Abstract

IMPORTANCE Prior studies have revealed gender differences in the milestone and clinical competency committee assessment of emergency medicine (EM) residents.

OBJECTIVE To explore gender disparities and the reasons for such disparities in the narrative comments from EM attending physicians to EM residents.

DESIGN, SETTING, AND PARTICIPANTS This multicenter qualitative analysis examined 10,488 narrative comments among EM faculty and EM residents between 2015 to 2018 in 5 EM training programs in the US. Data were analyzed from 2019 to 2021.

MAIN OUTCOMES AND MEASURES Differences in narrative comments by gender and study site. Qualitative analysis included deidentification and iterative coding of the data set using an axial coding approach, with double coding of 20% of the comments at random to assess intercoder reliability (κ, 0.84). The authors reviewed the unmasked coded data set to identify emerging themes. Summary statistics were calculated for the number of narrative comments and their coded themes by gender and study site. χ² tests were used to determine differences in the proportion of narrative comments by gender of faculty and resident.

RESULTS In this study of 283 EM residents, of whom 113 (40%) identified as women, and 277 EM attending physicians, of whom 95 (34%) identified as women, there were notable gender differences in the content of the narrative comments from faculty to residents. Men faculty, compared with women faculty, were more likely to provide either nonspecific comments (115 of 182 [63.2%] vs 40 of 95 [42.1%]), or no comments (338 of 10,496 [32.3%] vs 1,169 of 4,548 [25.7%]; P < .001) to men and women residents. Compared with men residents, more women residents were told that they were performing below level by men and women faculty (36 of 113 [31.9%] vs 43 of 170 [25.3%]), with the most common theme including lack of confidence with procedural skills.

CONCLUSIONS AND RELEVANCE In this qualitative study of narrative comments provided by EM attending physicians to residents, multiple modifiable contributors to gender disparities in assessment were identified, including the presence, content, and specificity of comments. Among women residents, procedural competency was associated with being conflated with procedural confidence. These findings can inform interventions to improve parity in assessment across graduate medical education.
Introduction

Feedback that is timely, specific, and actionable is key in creating an effective learning environment for resident physicians. Effective feedback improves learner motivation, engagement, and perceived support, lower rates of burnout, and clinical outcomes. However, delivering effective and unbiased feedback remains challenging with studies indicating that it is difficult to divorce gender and gendered expectations during the feedback process.

Recent studies reveal gender disparities within the assessment and feedback process across specialties, including family medicine, general surgery, internal medicine, and particularly notable within emergency medicine (EM) training programs. In 2013, a study across 8 EM training programs of competency-based assessments revealed that the rate of milestone attainment throughout training was higher for men as compared with women residents across all 23 EM sub-competencies. A follow-up single-site study of written feedback to postgraduate year (PGY) 3 residents revealed that, women compared with men, received inconsistent feedback when struggling. A national data set of EM resident milestone assessments by clinical competency committees (CCC) revealed that there were gender disparities in 4 of the 22 subcompetencies (ie, emergency stabilization, general approach to procedures, airway management, and vascular access) at the time of graduation. Gender bias has shown to affect women's career selection, advancement, and leadership opportunities; within the field of emergency medicine, gender parity has not yet been achieved, with women representing only 35.9% of EM residents, 28.3% of EM faculty, and 11% of EM department chairs.

While these prior studies revealed gender bias in assessment and feedback within EM training programs, specifically revealing that men achieve competency ratings faster than women, none examined what factors contribute to this disparity. Understanding modifiable contributors to gender inequities in the feedback and assessment process is integral to improving the learning environment for all trainees. Within the conceptual framework of sociocultural learning, the educational paradigm that suggests learning happens in and is informed by its context, gender inequity and potential biases in the clinical and nonclinical learning environment could affect how trainees learn and perform. Furthermore, addressing potential inequities in the assessment process within the learning environment would directly improve how men and women trainees learn. This is consistent with national policy guidelines from both the Accreditation Council of Graduate Medical Education (ACGME) and the Association of American Medical Colleges (AAMC), prioritizing the creation of a learning environment that acknowledges and addresses biases. In this qualitative study, we sought to explore gender disparities, the reasons for such disparities, and the association of faculty gender with the narrative comments provided to residents, derived from an online assessment system, from 5 training programs over 3 years to inform future interventions to mitigate this disparity.

Methods

The institutional review board at each of the 5 study sites (University of Pennsylvania, University of Washington, Oregon Health and Sciences University, University of Iowa, and University of Alabama) deemed the study exempt, and informed consent was waived because participant information was deidentified. We followed the Standards for Reporting Qualitative Research (SRQR) reporting guideline and the Sex and Gender Equity Research Guidelines.

Study Objectives

We conducted a qualitative analysis of narrative comments on EM resident assessment forms by EM attending physicians using an online assessment system (Medhub) during a 3-year period (July 1, 2015 to June 30, 2018) across 5 EM training programs: University of Pennsylvania, University of Washington, Oregon Health and Sciences University, University of Iowa, and University of Alabama.
These narrative assessment comments were used as the primary source of written feedback to EM trainees during the study period. The study sites did not mandate implicit bias training for the attending physicians during the study period.

