


Gendered Expectations: the Impact of Gender, Evaluation Language, and Clinical Setting on Resident Trainee Assessment of Faculty Performance



Virginia Sheffield, MD¹, Sarah Hartley, MD, MHPE¹, R. Brent Stansfield, PhD², Megan Mack, MD¹, Staci Blackburn, MD¹, Valerie M. Vaughn, MD, MSc^{1,3}, Lauren Heidemann, MD, MHPE¹, Robert Chang, MD¹, and Jennifer Reilly Lukela, MD¹ 

¹Department of Internal Medicine, University of Michigan, Ann Arbor, MI, USA; ²Graduate Medical Education, Wayne State University, Detroit, MI, USA; ³Department of Internal Medicine, University of Utah School of Medicine, Salt Lake City, UT, USA.

BACKGROUND: Gender inequity is pervasive in academic medicine. Factors contributing to these gender disparities must be examined. A significant body of literature indicates men and women are assessed differently in teaching evaluations. However, limited data exist on how faculty gender affects resident evaluation of faculty performance based on the skill being assessed or the clinical practice settings in which the trainee-faculty interaction occurs.

OBJECTIVE: Evaluate for gender-based differences in the assessment of general internal medicine (GIM) faculty physicians by trainees in inpatient and outpatient settings.

DESIGN: Retrospective cohort study

SUBJECTS: Inpatient and outpatient GIM faculty physicians in an Internal Medicine residency training program from July 1, 2015, to December 31, 2018.

MAIN MEASURES: Faculty scores on trainee teaching evaluations including overall teaching ability and Accreditation Council for Graduate Medical Education (ACGME) competencies (medical knowledge [MK], patient care [PC], professionalism [PROF], interpersonal and communication skills [ICS], practice-based learning and improvement [PBLI], and systems-based practice [SBP]) based on the institutional faculty assessment form.

KEY RESULTS: In total, 3581 evaluations by 445 trainees (55.1% men, 44.9% women) assessing 161 GIM faculty physicians (50.3% men, 49.7% women) were included. Male faculty were rated higher in overall teaching ability (male=4.69 vs. female=4.63, $p=0.003$) and in four of the six ACGME competencies (MK, PROF, PBLI, and SBP) based on our institutional evaluation form. In the inpatient setting, male faculty were rated more favorably for overall teaching (male = 4.70, female = 4.53, $p<0.001$) and across all ACGME competencies. The only observed gender difference in the outpatient setting favored female faculty in PC (male = 4.65, female = 4.71, $p=0.01$).

CONCLUSIONS: Male and female GIM faculty performance was assessed differently by trainees. Gender-based differences were impacted by the setting of evaluation, with the greatest difference by gender noted in the inpatient setting.

KEY WORDS: Graduate Medical Education; assessment/evaluation; gender bias; implicit bias; gender norms.

J Gen Intern Med 37(4):714–22
DOI: 10.1007/s11606-021-07093-w
© Society of General Internal Medicine 2021

INTRODUCTION

Gender inequity is pervasive in academic medicine. While the number of female physicians has steadily increased, a gender gap remains in both pay and promotion.^{1–3} Women have comprised more than 40% of medical students since 1995, yet continue to be underrepresented at higher levels of academic rank and leadership.^{4–6} Within internal medicine, women make up 52% of clinical instructors but only 38% of associate professors and 24% of full professors.^{5,6} Past work has demonstrated that the field of general internal medicine (GIM) is not immune to gender disparities.^{7–9} Even in the relatively “newer” field of hospital medicine within GIM, these disparities have been well documented.^{10–13} Despite abundant evidence regarding the existence of this “leaky pipeline” for women in academic medicine, the factors that contribute to its persistence are less well understood. To develop effective strategies to rectify gender inequities in academic medicine, we must first identify the factors that contribute to their existence.

Teaching evaluations are used to make decisions about promotion, rank, and leadership positions for clinician educators.¹⁴ However, the inherently subjective nature of evaluations introduces the risk of reflecting and amplifying evaluators’ underlying explicit (overt) or implicit (unconscious) biases.^{15–17} In previous work, learners rated instructors labeled as male significantly higher than instructors labeled as female regardless of the instructor’s actual gender, suggesting gender bias was playing a significant role in their evaluations.¹⁶ Others have demonstrated that descriptive language used in evaluation forms may influence how men and women are assessed.¹⁸ While the current literature is

This work was accepted as a poster submission to both the Society of Hospital Medicine Annual Conference 2020 and Alliance for Academic Internal Medicine 2020 (both conferences were cancelled due to COVID) and was presented as a poster at Associated of Program Directors of Internal Medicine Fall Meeting, APDIM Online, October 9, 2020

Drs. Sheffield and Hartley contributed equally as first authors.

Received November 30, 2020

Accepted July 30, 2021

Published online August 17, 2021

varied in terms of the impact of gender on teaching evaluations of clinical faculty in medical education, there is evidence that residents and medical students may also be vulnerable to such biases.^{19–23}

Gender-based social role theory and stereotype-based cognitive biases likely impact how women are evaluated and thus contribute to gender disparities. Specifically, agentic characteristics including decisiveness, instrumental competence, and assertiveness are considered more traditionally masculine, while communal characteristics such as compassion, empathy, and caring are considered more traditionally feminine.^{24,25} The wide-sweeping nature of these social norms results in preconceived “gendered expectations.” These expectations are further amplified in fields or roles that have been traditionally occupied by men, including procedural specialties and leadership roles in academic medicine.^{20,24} For women who demonstrate agentic behaviors as leaders of inpatient or cardiac resuscitation teams, these expectations may result in being penalized, a phenomenon previously described as “the double bind” or “role incongruity.”^{24,26–30}

Internal medicine residents who pursue a career in GIM may practice in the inpatient setting as a hospitalist or outpatient setting as a primary care physician. These distinct clinical settings also align with differing gender norms. Responsibilities traditionally seen as “agentic” including performing procedures and leading rapid response and resuscitation teams are more frequently performed by hospitalists,^{31,32} while primary care physicians practice in settings where “communal” traits such as strong physician-patient communication and collaborative care have greater emphasis.³³ The potential impact of gender-based expectations on learner assessment of faculty performance in these different clinical settings is unknown.

