

# Student Evaluation of Faculty Physicians: Gender Differences in Teaching Evaluations

Helen K. Morgan, MD,<sup>1,2</sup> Joel A. Purkiss, PhD,<sup>2,3</sup> Annie C. Porter, MD,<sup>4</sup>  
Monica L. Lypson, MD, MHPE,<sup>2,5</sup> Sally A. Santen, MD, PhD,<sup>2,6</sup>  
Jennifer G. Christner, MD,<sup>7</sup> Cyril M. Grum, MD,<sup>8</sup> and Maya M. Hammoud, MD<sup>1,2</sup>

## Abstract

**Purpose:** To investigate whether there is a difference in medical student teaching evaluations for male and female clinical physician faculty.

**Methods:** The authors examined all teaching evaluations completed by clinical students at one North American medical school in the surgery, obstetrics and gynecology, pediatrics, and internal medicine clinical rotations from 2008 to 2012. The authors focused on how students rated physician faculty on their “overall quality of teaching” using a 5-point response scale (1 = Poor to 5 = Excellent). Linear mixed-effects models provided estimated mean differences in evaluation outcomes by faculty gender.

**Results:** There were 14,107 teaching evaluations of 965 physician faculty. Of these evaluations, 7688 (54%) were for male physician faculty and 6419 (46%) were for female physician faculty. Female physicians received significantly lower mean evaluation scores in all four rotations. The discrepancy was largest in the surgery rotation (males = 4.23, females = 4.01,  $p = 0.003$ ). Pediatrics showed the next greatest difference (males = 4.44, females = 4.29,  $p = 0.009$ ), followed by obstetrics and gynecology (males = 4.38, females = 4.26,  $p = 0.026$ ), and internal medicine (males = 4.35, females = 4.27,  $p = 0.043$ ).

**Conclusions:** Female physicians received lower teaching evaluations in all four core clinical rotations. This comprehensive examination adds to the medical literature by illuminating subtle differences in evaluations based on physician gender, and provides further evidence of disparities for women in academic medicine.

## Introduction

MEDICAL STUDENT EVALUATIONS of teaching faculty are used to assess teaching quality, and medical schools frequently use teaching evaluations to guide decisions about reappointment, promotion, and pay increases.<sup>1,2</sup> These evaluations have been reported to be reliable<sup>3</sup> and have validity.<sup>4</sup> Many studies have sought to identify the attributes that learners associate with effective clinical teachers. Enthusiasm, active involvement of the learner, clinical competence, sensitivity to patients, and recognition of a student's limits have been reported as important qualities of effective teachers.<sup>5-8</sup> The past research has focused on largely *modifiable* characteristics leading to favorable teaching evaluations.

The potential relationship between the nonmodifiable characteristic of physician gender and medical student evaluations has not been comprehensively evaluated. The conceptual framework regarding the association between gender and teaching evaluations is complex; multiple variables, including the gender of the learner and the context of the learner-evaluator interaction, can contribute to disparate findings.<sup>9,10</sup> Students may also possess “gendered expectations” about the appropriate comportment for their teachers, with learners evaluating faculty differently based on the teacher's field of expertise.<sup>11-13</sup> It has been reported that small studies that only examine one discipline are more likely to find a gender bias. This is thought to be secondary to the methodological limitations of smaller studies.<sup>14</sup> For example, if a study only examined

Departments of <sup>1</sup>Obstetrics and Gynecology and <sup>2</sup>Learning Health Sciences, University of Michigan Medical School, Ann Arbor, Michigan.

<sup>3</sup>The Office of Medical Student Education, University of Michigan Medical School, Ann Arbor, Michigan.

<sup>4</sup>Department of Obstetrics and Gynecology, Washington University School of Medicine, St. Louis, Missouri.

Departments of <sup>5</sup>Internal Medicine and <sup>6</sup>Emergency Medicine, University of Michigan Medical School, Ann Arbor, Michigan.

<sup>7</sup>Department of Pediatrics, Baylor College of Medicine, Houston, Texas.

<sup>8</sup>Department of Internal Medicine, University of Michigan Medical School, Ann Arbor, Michigan.

