Abstract
This article was migrated. The article was marked as recommended.

There has long been an underrepresentation of women in medicine and biomedical research; this is often described as a "leaky pipeline," as the more senior the level of rank, the fewer women are appointed. In evaluating a faculty member for promotion, student evaluations of faculty teaching are often considered; therefore, if gender biases are reinforced by student evaluations of teaching, the gender gap in faculty promotion could remain or increase. In this study, we examine student evaluations of faculty teaching in a graduate biomedical research training program, using data gathered during two academic years. While female faculty received higher quantitative ratings of teaching, subtle gender differences in language existed in the student comments, indicating that implicit biases about women may be present in student evaluations of faculty teaching.

Keywords
research education, gender disparities, faculty, evaluations, graduate
Introduction
There has long been an underrepresentation of women in medicine and biomedical research. This gender gap is often attributed to the so-called “leaky pipeline”; that is, the further along in a career in medicine or biomedical research one goes, the greater the attrition of women in the field (Villablanca and Howell, 2016). These various “leaks” occur at multiple points in the career trajectory for those in biomedical research, beginning with lower rates of women admitted during the graduate education admission process; additional “leaks” are found during the processes of faculty selection, promotion, and retention (Liu et al., 2019). Most of these differences are perceived from the top-down (that is, leadership levels), as the gender gap widens with increasing levels of promotion. Reasons for women leaving biomedical research are highly varied; however, many issues often cited are related to implicit and explicit bias, hostile work environments, sexual harassment, and other climate issues (Kacik, 2019).

To date, little has been done to explore whether there are differences in how female faculty in graduate biomedical training programs are perceived by their students. As student evaluations of faculty are often considered when determining whether to promote a faculty member, understanding whether the gender gap is perceived and/or reinforced by students’ own gender stereotypes and implicit biases (Burns-Glover and Veith, 1995). It has been established at the undergraduate level that student evaluations of faculty teaching are often rife with gender bias that is observable both in quantitative ratings of faculty and in open-text comments (gendered language) (Sprague and Massoni, 2005; Basow et al., 2006; MacNell et al., 2015). Research has shown that differences exist in student evaluations of faculty in medical schools; specifically, female faculty receive lower teaching evaluations (Morgan et al., 2016); however, this phenomenon has never been explored in graduate biomedical research education.

In this study, we examined student evaluations of faculty teaching in a translational biomedical research program, using historical data gathered as standard of practice over the past two years. Based on prior work in related medical and health professions education, we hypothesized that women would receive lower teaching evaluation grades than their male counterparts, and that this would be even further pronounced between faculty with PhDs versus MDs. Additionally, we examined whether the composition of teaching teams (male-male, female-male, female-female) had any effect on student evaluation grades.

Methods
Program Description
For this study, faculty evaluations were used from all courses offered through Mayo Clinic Center for Clinical and Translational Science (CCaTS) track within the Mayo Clinic Graduate School of Biomedical Sciences (MCGSBS) for the 2017-2018 and 2018-2019 academic years. The program includes approximately ~52 unique courses delivered once or twice a year, or every other year, with each course having between 1 and 3 course directors. Coursework includes content across the clinical translational spectrum, from basic laboratory and clinical research to population and policy-focused work.

Student Evaluation Data
This study utilized data gathered as part of the standard CCaTS post-course evaluation process. A standard evaluation form is sent to students at the end of each quarter for each course in which they are enrolled. The evaluation is anonymous and optional, though with high completion rates. The primary item of interest for this study was “Please assign a grade for the co-director, [name here], and comment on what you liked or can be improved.” This item included both quantitative data (assigning faculty a grade) and qualitative data (commenting on areas for improvement in an open text box). Students could assign faculty a letter grade (A, B, C, D, F), which was assigned a corresponding number on a Likert scale (4, 3, 2, 1, and 0, respectively). Data were used in aggregate (average for each course delivery) for each faculty member; that is, each faculty member received a mean grade for each course, yielding faculty*course units.