The objective of this study was to determine whether gender-based differences exist in the assessment of teaching performance of GIM attending physicians by residents and to explore the extent the language used to describe the quality, skill, or behavior being evaluated (agentic vs. communal) and the environment of the interaction (inpatient vs. outpatient) impact gender differences in assessment. We hypothesized that male faculty would receive more favorable ratings overall, as well as for skills and behaviors related to agentic traits or described with more classically agentic language. Similarly, we hypothesized women would be more favorably assessed for skills or traits described with more communal language. Finally, we hypothesized that male faculty would be rated more favorably in the inpatient settings, whereas female faculty would be assessed more favorably in the outpatient settings where communal characteristics may be more highly valued.

METHODS

Setting and Participants

Participants included GIM faculty who served as teaching attendings on inpatient general medicine services and GIM faculty who

supervised outpatient continuity clinics at a single Midwestern academic tertiary care center and an affiliated Veteran’s Administration hospital (VA) between July 1, 2015, and December 31, 2018. Inpatient services included four general medicine teams and two resident-based hospital medicine teams at the University hospital and four general medicine teams at the VA. A general medicine team was comprised of one senior resident, two interns, and a faculty member. Trainees rotated monthly and faculty rotated every half month. A resident-based hospital medicine team included two senior residents and a faculty member who worked together in half-month increments. The outpatient continuity clinics occurred either at a traditional academic GIM practice based at the University Hospital, the VA, or in one of four University-affiliated community-based GIM practices. Residents averaged one half day per week in continuity clinic, working with the same one or two attending physicians over the 3 years of training.

Faculty Evaluations

At our institution, inpatient faculty are evaluated by trainees at the end of a half month rotation, while faculty supervising primary care continuity clinics are evaluated by trainees in their clinic twice annually. Evaluations are performed using MedHub™ (Minneapolis, MN).

Inpatient faculty are evaluated by all trainees on the inpatient team including residents from the categorical Internal Medicine program ($N=136$ total, 45–46 trainees per year), Medicine/Pediatrics ($N=32$ total, 8 trainees per year), interns completing a preliminary year ($N=8$), and Anesthesia interns ($N=28$). Outpatient faculty are evaluated by trainees in continuity clinic from the categorical Internal Medicine Program ($N=136$).

The faculty evaluation tool was developed internally by our residency program leadership team in 2014–2015. The form utilizes a 5-point Likert scale to answer prompts organized using the ACGME competencies: medical knowledge (MK), patient care (PC), interpersonal and communication skills (ICS), professionalism (PROF), practice-based learning and improvement (PBLI), and systems-based practice (SBP). The form also includes a global assessment of overall teaching ability (Fig. 1).

Study Design

The faculty evaluation form was reviewed by two members of the study team (JRL and SH) for language or skills that, based on prior literature, represented gender norms for men (agentic terms such as “leader,” “confident” or “autonomy” and skills such as procedures) or women (communal terms such as “collaborative,” “empathy” and “compassion” or skills such as history gathering and physical exam).^{34–39} The PC and ISC competency sections contained phrases weighted toward traditionally “feminine” characteristics, while the competencies of MK and PBLI were more weighted toward traditionally “masculine” characteristics (Fig. 1). The competencies of SBP

All skills are evaluated on a 5-point Likert scale (1- Strongly Disagree, 2- Disagree, 3- Uncertain, 4- Agree, 5- Strongly Agree)

Medical Knowledge (MK)

- Actively identifies and applies current **best evidence** to guide diagnostic and therapeutic decisions to ensure high quality patient care
- Teaches principles and approaches to guide junior members of the health team to diagnose and treat undifferentiated and **emergent** conditions
- Teaches the rationale and risks associated with common studies/**procedures** and anticipates potential complications that may result

Patient Care (PC)

- Role models gathering subtle and reliable information from the patient for junior members of the healthcare team
- Role models how to perform an effective physical examination and identifies key findings targeted to patient's complaints

Interpersonal and Communication Skills (ICS)

- Establishes therapeutic relationships and models cross-cultural communication including with persons of diverse socioeconomic backgrounds
- Role models effective communication skills in challenging situations
- Consistently and actively engages in collaborative communication with all members of the team

Professionalism (PROF)

- Demonstrates empathy and compassion to patients and caregivers in all situations
- Treats patients and caregivers with dignity, civility and respect, regardless of race, culture, gender, ethnicity, age or socioeconomic status
- Treats all members of the healthcare team (e.g., nurses, students, residents) with dignity, civility, and respect, regardless of race, culture, gender, ethnicity or age.
- Allows appropriate **autonomy** and guidance to support House Officer in order to provide effective and safe patient care
- Creates an environment that supports learning and the well-being of trainees including awareness and attention to trainee fatigue, duty hour limitations and training program expectations
- Assumes **professional responsibility** regardless of the situation
- Demonstrates integrity, honesty, and accountability to patients, society and the profession

Practice-Based Learning and Improvement (PBLI)

- Maintains awareness of the situation in the moment, and **responds to meet situational needs** to ensure safe patient care
- Demonstrates a **commitment to lifelong learning** through pursuing emerging clinical questions identified when caring for patients (e.g., identification and application of current, relevant **scientific** literature)