Poster presentation of initial data presented at APGO/CREOG national conference, March 2013, and AAMC GWIMS Poster Session, November, 2013.

evaluations for a small number of faculty, then individual personal differences could have a larger contribution to differences in evaluation than gender. At this current time, the examinations of physician gender and learner evaluations have predominantly been limited to small studies. Female physician faculty received lower teaching evaluations from medical students in an ambulatory care setting<sup>15</sup>; likewise, psychiatry and gastroenterology female physician faculty received lower teaching evaluations from residents in their respective fields.<sup>16,17</sup> An analysis of medical student evaluations of obstetrics and gynecology residents revealed that male gender was one of the strongest predictors of being identified as an excellent teacher.<sup>18</sup> The purpose of our study was to investigate whether a difference exists in clinical medical students' evaluations of male and female physician faculty on four required clinical rotations. We analyzed ratings of overall teaching quality for significant differences based on the gender of both the physician faculty and the medical student evaluator. This comprehensive examination adds to the medical literature by illuminating possible subtle differences in evaluations based on physician gender.

### Materials and Methods

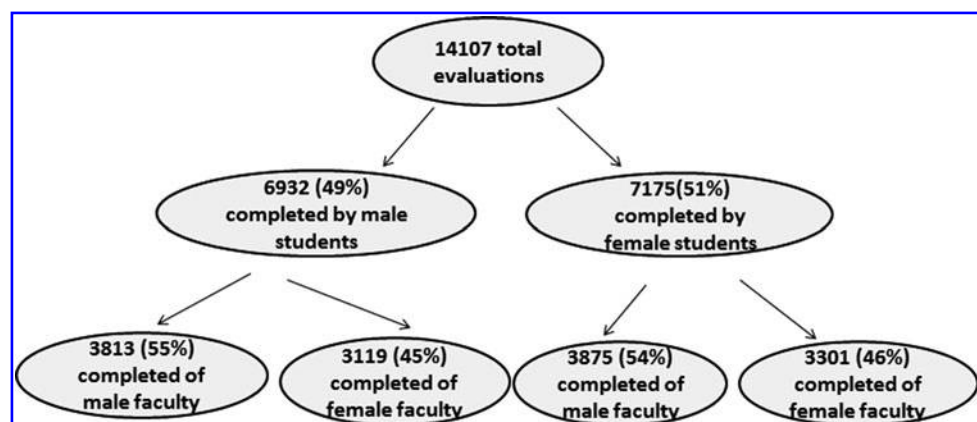
This retrospective study was performed at a single, large, public North American medical school. The third year of the 4-year curriculum consists entirely of seven required core clinical rotations (surgery, internal medicine, pediatrics, obstetrics and gynecology, neurology, family medicine, and psychiatry). While we theoretically could have examined evaluations in all of the seven clerkships during the study period, this would have involved an examination of a large number of specialties. We decided to focus on two surgical rotations (obstetrics and gynecology and surgery) and two nonsurgical rotations (internal medicine and pediatrics). During the rotations, medical students are integrated into the healthcare teams in both the inpatient and outpatient settings. Each rotation lasts 6–8 weeks, and the students complete a Clinical Teaching Assessment (Supplementary Data; Supplementary Data are available online at [www.liebertpub.com/jwh](http://www.liebertpub.com/jwh)) of faculty that they have worked with at the completion of each of their clinical rotations. The evaluations are completed confidentially through a web-based system. The students rate faculty on multiple dimensions, including an item on "Overall quality of teaching," which utilizes a 5-point response scale (1=Poor to 5=Excellent). All evaluations completed by clinical medical students for the surgery, ob-

stetrics and gynecology, pediatrics, and internal medicine rotations were examined from 2008 to 2012. The study was exempt from review by the medical school's Institutional Review Board.

Given the nature of the data, responses to the teaching evaluation items could not be treated as independent observations since individual faculty were evaluated by multiple students across the 5 years of the study. Furthermore, students also evaluated multiple instructors. To address the resulting issue of crossed random effects in our data, we fit and interpreted linear mixed-effects models using the *lmer* and *language R* functions available in R statistical software (R version 2.15.0, R Foundation for Statistical Computing). These models correct for the issue of nonindependence of observations, and provide estimated mean differences in evaluation outcomes by faculty gender. To examine possible gender-based evaluation differences by discipline, we ran four models—one each for surgery, obstetrics and gynecology, pediatrics, and internal medicine. We included student gender in our models to control for any effect it may have on the relationship between faculty gender and teaching evaluation outcomes. We used Markov chain Monte Carlo (MCMC) estimation to derive *p*-values for the parameters generated by our linear mixed-effects models. We interpreted outcomes with *p*-values below the conventional  $\alpha=0.05$  level to be statistically significant.