Faculty Information
For each of the courses included, faculty degree, rank, gender, and team composition were gathered and incorporated into student evaluation feedback. Several faculty members teach multiple courses; this was taken into account during data cleaning and analysis. Once data were merged for the quantitative and qualitative data sets, faculty names and course numbers were removed.

Data Analysis
Quantitative data were analyzed to generate descriptive statistics, and to test for differences between males and females using chi-squared/fisher’s exact and/or one-way and two-way ANOVA tests, depending on the variable and data type.

For qualitative data, two reviewers separately coded each as having gendered language (Y/N); containing positive, negative, or mixed feedback; and key words (adjectives, adverbs specifically) were noted.
**Results/Analysis**

There were 57 unique faculty members who taught at least one course during the 2017-2018 and 2018-2019 academic years; a total of 88 course offerings were delivered within the CCaTS program during those two years (Table 1). As most courses are team-taught by 2 or 3 faculty members, this yielded 144 faculty*course units. There were statistically significant differences in terms of highest degree attained by gender (p=0.02). Of these, 24 (42%) were female and 33 (58%) were male. Among the 32 faculty with a PhD, 19 (59.4%) were female and 13 (40.6%) were male; among the 17 faculty with MDs, 4 (23.5%) were female and 13 (76.5%) were male. There were no female faculty with dual MD/PhD degrees; however, there were 3 male faculty members who held both.

There were also statistically significant differences in rank by gender (p=0.018; Table 1). Of the 10 who held the rank of Assistant Professor, 8 (80%) were female; 7 of the 14 (50%) who held the rank of Associate Professor were female; and only 6 of 24 (25%) of individuals who held the rank of full Professor were women. There were not enough data to analyze fields of rank for each faculty member; however, more women held rank in epidemiology and health services researchers.

**Table 1. Demographic characteristics of unique faculty within the Clinical Translational Science Track at Mayo Clinic Graduate School of Biomedical Sciences who taught a course in the 2017-2018 and/or 2018-2019 academic years (n=57).**

|                      | Female n (%) | Male n (%) | p-value |
|----------------------|--------------|------------|---------|
| **Total**            | 24           | 33         |         |
| **Highest Degree**   |              |            | 0.020*  |
| Masters              | 1 (20.0%)    | 4 (80.0%)  |         |
| MD                   | 4 (23.5%)    | 13 (76.5%) |         |
| PhD                  | 19 (59.4%)   | 13 (40.6%) |         |
| MD/PhD               | 0 (0%)       | 3 (100.0%) |         |
| **Rank**             |              |            | 0.018*  |
| Instructor           | 0 (0%)       | 3 (100%)   |         |
| Assistant Professor  | 8 (80.0%)    | 2 (20.0%)  |         |
| Associate Professor  | 7 (50.0%)    | 7 (50.0%)  |         |
| Professor            | 6 (25.0%)    | 18 (75.0%) |         |
| NA                   | 3 (50.0%)    | 3 (50.0%)  |         |
| **Primary Field of Rank** |          |            | 0.149*  |
| Anesthesiology       | 0 (0%)       | 2 (100%)   |         |
| Biochemistry and Molecular Biology | 0 (0%) | 2 (100%) |         |
| Biomedical Ethics    | 0 (0%)       | 3 (100%)   |         |
| Biomedical Informatics | 0 (0%)       | 1 (100%)   |         |
| Biostatistics        | 5 (45.5%)    | 6 (54.5%)  |         |
| Epidemiology         | 5 (83.3%)    | 1 (16.7%)  |         |
| Health Services Research | 4 (66.7%) | 2 (33.3%) |         |
| Medical Education    | 1 (100%)     | 0 (0%)     |         |
| Medicine             | 3 (27.3%)    | 8 (72.7%)  |         |
| Neurology            | 0 (0%)       | 2 (100%)   |         |
| Pediatrics           | 1 (50.0%)    | 1 (50.0%)  |         |
| Pharmacology         | 1 (50.0%)    | 1 (50.0%)  |         |
| Physiology           | 0 (0%)       | 1 (100%)   |         |
| Psychology           | 2 (100%)     | 0 (0%)     |         |
| NA                   | 2 (40.0%)    | 3 (60.0%)  |         |

