? (via email, office hours, etc.)**

- [ ] 1-3 times
- [ ] 4-7 times
- [ ] More than 7 times
- [ ] I did not interact with the instructor outside of class

**Self-Assessment**  In this section, you are asked to rate your own attainment of each of the course objectives using the standard Ohio State grading scale. Your responses in this section
are one way that the department can monitor how much is being learned in each class.

<table>
<thead>
<tr>
<th>ABET Course Objective – ME 3870</th>
<th>A (4)</th>
<th>B (3)</th>
<th>C (2)</th>
<th>D (1)</th>
<th>E (0)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Understand analog and digital wave forms and data acquisition.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Understand signal conditioning.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Understand noise sources and their mitigation.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Understand time and frequency domain analysis.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Understand sensor selection and spec sheets, uncertainty analysis, and error propagation.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Understand basic probability and statistics, emphasizing the statistical nature of measurements.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Understand a variety of specific measurement transducers/instruments/techniques commonly used in both industry settings and academic research in mechanical engineering.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A complimentary understanding of the core material from fluid, thermal, mechanical and dynamic systems in mechanical engineering and their measurement.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
**Overall Evaluation**  This section considers the course as a whole, taking aspects of its design, instruction, and your participation together.

<table>
<thead>
<tr>
<th></th>
<th>Strongly Disagree (1)</th>
<th>Disagree (2)</th>
<th>Neither Agree nor Disagree (3)</th>
<th>Agree (4)</th>
<th>Strongly Agree (5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>I have learned perspectives, principles, or practices from this course that I expect to apply to new situations.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>This course challenged me to think critically and communicate clearly about the subject.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>This course provided me with information that is directly applicable to my career and academic goals.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Please describe the most valuable aspect of the course.**

_________________________________________________________________
_________________________________________________________________
_________________________________________________________________
_________________________________________________________________
Please describe the least valuable aspect of the course.

________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________

Overall rating for the instructor this semester:

○ 1, Poor
○ 2, Below Average
○ 3, Average
○ 4, Above Average
○ 5, Excellent

Is there additional feedback that you would like to share with the course instructor?

________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________

We thank you for taking the time to complete this survey. Please click the button below to submit your response.
Appendix C. Summary of Peterson et al. (2019)

Peterson et al. (2019) used the standard faculty on-line evaluation used at Iowa State University and added a disclaimer regarding implicit bias to a control group. In this randomized experiment, student evaluations from 4 large classes, 2 with male instructors and 2 with female instructors, were analyzed to discover the impact anti-bias language had on ratings of professors. The experiment found the anti-bias language resulted in higher ratings of female instructors, compared to the students in the standard treatment group. The survey was conducted in an introductory Biology course and an introductory American Politics course in Spring 2018. All instructors were white. The language used in the survey for the treatment condition is as follows:

Student evaluations of teaching play an important role in the review of faculty. Your opinions influence the review of instructors that takes place every year. Iowa State University recognizes that student evaluations of teaching are often influenced by students’ unconscious and unintentional biases about the race and gender of the instructor. Women and instructors of color are systematically rated lower in their teaching evaluations than white men, even when there are no actual differences in the instruction or in what students have learned.

As you fill out the course evaluation please keep this in mind and make an effort to resist stereotypes about professors. Focus on your opinions about the content of the course (the assignments, the textbook, the in-class material) and not unrelated matters (the instructor’s appearance). (Peterson et al. 2019:3)

The above prompt was created to “overcome bias because it makes the student aware of the possibility of bias (step 1), attempts to motivate them to suppress its effects (step 2), and
provides cues about what other considerations he or she could use when answering the questions about their instructor (step 3).” (Peterson et al. 2019:3) The findings of the experiment revealed the intervention of informing students about gender bias can significantly influence the evaluation of female instructors. The difference in rating was as much as half a point on a five-point scale (p. 8). In future studies, a larger sample size is needed.