Systems-Based Practice (SBP)

- Understands the roles and responsibilities of and effectively partners with all members of the team
- Provides timely and constructive feedback to learners
- Anticipates needs of patient, caregivers and future care providers and takes appropriate steps to address those needs
- Viewed by other team members as a **leader** in the delivery of high-quality care

Overall evaluation of teaching ability on a 5-point Likert scale (1-Unsatisfactory, 2- Below Expectation, 3- At Expectation, 4- Above Expectation, 5- Outstanding)

Overall

- Please rate the attending's overall teaching ability

Figure 1 Assessment form of attending general internal medicine physicians in the inpatient and outpatient setting. Words or skills previously associated positively with men/classic agentic characteristics are in bold; words or skills previously associated positively with women/classic communal characteristics are underlined and italicized.

and PROF contain a mix of both agentic and communal key words or skills.

The results of trainee evaluations of faculty were compiled from rotations on inpatient general medicine, resident-based hospital medicine, and continuity clinic. Evaluations of subspecialist faculty were excluded. Faculty characteristics including gender and year of medical school graduation were added from a departmental database. The variables of interest

included the gender of the trainee who completed the evaluation (evaluator) and of the faculty (evaluatee) and de-identified prior to analysis.

Primary Outcome

Our primary outcome of interest was the mean rating of the faculty's overall teaching ability and the mean rating of each

ACGME competency based on our assessment form. Mean ratings were then compared by faculty gender.

Data Analysis

Descriptive statistics including means with standard deviations were reported for men and women for overall teaching ability and each ACGME competency. Differences in evaluation scores between male and female faculty were assessed using a multilevel model with resident evaluator identity as a random factor and attending gender (male or female) as a fixed factor. The inclusion of the random factor nullified any effect of an individual evaluators' tendency to give low or high ratings ("hawk" vs. "dove" bias), which is a possible confounder given the unbalanced design of these observational data. By nullifying this effect, the estimated means of male and female scores were free of possible confounding and provided a stronger test of gender effects. The effects of gender on overall rating and each competency rating were evaluated in separate models. Evaluations with missing data were excluded casewise from the multilevel analysis.

We evaluated the interaction of setting (inpatient vs. outpatient) on the effect of faculty gender on assessment scores using a multilevel model with trainee evaluator identity as a random factor and a full factorial of attending gender (male or female) and clinical site type (inpatient vs. outpatient). Again, the inclusion of trainee identity nullified rater bias. A significant interaction of gender and site type would indicate a difference in effect of gender based on clinical site. Least-squares means (classical Yates contrasts) were used as post hoc tests to compare the gender effect in each site type. The interaction effects on overall rating and for each competency rating were assessed in separate models.

All analyses were conducted using R version 3.6.3. Multi-level models were conducted using the "lmerTest" (version 3.1-2) addition to the "lme4" (version 1.1-23) package. Post hoc Yates contrasts were performed using the "ls means" command in "lmerTest."

The study was determined to be exempt by the University of Michigan Institutional Review Board (HUM00160043).

RESULTS

In total, 4081 faculty teaching evaluations from inpatient and outpatient general medicine services were completed by trainees. Five hundred assessments were of subspecialty faculty and excluded. One hundred thirty (3.6%) evaluations were missing data for at least one evaluation measure and excluded casewise from the multilevel analysis.

Of the final 3581 evaluations included, 2046 evaluations were of male faculty (57.1%) and 1535 (42.9%) evaluations were of female faculty (Fig. 2). Among these, 445 total trainees (245 male, 55.1%, and 200 female, 44.9%) assessed 161 distinct attending GIM physicians (81 male, 50.3% and 80 female, 49.7%) with 2365 unique rater-attending pairs. The majority of pairs involved a single assessment of an attending physician by a single resident ($N=1861$, 78.7%). In a minority of cases ($N=302$, 12.8%), a resident assessed the same attending three or more times, mostly in the outpatient setting ($N=298$, 98.9%).

Among all faculty included in our analysis, 83% were on the clinician educator track (85% of total female faculty, 81% of total male faculty) and 17% were on the clinician investigator (i.e., tenure) track. Among the inpatient faculty included in our analysis, 90% were clinician educators, while 75% of

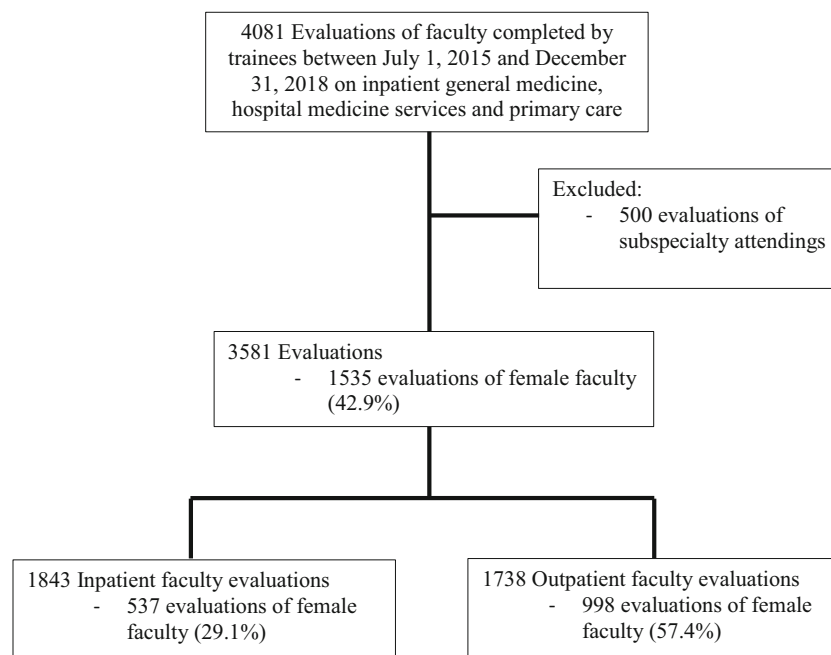


Figure 2 Evaluations of general internal medicine faculty.