*Fisher’s Exact Test used (Pr <= P)
Table 2. Mean faculty grades by gender, degree, rank, and course director team composition within the Clinical Translational Science Track at Mayo Clinic Graduate School of Biomedical Sciences for academic years 2017-2018 and 2018-2019 combined (n=144 faculty*courses, based on 1,352 evaluations).

|                         | Overall (n=144) | p-value (overall) | Female (n=61) | Male (n=83) | p-value (Male/Female) |
|-------------------------|----------------|------------------|---------------|-------------|----------------------|
|                         | n  | Mean | Median | SD | n  | Mean | Median | SD | n  | Mean | Median | SD | p-value |
| Overall                 | 144 | 3.64 | 3.72   | 0.39 | -  | -    | -      | -  | 61  | 3.72 | 3.78   | 0.30 | 83  | 3.59 | 3.67 | 0.43 | 0.051*  |
| Highest Degree          | 0.188 | 0.147 |        |     | 0.147 |        |     |     |
| Masters                 | 22  | 3.57 | 3.59   | 0.27 | 2  | 3.70 | 3.70   | 0.05 | 20  | 3.56 | 3.57   | 0.28 | |
| MD                      | 37  | 3.71 | 3.80   | 0.35 | 7  | 3.95 | 4.00   | 0.13 | 30  | 3.65 | 3.71   | 0.36 | 0.147 |
| PhD                     | 78  | 3.65 | 3.76   | 0.35 | 52 | 3.69 | 3.78   | 0.31 | 26  | 3.59 | 3.70   | 0.41 | |
| MD/PhD                  | 7   | 3.40 | 3.71   | 0.93 | 0  | -    | -      | -  | 7   | 3.40 | 3.71   | 0.93 | |
| Rank                    | 0.969 | 0.647 |        |     | 0.647 |        |     |     |
| Instructor              | 12  | 3.57 | 3.57   | 0.27 | 0  | -    | -      | -  | 12  | 3.59 | 3.57   | 0.27 | |
| Assistant Professor     | 23  | 3.65 | 3.76   | 0.34 | 19 | 3.71 | 3.80   | 0.33 | 4   | 3.33 | 3.31   | 0.13 | |
| Associate Professor     | 28  | 3.68 | 3.78   | 0.40 | 18 | 3.70 | 3.78   | 0.32 | 10  | 3.64 | 3.77   | 0.52 | |
| Professor               | 60  | 3.73 | 3.73   | 0.45 | 15 | 3.71 | 3.78   | 0.34 | 45  | 3.60 | 3.71   | 0.48 | |
| NA                      | 21  | 3.71 | 3.71   | 0.30 | 9  | 3.77 | 3.74   | 0.16 | 12  | 3.57 | 3.63   | 0.35 | |
| Course Director Team Composition | 0.033* |        |     |     |     |     |     |     |
| Male (Solo)             | 19  | 3.55 | 3.58   | 0.41 | -  | -    | -      | -  | -   | -    | -      | -    | -    | |
| Female (Solo)           | 21  | 3.61 | 3.61   | 0.34 | -  | -    | -      | -  | -   | -    | -      | -    | -    | |
| Male (Team)             | 42  | 3.54 | 3.65   | 0.46 | -  | -    | -      | -  | -   | -    | -      | -    | -    | |
| Female (Team)           | 21  | 3.82 | 3.79   | 0.13 | -  | -    | -      | -  | -   | -    | -      | -    | -    | |
| Mixed Gender Team       | 41  | 3.72 | 3.88   | 0.37 | -  | -    | -      | -  | -   | -    | -      | -    | -    | |
than men, and more men held rank in medicine. Notably, biostatistics was fairly evenly split (5 [45.5%] female and 6 [54.5%] male).