### Results

There were 14,107 teaching evaluations of 965 faculty members. Of these evaluations, 7688 (54%) were for male physician faculty and 6419 (46%) were for female physician faculty. 6932 (49%) of the evaluations were completed by male medical students and 7175 (51%) were completed by female medical students. Figure 1 demonstrates the proportion of total evaluations that were completed by male and female medical students, and the proportion of evaluations that were completed within the student's gender. Of the 965 faculty members, 527 (55%) were male and 438 (45%) were female. The number of male and female faculty who received evaluations in each of the four clinical rotations is demonstrated in Table 1. The internal medicine (62% male, 38% female) and surgery (69% male, 31% female) rotations had higher proportions of male faculty. The obstetrics and gynecology (44% male, 56% female) and pediatric (37% male, 63% female) rotations had higher proportions of female faculty.



**FIG. 1.** Proportion of teaching evaluations submitted by male and female medical students, for male and female faculty, 2008–2012.

TABLE 1. NUMBER OF TEACHING EVALUATIONS OF PHYSICIAN GENDER BY DIFFERENT CLINICAL ROTATIONS, 2008–2012

<i>Clinical rotation</i>	<i>Male physician faculty</i>	<i>Female physician faculty</i>
Surgery	87	39
OBGYN	94	118
Pediatrics	61	105
Internal medicine	285	176
Overall	527	438

Female physicians received lower scores on the evaluation item “Overall quality of teaching” in all four clinical rotations, with the results demonstrated in Table 2. Linear mixed regression model results are further outlined in Table 2, along with Bayesian highest posterior density confidence intervals and MCMC-estimated *p*-values. The discrepancy was largest in the surgery rotation, with a mean score for male physicians of 4.23 and female physicians of 4.01 ( $p=0.003$ ). Pediatrics showed the next greatest difference (males: 4.44, females: 4.29,  $p=0.009$ ), followed by obstetrics and gynecology (males: 4.38, females: 4.26,  $p=0.026$ ), and internal medicine (males: 4.35, females: 4.27,  $p=0.043$ ). An analysis of faculty evaluations based on medical student gender revealed no differences.

### Discussion

This is the first comprehensive study to examine differences in medical students’ evaluations of clinical physician faculty by the same cohorts of students across multiple disciplines. We found that female faculty physicians received lower evaluations in four clinical rotations. The discrepancy was present in both surgical and nonsurgical clerkships and in both male- and female-predominant specialties. There was no difference in faculty evaluations based on medical student gender. Strength of our study was that we examined all Clinical Teaching Assessments completed by third-year medical students during this 5-year period on the four clerkships. Students voluntarily completed these evaluations for faculty that they have worked with. Since we do not capture the names of the individual faculty that each student works with, it is not possible to determine a response rate. The mean number of faculty evaluated per year, and the mean number of faculty evaluated per student, across the 5 years of study are available in Supplementary

Table 1. Our study found that there were gender disparities in faculty evaluations and should raise additional questions about future areas of inquiry.

Our findings differ from a large study of resident evaluations of faculty, which demonstrated no significant differences based on faculty gender.<sup>19</sup> That study also examined faculty across multiple disciplines and included a similar numbers of evaluations in the analysis. The inherent differences in context and time periods of interactions between faculty and medical students versus residents may have contributed to the dissimilar findings of our study. Medical students rotate on clinical rotations for a transient time period and often form the basis for their evaluation of faculty on limited interactions. In contrast, residents interact with faculty for longer periods of time, in multiple contexts. Residents therefore may be able to formulate their evaluations of faculty based on more substantive interactions. Preliminary data from a qualitative study of resident and medical students’ perceptions of faculty teaching did reveal differences between the two levels of learners, with residents appearing to value efficiency more than the students in that study.<sup>20</sup> Future studies should further compare how medical students and residents perceive differences in effective clinical teaching.

There are likely multiple etiologies that contributed to the differences; a limitation of our study was that continuous variables such as faculty age and seniority were not included in our analysis. Medical schools are now balanced in the ratio of male and female students, but senior faculty members are predominantly male<sup>21</sup> and the influence of seniority as a reflection of leadership may have contributed to these results. Future studies will need more in-depth measurements and analysis of multiple variables, including faculty age, gender, and seniority. It is important to also discuss and question the practical significance of these differences. While the overall evaluation discrepancy of 0.09 is small, and unlikely to be the sole contributor of a department’s decision to promote an individual, teaching quality is becoming increasingly important in promotion decision-making.<sup>22</sup> Furthermore, the differences in clinical teaching evaluations may also contribute to differences in the selection of teaching awards and recognitions within departments. Future studies should examine whether there is a difference in the gender distribution of teaching awards by medical students for clinical physician faculty. Unconscious bias refers to social stereotypes about certain demographics or groups of people that individuals form outside of their own conscious awareness.<sup>23</sup> It is unclear