the outpatient faculty were clinician educators. A total of seven faculty in our cohort attended in both the inpatient and outpatient setting (five men and two women). Male faculty on average were 20.2 years post-medical school graduation, while female faculty were on average of 15.7 years post-graduation. The number of years since training for individual male attendings ranged from a mean of 14.7 at their earliest evaluation during the analyzed period to 17.3 at their last evaluation compared to individual female attendings who ranged from a mean of 12.8 to 14.8 years since training.

Teaching Assessments by Gender and ACGME Competency

Faculty of both genders were rated by trainees as having excellent clinical performance and teaching ability (Fig. 3). After controlling for rater gender, male faculty were rated as having higher overall teaching ability compared to their female colleagues (male=4.69 vs. female=4.63, $p=0.003$). Male and female faculty were rated similarly in PC (male=4.67 vs. female=4.67, $p=0.94$) and ICS (male=4.72 vs. female=4.72, $p=0.79$). In contrast, male faculty received higher scores than female faculty in the competencies of MK (male = 4.73 vs. female = 4.67, $p<0.001$), PROF (4.79 vs. 4.76, $p=0.02$), PBLI (4.76 vs. 4.73, $p = 0.04$), and SBP (4.75 vs. 4.71, $p = 0.01$).

Impact of Clinical Setting on Gender Differences in Assessment

A total of 1843 evaluations were from inpatient experiences (70.9% male and 29.1% female) and 1738 evaluations were from outpatient experiences (42.6% male and 57.4% female). For all competencies, there was a significant interaction of attending gender and clinical setting (inpatient vs. outpatient) (Fig. 4). This was predominantly due to a larger gender difference in the inpatient setting where male faculty received

higher teaching ratings than female faculty overall and in each of the six competencies (Fig. 4). By contrast, there was no difference in the overall rating of male and female faculty in the outpatient setting or in the competencies of MK, PROF, PBLI, SBP, or ICS. Female faculty in the outpatient setting were rated higher than male faculty in the competency of PC (Fig. 4).

DISCUSSION

This study adds to the growing literature addressing gender disparities in academic medicine. In our cohort, female GIM faculty received lower overall teaching scores than their male counterparts. This difference was largely attributable to evaluations from the inpatient setting. Teaching evaluations for clinician educators play an important role in promotion and compensation; thus, differences in evaluation may be contributing to the “leaky pipeline” of academic medicine. While the absolute differences are small, they should not be discredited as “unimportant.” Prior work on the phenomenon of *amplification cascade* (small differences in evaluations leading to large differences in overall assessment) and *bias accumulation* (multiple subtle biases adding up to overt discrimination) support the theory that gender disparities in GIM are a culmination of countless “small” differences like the ones found here.^{40,41}

When evaluating assessments of GIM faculty independent of setting, male faculty scored higher in overall teaching and in four ACMGE competencies (MK, PROF, PBLI, SBP). While we had hypothesized men would score higher in the competencies with evaluation prompts that included more traditionally agentic language, traits, or skills (MK and PBLI), men were rated higher both in these competencies and in those using both agentic and communal evaluation prompts (PROF

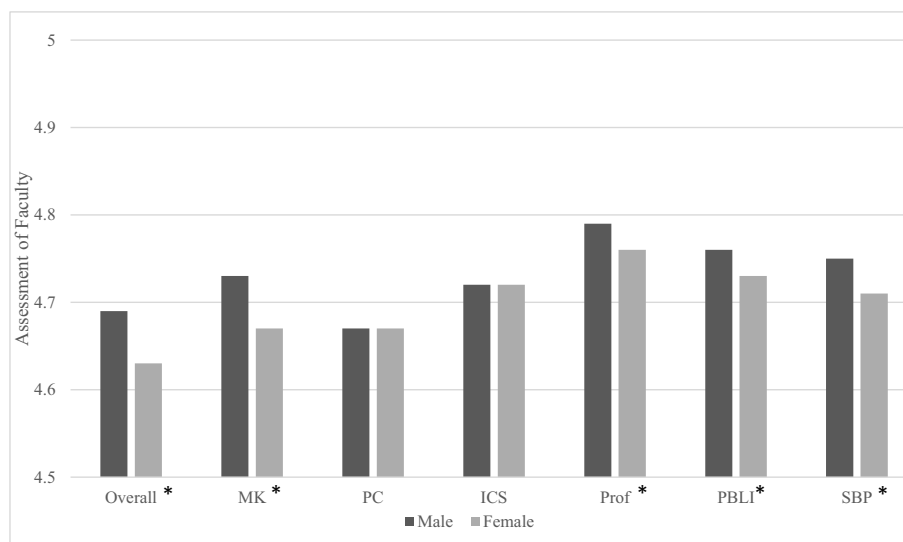


Figure 3 Mean male and female general internal medicine faculty evaluations by trainees using combined inpatient and outpatient settings. MK, medical knowledge; PC, patient care; ICS, interpersonal and communication skills; PROF, professionalism; PBLI, practice-based learning and improvement; SBP, systems-based practice. * $p<0.05$.