Overall, there was a borderline statistically significant difference (p=0.051) in faculty grades between female and male faculty members, with female faculty members scoring slightly higher (Table 2). No statistically significant difference was found when adding highest degree earned to the analysis (p = 0.647; Table 2). However, when examining the composition of the teaching team for each class, significant differences were seen in grades, with male course directors teaching solo earning the lowest grades, and mixed gender course director teams earning the highest grades (p=0.033).

Review of qualitative comments revealed that there were few explicitly gendered comments made by students in providing faculty feedback. Our team noted that comments about female faculty tended to focus on personality and perceived authority, while comments about male faculty addressed expertise and included more superlatives. For example, students suggested that some female faculty “take more ownership,” appreciated that they were “passionate,” and were “approachable” and “nice.” Male faculty members were more frequently described as “dismissive,” “thoughtful,” “knowledgeable,” and “excellent.”

Discussion

Our findings are in stark contrast to those reported previously for faculty in medical schools, in which males receive more favorable grades for their teaching than women (Morgan et al., 2016). In this study, we found that women in the graduate research program received higher grades on average for teaching than their male counterparts. This is of particular interest as women with advanced degrees in biomedical science are increasingly leaving academia for industry and government due to personal, interpersonal, and professional factors (the proverbial “leaky pipeline”) (Sarseke, 2018; Yseldyk et al., 2019). Female faculty in the biomedical research space may face a catch-22 in that, while they are viewed more favorably by their students, they have a steeper climb for promotion and tenure in the academic world than their male counterparts (Carr et al., 2013). Despite the higher ratings quantitatively, subtle differences in language in student comments reinforce that gender patterns and biases still exist; this has been established previously in several other fields of academia (Basow et al., 2006).

Furthermore, these results support the fact that teaching teams involving women in the biomedical/translational sciences are rated more highly by students than male or female faculty teaching alone. This is in line with studies that have shown that teams involving women are more successful than those without women, as measured by collective intelligence, and that diversity is associated with creativity and productivity (Woolley and Malone, 2011; Rock and Grant, 2016; Kubik-Huch et al., 2019). Therefore, seeing this reflected in our education data shows that having women as part of the team and team success translates to graduate biomedical research education.

Conclusion

This is the first study that our team is aware of to explore gender differences in student evaluations of graduate biomedical research faculty. A strength of this study is that the CCaTS faculty is diverse in their areas of expertise which is reflected in the variety of courses offered through the center; therefore, this study encompasses faculty who teach basic, clinical, population health, and policy-focused courses. Limitations of this study include those associated with small sample sizes, such as limitations in the type of analyses we were able to conduct. Additionally, while the content taught by faculty spans the entire translational spectrum, these findings are only representative of one track within MCGSBS; data were not available school-wide, as CCaTS has led the way with implementing these evaluation procedures. Finally, we did not have data as to which students completed the evaluation surveys, so could not determine whether there were differences on faculty rating based on whether the students were male or female. This factor has been shown in other studies to influence the association between gender and faculty teaching evaluations (Santhanam and Hicks, 2002; Sprague and Massoni, 2005). Future work should emphasize faculty in more “basic” science settings and include data to explore whether faculty ratings depend on student gender as well as the gender of the faculty, and should be expanded to include non-binary gender categories (MacNell et al., 2015).

Take Home Messages

- There may be gender disparities in how students evaluate faculty in graduate biomedical research programs
- This may impact recruitment, retention, and promotion of female faculty
- These disparities may present either as quantitative grading of teaching or more subtly in word choices within qualitative feedback
Most of these disparities appear to be attributable to implicit biases held by students, rather than explicit biases.

Notes On Contributors
Alexandra Joelle Greenberg-Worisek, PhD, MPH is an Assistant Professor of Epidemiology and the Associate Director for Graduate Curriculum within the Center for Clinical and Translational Science at Mayo Clinic College of Medicine and Science in Rochester, Minnesota, USA.