TABLE 2. MEAN EVALUATIONS FOR MALE AND FEMALE FACULTY IN “OVERALL QUALITY OF TEACHING” 5-POINT RESPONSE SCALE (1=POOR TO 5=EXCELLENT) LINEAR MIXED REGRESSION, UNIVERSITY OF MICHIGAN MEDICAL SCHOOL, 2008–2012

	<i>Female faculty</i>	<i>Male faculty</i>	<i>Mean difference</i>	<i>95% CI</i>		<i>p<sup>b</sup></i>	<i>Number of evaluations</i>
				<i>Lower bound<sup>a</sup></i>	<i>Upper bound<sup>a</sup></i>		
Overall	4.24	4.33	−0.09	−0.15	−0.05	<0.001	14,107
Surgery	4.01	4.23	−0.22	−0.39	−0.07	0.003	3565
Obstetrics/Gynecology	4.26	4.38	−0.12	−0.24	−0.01	0.026	3144
Pediatrics	4.29	4.44	−0.15	−0.27	−0.04	0.009	3592
Internal medicine	4.27	4.35	−0.08	−0.17	−0.003	0.043	3854

<sup>a</sup>Bayesian highest posterior density confidence intervals.

<sup>b</sup>Markov-chain Monte Carlo estimated *p*-value.

whether the results of our study are due to unconscious biases of the medical students, and this question should be explored further with qualitative as well as quantitative analyses.

Our study raises many questions about the implications of our findings. For example, one major concern is that lower evaluation scores of female faculty may contribute to a promotion gap for women. Promotion rates are lower for female faculty compared to their male counterparts and they continue to be underrepresented in medical school leadership positions.<sup>24</sup> Stereotype threat is when individuals who are members of a group feel that they are at risk of confirming a negative stereotype about their social group. Our goal is not to exacerbate stereotype threat; which has been hypothesized as a potential contributor to this leadership underrepresentation.<sup>25</sup> Our hope is that these findings will raise awareness of gender discrepancies for faculty development, with the goal of increasing effective medical student teaching. The possibility of creating additional faculty development opportunities for teaching can be considered for female faculty. Similarly, medical students should be made aware that these differences in perceptions exist to improve their evaluation processes. There continues to be inadequate progress for women in academic medicine.<sup>26</sup> This comprehensive examination has illuminated subtle differences in evaluations based on physician gender and provides further evidence of disparities for women in academic medicine.

#### Author Disclosure Statement

No competing financial interests exist.