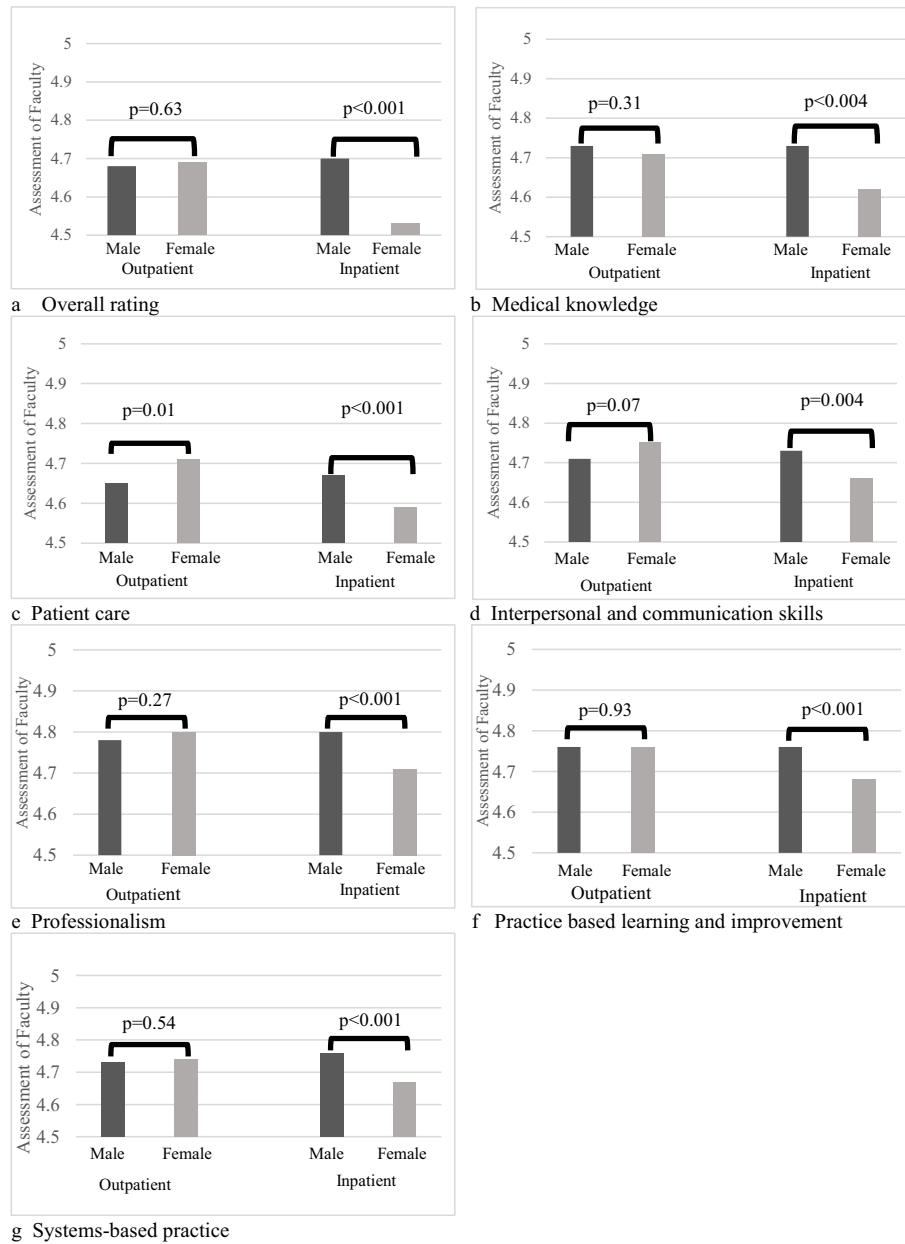


Figure 4 Mean male and female internal medicine faculty evaluations by trainees by clinical setting (a overall rating: outpatient male vs. female faculty (4.68 vs. 4.69) and inpatient male vs. female faculty (4.70 vs. 4.53); b medical knowledge: outpatient male vs. female faculty (4.73 vs. 4.71) and inpatient male vs. female faculty (4.73 vs. 4.62); c patient care: outpatient male vs. female faculty (4.65 vs. 4.71) and inpatient male vs. female faculty (4.67 vs. 4.59); d communication: outpatient male vs. female faculty (4.71 vs. 4.75) and inpatient male vs. female faculty (4.73 vs. 4.66); e professionalism: outpatient male vs. female faculty (4.78 vs. 4.80) and inpatient male vs. female faculty (4.80 vs. 4.71); f practice-based learning and improvement: outpatient male vs. female faculty (4.76 vs. 4.76) and inpatient male vs. female faculty (4.76 vs. 4.68); g systems-based practice: outpatient male vs. female faculty (4.73 vs. 4.74) and inpatient male vs. female faculty (4.76 vs. 4.67)).

and SBP). Conversely, the PC and ICS prompts contained language, traits, or skills considered more communal; while we hypothesized this would result in higher ratings of female faculty, male and female faculty scored no differently in these competencies (Fig. 3). These findings are important — while prior work has shown that gender-biased language is common in narrative comments on evaluations in academic medicine,^{34,35,37,38,42,43} our findings also suggest a potential impact of gendered language within the assessment tools themselves.