Katherine Elizabeth Cornelius, MPH, is a program analyst and evaluation specialist within the Center for Clinical and Translational Science at Mayo Clinic College of Medicine and Science in Rochester, Minnesota, USA.

Karen Weavers, MEd, is a program manager within the Office of Applied Scholarship and Education Science at Mayo Clinic College of Medicine and Science in Rochester, Minnesota, USA.

Felicity Enders, PhD, MPH, is a Professor of Biostatistics and the Director for Curriculum within the Center for Clinical and Translational Science at Mayo Clinic College of Medicine and Science in Rochester, Minnesota, USA.

Declarations
The author has declared that there are no conflicts of interest.

Ethics Statement
This project was conducted as part of regular educational programming and was deemed by the Mayo Clinic IRB to be a “quality improvement” project.

External Funding
This publication was supported by Grant Number UL1 TR002377 from the National Center for Advancing Translational Sciences (NCATS). Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the National Institutes of Health.

Acknowledgments
The authors wish to acknowledge the support of CCaTS and MCGSBS for this project.

Bibliography/References

Basow, S. A., Phelan, J. E. and Capotosto, L. (2006) Gender patterns in college students’ choices of their best and worst professors. Psychology of Women Quarterly. 30(1), pp. 25–35. Reference Source

Burns-Glover, A. L. and Veith, D. J. (1995) Revisiting gender and teaching evaluations: Sex still makes a difference. Journal of Social Behavior and Personality. 10(4), p. 69. Reference Source

Carr, P. L., Gunn, C. M., Kaplan, S. A., Raj, A., et al. (2015) Inadequate progress for women in academic medicine: findings from the National Faculty Study. Journal of Women’s health. 24(3), pp. 190–199. Reference Source

Kacik, A. (2019) Time’s Up launches healthcare branch to address harassment. Available at: Reference Source (Accessed: 18 December 2019).

Kubik-Huch, R. A., Vilgrain, V., Krestin, G. P., Reiser, M. F., et al. (2019) Women in radiology: gender diversity is not a metric—it is a tool for excellence. European Radiology. pp. 1–9. Reference Source

Liu, S.-N. C., Brown, S. E. and Sabat, I. E. (2019) Patching the “leaky pipeline”: Interventions for women of color faculty in STEM academia. Archives of Scientific Psychology. 7(1), p. 32. Reference Source

MacNeill, L., Driscoll, A. and Hunt, A. N. (2015) What’s in a name: Exposing gender bias in student ratings of teaching. Innovative Higher Education. 40(4), pp. 291–303. Reference Source

Morgan, H. K., Purkiss, J. A., Porter, A. C., Lypson, M. L., et al. (2016) Student evaluation of faculty physicians: gender differences in teaching evaluations. Journal of Women’s Health. 25(5), pp. 453–456. Reference Source

Rock, D. and Grant, H. (2016) Why diverse teams are smarter. Harvard Business Review. 4(4), pp. 2–5. Reference Source

Santhanam, E. and Hicks, O. (2002) Disciplinary, gender and course year influences on student perceptions of teaching: Explorations and implications. Teaching in Higher Education. 7(1), pp. 17–31. Reference Source

Sarseke, G. (2018) Under-Representation of Women in Science: From Educational, Feminist and Scientific Views. NASPA Journal About Women in Higher Education. 11(1), pp. 89–101. Reference Source

Sprague, J. and Massoni, K. (2005) Student evaluations and gendered expectations: What we can’t count can hurt us. Sex Roles. 53(11-12), pp. 779–793. Reference Source

Villablanca, A. C. and Howell, L. P. (2016) Retention of women faculty in academia: Ensuring success. Mary Ann Liebert, Inc. 140 Huguenot Street, 3rd Floor New Rochelle, NY 10801 USA. Reference Source

Woolley, A. and Malone, T. (2011) What makes a team smarter? More women. Harvard business review. 89(6), pp. 32–33. Reference Source

Ysseldyk, R., Greenaway, K. H., Hassinger, E., Zutrauen, S., et al. (2019) A Leak in the Academic Pipeline: Identity and Health Among Postdoctoral Women. Frontiers in psychology. 10, p. 1297. Reference Source
This study is quite interesting that they examine student evaluations of faculty teaching in a graduate biomedical research training program, using data gathered through Mayo Clinic Center for Clinical and Translational Science (CCaTS) track within the Mayo Clinic Graduate School of Biomedical Sciences (MCGSBS) during two academic years. It indicates that female faculty play such important roles in the teaching team. Looking forward for their further study!