#### References

1. Beasley BW, Wright SM, Cofrancesco Jr J, Babbott SF, Thomas PA, Bass EB. Promotion criteria for clinician-educators in the United States and Canada: A survey of promotion committee chairpersons. *JAMA* 1997;278:723–728.
2. Atasoylu AA, Wright SM, Beasley BW, et al. Promotion criteria for clinician-educators. *J Gen Intern Med* 2003;18:711–716.
3. Solomon DJ, Speer AJ, Rosebraugh CJ, DiPette DJ. The reliability of medical student ratings of clinical teaching. *Eval Health Prof* 1997;20:343–352.
4. Williams B, Litzelman DK, Babbott SF, Lubitz RM, Hofer TP. Validation of a global measure of faculty's clinical teaching performance. *Acad Med* 2002;77:177–180.
5. Arah OA, Heineman MJ, Lombarts KMJM. Factors influencing post-graduate trainees' evaluations of clinical faculty member teaching qualities and role model status. *Med Educ* 2012;46:381–389.
6. Irby D, Rakestraw P. Evaluating clinical teaching in medicine. *J Med Educ* 1981;56:181–186.
7. Irby DM, Ramsey PG, Gillmore GM, Schaad D. Characteristics of effective clinical teachers. *Acad Med* 1991;66:54–55.
8. Mullan P, Sullivan D, Dielman T. What are raters rating? Predicting medical student, pediatric post-graduate trainee, and faculty ratings of clinical teachers. *Teach Learn Med* 1993;5:221–226.
9. Basow S. Student evaluations of college professors: When gender matters. *J Educ Psychol* 1995;87:656–665.
10. Sprague J, Massoni K. Student evaluations and gendered expectations: What we can't count can hurt us. *Sex Roles* 2005;53:779–793.
11. Burns-Glover AL, Veith DJ. Revisiting gender and teaching evaluations: Sex still makes a difference. *J Soc Behav Pers* 1995;10:69–80.
12. Basow SA, Phelan JE, Capotosto L. Gender patterns in college students' choices of their best and worst professors. *Psychol Wom Q* 2006;30:25–35.
13. Baker P, Copp M. Gender matters most: The interaction of gendered expectations, feminist course content and pregnant in student course evaluations. *Teach Sociol* 1997;25: 29–43.
14. Santhanam E, Hicks O. Disciplinary, gender and course year influences on student perceptions of teaching: Explorations and implications. *Teach Higher Educ*. 2002;7:17–31.
15. Leone-Perkins M, Schnuth R, Kantner T. Preceptor-student interactions in an ambulatory clerkship: Gender differences in student evaluations of teaching. *Teach Learn Med*. 1999; 11:164–167.
16. DeGroot J, Brunet A, Kaplan AS, Bagby M. A comparison of evaluations of male and female psychiatry supervisors. *Acad Psychiatry*. 2003;27:39–43.
17. Thackeray EW, Halvorsen AJ, Ficalora R, Engstler GJ, McDonald FS, Oxentenko AS. Gender differences in trainee evaluations of gastroenterology faculty. *Clin Gastroenterol Hepatol* 2011;9:186.
18. Ogburn JA, Espey EL, Dorin MH, Ming C, Rayburn WF. Obstetrics and gynecology post-graduate trainees as teachers of medical students: Predictors of excellence. *Am J Obstet Gynecol* 2005;193:1831–1834.
19. McOwen KS, Bellini L, Guerra C, Shea J. Evaluation of clinical faculty: Gender and minority implications. *Acad Med* 2007;82:S94–S96.
20. Huff NG, Roy B, Estrada CA, et al. Teaching behaviors that define highest rated attending physicians: A study of the post-graduate trainee perspective. *Med Teach* 2014;36:991–996.
21. Association of American Medical Colleges. Analysis in brief: Women in U.S. Academic Medicine and Science: Statistics and Benchmarking Report, 2012:2–58.
22. Curran DS, Stalburg CM, Xu X, Dewald SR, Quint EH. Effect of resident evaluations on obstetrics and gynecology faculty on promotion. *J Grad Med Educ* 2013;5:620–624.
23. Association of American Medical Colleges. Unconscious bias in faculty and leadership recruitment: A literature review, Volume 9, number 2. In: *Analysis in Brief*. Washington, DC: Association of American Medical Colleges, 2009.
24. Lautenberger D, Raezer C, Bunton S. The underrepresentation of women in leadership positions at US Medical Schools, volume 15, number 1–2. In: *Analysis in Brief*. Washington, DC: AAMC, 2015.
25. Burgess D, Joseph A, van Ryn M, Canes M. Does stereotype threat affect women in academic medicine. *Acad Med* 2012;87:506–512.
26. Carr PL, Gunn CM, Kaplan SA, Raj A, Freund KM. Inadequate progress for women in academic medicine: Findings from the national faculty study. *J Womens Health (Larchmt)* 2015;24: 190–199.

Address correspondence to:

Helen K. Morgan, MD  
 Department of Obstetrics and Gynecology  
 University of Michigan Medical School  
 1500 E. Medical Center Drive  
 L4000 Von Voigtlander Women's Hospital  
 Ann Arbor, MI, 48109

E-mail: hjkang@med.umich.edu

**This article has been cited by:**

1. Eglė Vaižgėlienė, Žilvinas Padaiga, Daiva Rastenytė, Algimantas Tamelis, Kęstutis Petrikonis, Cornelia Fluit. 2017. Evaluation of clinical teaching quality in competency-based residency training in Lithuania. *Medicina* . [[Crossref](#)]
2. Cody A. Nebeker, Marc D. Basson, Pam S. Haan, Alan T. Davis, Muhammad Ali, Rama N. Gupta, Robert L. Osmer, John C. Hardaway, Andi N. Peshkepija, Michael K. McLeod, Cheryl I. Anderson, Karen A. Chojnacki, Charles J. Yeo, Francesco Palazzo, Jeffrey M. Gauvin, Anthony S. Pozzessere, Rondi B. Gelbard, Keith A. Delman, Denny R. Martin, Joanna Y. Woo, Laura E. Tate, Nicolas Elliott, Robert G. Molnar, Christopher C. Pfeifer, Lawrence Narkiewicz, Shawn H. Obi, Daniel E. Smith. 2017. Do female surgeons learn or teach differently?. *The American Journal of Surgery* **213**:2, 282-287. [[Crossref](#)]
3. Templeton Kim. 2016. Impact of Gender on Teaching Evaluations of Faculty: Another Example of Unconscious Bias?. *Journal of Women's Health* **25**:5, 420-421. [[Citation](#)] [[Full Text HTML](#)] [[Full Text PDF](#)] [[Full Text PDF with Links](#)]