While we strive for all faculty to demonstrate each of the attributes described on the assessment form, it is important to understand the impact context-based and gender-based behavior expectations may have on how faculty are evaluated by trainees. In our cohort, men were rated higher than their female peers in overall teaching and across all competency groupings in the inpatient setting, a clinical environment which has been historically male dominated and where traditionally agentic characteristics are more highly valued. Conversely, female GIM faculty in the outpatient setting received higher ratings

in PC compared to male faculty, with no difference in ratings for overall teaching, MK, ICS, PROF, PBLI, and SBP. We hypothesize that the gender disparity in assessment of faculty performance in the inpatient setting may represent the discordance between the expected gender norms for female physicians and the clinical requirements of the inpatient setting, where decisiveness, assertiveness, and urgency of clinical situations may contradict the communal gender-based expectations others have for female faculty. Prior literature has described how female hospitalists leading inpatient teaching services intentionally work to navigate the “too nice” versus “too aggressive” discord between societal gender-based behavior expectations and the more “masculine” expectations of the inpatient clinical setting.⁴⁴

Female faculty may pay a “gender-tax” when being evaluated by trainees in the inpatient setting due to this discordance between expected and observed behaviors. The congruence between expected gender norms and the emphasis on communal traits, such as collaboration, interpersonal sensitivity, and communication, in the outpatient clinical setting is a potential explanation for the equivalent or higher evaluation scores for female GIM faculty in this clinical setting. Like female GIM inpatient faculty, male GIM faculty in the outpatient setting may also pay a “gender-tax” in performance assessment, where traditionally male gender norms could be incongruent with the expectations for the clinical setting.

Overall, these findings highlight the complex interplay between gender norms and the potential impact of clinical setting on teaching evaluations for female faculty. Previously, others have hypothesized that gender disparities may not be as prevalent in hospital medicine due to the near equal number of men and women faculty practicing as academic hospitalists.^{10,11,45} However, the amount of time an individual academic hospitalist spends clinically on teaching vs. non-teaching services varies.⁴⁶ In our study, it is notable only 29% of the total evaluations completed by trainees in the inpatient setting were of women faculty, despite the fact that women made up 44% of the total GIM faculty practicing as hospitalists at our institution, highlighting potential differences in the gender makeup among faculty on teaching vs. non-teaching services. An important next step will be to examine representation of female faculty in inpatient internal medicine teaching roles as this may contribute to trainees’ expectations for behaviors and result in gender differences in teaching evaluations, similar to what has been seen in other male-dominated fields in medicine.²⁰

The fact that the male faculty in our cohort were further out from medical school graduation raises the question of whether there is a confounding disparity in seniority and experience that explains the differences in performance ratings between male and female faculty. While many hypothesize that more senior faculty would be assessed more favorably by learners, this theory is not supported in the literature.^{21,47} We were not able to examine the impact of seniority in our data set due to a

lack female faculty at the most senior level to compare to their male colleagues.

Our study has several limitations. First, this represents the experience of a single institution, using an internally developed evaluation form, and only includes faculty from GIM; our findings may not be representative of all GIM programs or other specialties. Additionally, we recognize that the relationship between faculty and learners is not identical in inpatient and outpatient environment and that this may impact resident assessment of faculty performance. While uncommon for a resident to evaluate the same faculty more than once in the inpatient setting, it does occur in the outpatient setting due to the longitudinal nature of resident continuity clinic. This longitudinal relationship may impact how a resident perceives their attending, independent of the attending’s gender.

In conclusion, our findings suggest that female GIM faculty in academic medicine may be evaluated less favorably by trainees compared to their male colleagues. Our work suggests that gender disparity in evaluations may be heightened based on the language used in the evaluation tools and influenced by the clinical setting of the evaluation. Implicit bias, stereotype-threat, and role incongruity all likely play a role in these observed disparities. Because of the potential impact of teaching evaluations on faculty promotion, advancement, and salary, recognition of gender-based biases in teaching and performance evaluations is essential, especially for female faculty in divisions of hospital medicine and female faculty in other inpatient-focused specialties.

Corresponding Author: Jennifer Reilly Lukela, MD; Department of Internal Medicine, University of Michigan, Ann Arbor, MI, USA (e-mail: jrlreilly@med.umich.edu).

Declarations:

Conflict of Interest: The authors declare that they do not have a conflict of interest.

REFERENCES

1. Jena AB, Olenski AR, Blumenthal DM. Sex Differences in Physician Salary in US Public Medical Schools. *JAMA Internal Medicine*. 2016;176(9):1294. doi:<https://doi.org/10.1001/jamainternmed.2016.3284>
2. aa-data-reports-state-of-women-full-time-faculty-gender-2009-2018_0.jpg (1584×1224). Accessed September 20, 2020. https://www.aamc.org/sites/default/files/aa-data-reports-state-of-women-full-time-faculty-gender-2009-2018_0.jpg
3. Richter KP, Clark L, Wick JA, et al. Women Physicians and Promotion in Academic Medicine. *New England Journal of Medicine*. Published online November 25, 2020. <https://doi.org/10.1056/NEJMsa1916935>
4. Table A-7.2: Applicants, First-Time Applicants, Acceptees, and Matriculants to U.S. Medical Schools by Sex, 2010-2011 through 2019-2020. Published online 10-04 2019. https://www.aamc.org/system/files/2019-10/2019_FACTS_Table_A-7.2.pdf
5. Lautenberger D, Dandar V, Raezer C, Sloane R. The State of Women in Academic Medicine: The Pipeline and Pathways to Leadership. Published online 2014. https://www.hopkinsmedicine.org/women_science_medicine/_pdfs/The%20State%20of%20Women%20in%20Academic%20Medicine%202013-2014%20FINAL.pdf