**Competing Interests:** No conflicts of interest were disclosed.

---

I enjoyed reading this paper and the analysis of the students' comments meant that it is an important contribution to this area. I hope the authors take this work further, possibly by undertaking focus groups with their students to explore in more detail this area of language and why student use the terms they do...
in relation to gender. Is it merely stereotyping or something more. Using student evaluations for promotional purposes is fraught with issues and concerns about bias but that's another study!

**Competing Interests:** No conflicts of interest were disclosed.

Reviewer Report 22 March 2020

https://doi.org/10.21956/mep.19994.r30442

© 2020 Shankar P. This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

**P Ravi Shankar**
American International Medical University

This review has been migrated. The reviewer awarded 4 stars out of 5

This is an interesting study examining gender differences in student evaluations of faculty. Contrary to earlier published research the authors found female faculty members had higher scores. The authors have mentioned some of the limitations of their study and also mentioned the strengths of the present study. The authors could consider providing some anonymized free text comments of student evaluation of faculty.

**Competing Interests:** No conflicts of interest were disclosed.

Reviewer Report 21 March 2020

https://doi.org/10.21956/mep.19994.r30440

© 2020 Silwimba F. This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

**Felix Silwimba**
University of Lusaka

This review has been migrated. The reviewer awarded 5 stars out of 5

this is a good study and is to follow through. it confirms the need for balanced faculty in terms of gender.

**Competing Interests:** No conflicts of interest were disclosed.
This review has been migrated. The reviewer awarded 4 stars out of 5

This article describes a really interesting study and contributes to ongoing work around gender differences and the representation of women in higher education generally. The results are heartening, and the differences between this study and earlier work are interesting. As is indicated in the article, however, further work is now needed to inform our understanding of where these differences come from, and why they are not seen to be filtering through to everyday discourse (as represented in the language used in the student comments). However, one obvious difference in this case is that this study was undertaken among postgraduate students versus previous work focusing on undergraduates and I did wonder whether the authors might not have looked to explore that further in the discussion. All in all, provocative early work – I look forward to seeing where it goes from here.

**Competing Interests:** No conflicts of interest were disclosed.

---

This review has been migrated. The reviewer awarded 4 stars out of 5

The result of the article echoed the situation in our institution that we have many women as heads of school. Great news that challenges the so-called "glass ceiling" and "leaky pipe".

**Competing Interests:** No conflicts of interest were disclosed.
This review has been migrated. The reviewer awarded 5 stars out of 5

Considering the so-called "glass ceiling" as a common barrier in the careers of women around the world, this article is further evidence that the field of biomedical research is also not free from gender stereotypes and bias. Interestingly, the study revealed conflicting results that women at the Mayo Clinic received, on average, higher grades for teaching than their male counterparts; Despite these amazing results, differences in ranks (according to the data presented) confirm the existence of a "glass ceiling", at least as it seems in this school, for example, by the figures, the number of PhDs among women prevails over PhD men, but only one third of women PhD are professors. A gender imbalance in “favor” of women with a lower rank of associate professor is also very noticeable. As you know, an important role in the formation of gender stereotypes and attitudes is played by organizational culture and spirit. There is no need to talk about the necessity to mainstream gender equality at all levels of any structure, of course, this can affect the student's assessment, even at an unconscious level. It would be interesting to continue research on how a student's gender affects the assessment process. Many thanks to the authors for sharing this very interesting survey. I will wait for further results.

**Competing Interests:** No conflicts of interest were disclosed.