6. 2019 U.S. Medical School Faculty. AAMC. Accessed August 16, 2020. <https://www.aamc.org/data-reports/faculty-institutions/interactive-data/2019-us-medical-school-faculty>
7. Weeks WB, Wallace AE. Race and Gender Differences in General Internists' Annual Incomes. *J Gen Intern Med*. 2006;21(11):1167-1171. doi:<https://doi.org/10.1111/j.1525-1497.2006.00592.x>
8. Freund KM. Gender Equity in Leadership: SGIM, It's Our Problem! *J GEN INTERN MED*. 2020;35(6):1631-1632. doi:<https://doi.org/10.1007/s11606-020-05710-8>
9. Blazey-Martin D, Carr PL, Terrin N, et al. Lower Rates of Promotion of Generalists in Academic Medicine: a Follow-up to the National Faculty Survey. *J GEN INTERN MED*. 2017;32(7):747-752. doi:<https://doi.org/10.1007/s11606-016-3961-2>
10. Burden M, Frank MG, Keniston A, et al. Gender disparities in leadership and scholarly productivity of academic hospitalists: Gender Disparities for Academic Hospitalists. *J Hosp Med*. 2015;10(8):481-485. doi:<https://doi.org/10.1002/jhm.2340>
11. Herzke C, Bonsall J, Bertram A, Yeh H-C, Apfel A, Cofrancesco J. Gender Issues in Academic Hospital Medicine: a National Survey of Hospitalist Leaders. *J GEN INTERN MED*. 2020;35(6):1641-1646. doi:<https://doi.org/10.1007/s11606-019-05527-0>
12. Reid MB, Misky GJ, Harrison RA, Sharpe B, Auerbach A, Glasheen JJ. Mentorship, Productivity, and Promotion Among Academic Hospitalists. *J GEN INTERN MED*. 2012;27(1):23-27. doi:<https://doi.org/10.1007/s11606-011-1892-5>
13. Bhandari S, Jha P, Cooper C, Slawski B. Gender-Based Discrimination and Sexual Harassment Among Academic Internal Medicine Hospitalists. *J Hosp Med*. 2021;16(2):84-89. doi:<https://doi.org/10.12788/jhm.3533>
14. Atasoylu AA, Wright SM, Beasley BW, et al. Promotion Criteria for Clinician-educators. *J Gen Intern Med*. 2003;18(9):711-716. doi:<https://doi.org/10.1046/j.1525-1497.2003.10425.x>
15. Mitchell KMW, Martin J. Gender Bias in Student Evaluations. APSC. 2018;51(03):648-652. doi:<https://doi.org/10.1017/S104909651800001X>
16. MacNeil L, Driscoll A, Hunt AN. What's in a Name: Exposing Gender Bias in Student Ratings of Teaching. *Innovative Higher Education*. 2015;40(4):291-303. doi:<https://doi.org/10.1007/s10755-014-9313-4>
17. Sprague J, Massoni K. Student Evaluations and Gendered Expectations: What We Can't Count Can Hurt Us. *Sex Roles*. 2005;53(11-12):779-793. doi:<https://doi.org/10.1007/s11199-005-8292-4>
18. Smith DG, Rosenstein JE, Nikolov MC, Chaney DA. The Power of Language: Gender, Status, and Agency in Performance Evaluations. *Sex Roles*. 2019;80(3-4):159-171. doi:<https://doi.org/10.1007/s11199-018-0923-7>
19. McOwen KS, Bellini LM, Guerra CE, Shea JA. Evaluation of Clinical Faculty: Gender and Minority Implications: *Academic Medicine*. 2007;82(Suppl):S94-S96. doi:<https://doi.org/10.1097/ACM.0b013e3181405a10>
20. Fassiotto M, Li J, Maldonado Y, Kothary N. Female Surgeons as Counter Stereotype: The Impact of Gender Perceptions on Trainee Evaluations of Physician Faculty. *Journal of Surgical Education*. 2018;75(5):1140-1148. doi:<https://doi.org/10.1016/j.jsurg.2018.01.011>
21. Thackeray EW, Halvorsen AJ, Ficalora RD, Engstler GJ, McDonald FS, Oxentenko AS. The Effects of Gender and Age on Evaluation of Trainees and Faculty in Gastroenterology. *Official journal of the American College of Gastroenterology | ACG*. 2012;107(11):1610-1614. doi:<https://doi.org/10.1038/ajg.2012.139>
22. Morgan HK, Purkiss JA, Porter AC, et al. Student Evaluation of Faculty Physicians: Gender Differences in Teaching Evaluations. *Journal of Women's Health*. 2016;25(5):453-456. doi:<https://doi.org/10.1089/jwh.2015.5475>
23. Leone-Perkins M, Schnuth R, Kantner T. Preceptor-Student Interactions in an Ambulatory Clerkship: Gender Differences in Student Evaluations of Teaching. *Teaching and Learning in Medicine*. 1999;11(3):164-167. doi:<https://doi.org/10.1207/S15328015TL110307>
24. Carnes M, Bartels CM, Kaatz A, Kolehmainen C. Why is John More Likely to Become Department Chair Than Jennifer? *Trans Am Clin Climatol Assoc*. 2015;126:197-214.
25. Eagly AH, Wood W. Social Role Theory of Sex Differences. In: *The Wiley Blackwell Encyclopedia of Gender and Sexuality Studies*. American Cancer Society; 2016:1-3. doi:<https://doi.org/10.1002/9781118663219.wbeggs183>
26. Carnes M, Bland C. Viewpoint: A Challenge to Academic Health Centers and the National Institutes of Health to Prevent Unintended Gender Bias in the Selection of Clinical and Translational Science Award Leaders: *Academic Medicine*. 2007;82(2):202-206. doi:<https://doi.org/10.1097/ACM.0b013e31802d939f>
27. Chadwick AJ, Baruah R. Gender Disparity and Implicit Gender Bias Amongst Doctors in Intensive Care Medicine: a 'Disease' We Need to Recognise and Treat. *J Intensive Care Soc*. 2020;21(1):12-17. doi:<https://doi.org/10.1177/1751143719870469>
28. Bartels C, Goetz S, Ward E, Carnes M. Internal Medicine Residents' Perceived Ability to Direct Patient Care: Impact of Gender and Experience. *J Womens Health (Larchmt)*. 2008;17(10):1615-1621. doi:<https://doi.org/10.1089/jwh.2008.0798>
29. Kolehmainen C, Brennan M, Filut A, Isaac C, Carnes M. Afraid of Being "Witchy With a B": a Qualitative Study of How Gender Influences Residents' Experiences Leading Cardiopulmonary Resuscitation. *Academic Medicine*. 2014;89(9):1276-1281. doi:<https://doi.org/10.1097/ACM.0000000000000372>
30. Linden JA, Breaud AH, Mathews J, et al. The Intersection of Gender and Resuscitation Leadership Experience in Emergency Medicine Residents: a Qualitative Study. *AEM Educ Train*. 2018;2(2):162-168. doi:<https://doi.org/10.1002/aet2.10096>
31. Wachter RM, Goldman L. The Emerging Role of "Hospitalists" in the American Health Care System. *N Engl J Med*. 1996;335(7):514-517. doi:<https://doi.org/10.1056/NEJM199608153350713>
32. Dressler DD, Pistoria MJ, Budnitz TL, McKean SCW, Amin AN. Core Competencies in Hospital Medicine: Development and Methodology. *Journal of Hospital Medicine*. 2006;1(1):48-56. doi:<https://doi.org/10.1002/jhm.6>
33. Rothman AA, Wagner EH. Chronic Illness Management: What Is the Role of Primary Care? *Annals of Internal Medicine*. 2003;138(3):256-261. doi:<https://doi.org/10.7326/0003-4819-138-3-200302040-00034>
34. Mueller AS, Jenkins TM, Osborne M, Dayal A, O'Connor DM, Arora VM. Gender Differences in Attending Physicians' Feedback to Residents: a Qualitative Analysis. *J Grad Med Educ*. 2017;9(5):577-585. doi:<https://doi.org/10.4300/JGME-D-17-00126.1>
35. Arkin N, Lai C, Kiwakyou LM, et al. What's in a Word? Qualitative and Quantitative Analysis of Leadership Language in Anesthesiology Resident Feedback. *J Grad Med Educ*. 2019;11(1):44-52. doi:<https://doi.org/10.4300/JGME-D-18-00377.1>
36. Santen S, Yamazaki K, Holmboe E, Yarris L, Hamstra S. Comparison of Male and Female Resident Milestone Assessments During Emergency Medicine Residency Training: a National Study. *Acad Med*. 2020;95(2):263-268. doi:<https://doi.org/10.1097/ACM.0000000000002988>
37. Axelson RD, Solow CM, Ferguson KJ, Cohen MB. Assessing Implicit Gender Bias in Medical Student Performance Evaluations. *Eval Health Prof*. 2010;33(3):365-385. doi:<https://doi.org/10.1177/0163278710375097>
38. Gerull K, Loe M, Seiler K, McAllister J, Salles A. Assessing Gender Bias in Qualitative Evaluations of Surgical Residents. *Am J Surg*. 2018;217(2):306-313. doi:<https://doi.org/10.1016/j.amjsurg.2018.09.029>
39. Smith DG, Rosenstein JE, Nikolov MC. The Different Words We Use to Describe Male and Female Leaders. *Harvard Business Review*. Published online May 25, 2018. Accessed April 25, 2021. doi:<https://hbr.org/2018/05/the-different-words-we-use-to-describe-male-and-female-leaders>
40. Teherani A, Hauer KE, Fernandez A, King TE, Lucey C. How Small Differences in Assessed Clinical Performance Amplify to Large Differences in Grades and Awards: a Cascade With Serious Consequences for Students Underrepresented in Medicine. *Academic Medicine*. 2018;93(9):1286-1292. doi:<https://doi.org/10.1097/ACM.0000000000002323>
41. Page S. *The Diversity Bonus*. Princeton: Princeton University Press; 2017.
42. Madera JM, Hebl MR, Martin RC. Gender and Letters of Recommendation for Academia: Agentive and Communal Differences. *The Journal of Applied Psychology*. 2009;94(6):1591-1599. doi:<https://doi.org/10.1037/a0016539>
43. Rojek AE, Khanna R, Yim JWL, et al. Differences in Narrative Language in Evaluations of Medical Students by Gender and Under-represented Minority Status. *J GEN INTERN MED*. 2019;34(5):684-691. doi:<https://doi.org/10.1007/s11606-019-04889-9>

44. **Houchens N, Quinn M, Harrod M, Cronin DT, Hartley S, Saint S.** Strategies of Female Teaching Attending Physicians to Navigate Gender-Based Challenges: an Exploratory Qualitative Study. *J Hosp Med.* 2020;15(8):454-460. <https://doi.org/10.12788/jhm.3471>
45. The Society of General Internal Medicine Membership Committee, Miller CS, Fogerty RL, Gann J, Bruti CP, Klein R. The Growth of Hospitalists and the Future of the Society of General Internal Medicine: Results from the 2014 Membership Survey. *J GEN INTERN MED.* 2017;32(11):1179-1185. doi:<https://doi.org/10.1007/s11606-017-4126-7>
46. **Glasheen J, Misky G, Reid M, Harrison R, Sharpe B, Auerbach A.** Career satisfaction and burnout in academic hospital medicine. *ARCH INTERN MED.* 2011;171(8):4.
47. **Arah OA, Heineman MJ, Lombarts KMJM.** Factors Influencing Residents' Evaluations of Clinical Faculty Member Teaching Qualities and Role Model Status. *Medical Education.* 2012;46(4):381-389. doi:<https://doi.org/10.1111/j.1365-2923.2011.04176.x>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.