Our study examined gender differences in the (1) total number of narrative comments by study site, and study year; (2) thematic content of the narrative comments; (3) specificity of feedback; and (4) faculty perception of resident performance as compared with their peers. Because prior studies have incompletely characterized why EM women residents lag behind their male colleagues in attaining EM milestones, we examined a subset of narrative comments among residents who were assessed as performing below level to understand differences by gender.

**Study Participants and Setting**

Our study population included all EM faculty and trainees at the study sites. Gender of the residents and attending physicians were self-identified and provided by the site investigators to the research team. No participants self-identified as nonbinary during the study period. Study sites were chosen to ensure a geographic diversity and adequate site participation within the online evaluation system. Study sites participated within the years 2015 to 2018.

**Data Analysis**

**Qualitative**

The qualitative analysis for this study was performed from 2019 to 2021 by the Mixed Methods Research Lab (MMRL) at University of Pennsylvania. The MMRL is composed of a team of experts in qualitative and mixed methods research methods who use their specialized knowledge and skills to analyze and manage qualitative data sets. We extracted assessment data inclusive of 283 EM residents during the study period from Medhub at the 5 study sites. We categorized assessment data as no comment when only competency scoring was present and with comment if both competency scoring and narrative comments were present. After linking qualitative assessment data and resident gender data, the entire data set was deidentified for names and pronouns that may indicate gender and thereafter imported into NVivo™ for coding and analysis.

**Coding**

The study authors provided the MMRL with the codebook by Mueller et al as a guide prior to coding. The MMRL used a multistage, multianalyst iterative process that included (1) refining the code book through an iterative review of the data set (eAppendix in the Supplement); (2) coding the de-identified data using an axial coding approach; and (3) double coding by separate coder of 20% of the data set at random to assess intercoder reliability. In lieu of the more traditional qualitative study approach of analyzing data until thematic saturation has been achieved, we coded the entire data set to better characterize nuances in narrative comments to men and women residents. We also had co-codes related to (1) faculty assessment of resident performance as compared with their peers designated as at level, above level, or below level and (2) specificity of narrative comments, designated as nonspecific vs specific when there was feedback on a specific case or behavior. Notably, both men and women coders, who were not part of the study team, were employed by the MMRL to provide gender balance and avoid confirmatory bias in the coding process. The coders were trained by the MMRL prior to initiation of coding. Because the coders were nonphysicians, EM attending physicians reviewed the coded data set with the coders to provide clinical context and ensure trustworthiness. All coding discrepancies were resolved by group consensus. Intercoder reliability using the k coefficient revealed acceptable agreement among coders (k, 0.84).

**Thematic Analysis**

Data were unmasked to gender after coding. We used a postpositivist deductive approach when reviewing the unmasked coded data set to identify emerging themes. We examined (1) the gender
differences in the percent of faculty that provided narrative comments within these codes, and (2) the gender differences in the percent of residents that received narrative comments within these codes. We highlighted gender differences that were greater than or equal to 5%, to delve into the areas with the greatest differences and reported these differences in summary form with identified emerging themes.

Descriptive Analysis of Qualitative Themes

Our qualitative data comprised 39,776 codes for all narrative comments during the study period. We calculated summary statistics (i.e., frequencies and percentages) of the number of narrative comments and their themes by gender, study site, and overall. We used \( \chi^2 \) tests to identify differences in the proportion of narrative comments by gender in 4 separate dyads (i.e., men faculty to men residents, men faculty to women residents, women faculty to men residents, and women faculty to women residents). Descriptive analyses were conducted using SAS statistical software version 9.4 (SAS Institute).

Results

Gender Differences in Providing Narrative Comments

The study population included 283 EM residents, of whom 170 (60%) identified as men and 113 (40%) as women, and 277 EM attending physicians, of whom 182 (66%) identified as men and 95 (34%) women. Of the total 15,044 instances of assessment during the study period, 10,488 (69.7%) included both competency scorings and narrative comments (with comments) and 4,556 (30.3%) included only competency scorings (no comments). Of the data with no comments, 3,387 of 10,496 (32.3%) came from men faculty compared with 1,169 of 4,548 (25.7%) of women faculty \( (P < .001) \). Overall, women faculty provided more narrative comments to women residents compared with men residents (1,814 of 2,576 [70.4%] vs 1,565 of 1,972 [70.4%]; \( P < .001 \)). Similarly, men faculty provided more comments to women vs men residents (2,631 of 3,713 [70.9%] vs 4,478 of 6,783 [66.0%]; \( P < .001 \)).

The gender distribution of residents and faculty as well as the number of narrative comments varied across study sites (Figure 1 and Figure 2). Site 1 had the fewest narrative comments (men and women faculty provided a median [IQR] of 5 [1-13] and 5 [2-10], respectively), while site 3 had the most narrative comments (men and women faculty provided a median [IQR] of 50 [13-97] and 57 [31-124], respectively). A greater proportion of attending physicians were men at each of the study sites. Within sites, we found no differences in the number of comments between men and women.
faculty or residents, except for site 2 which had more comments directed to men residents vs women residents (median [IQR], 50 [19-82] vs 17 [6-34]).

**Differences in the Content of Narrative Comments Provided by Men vs Women Faculty**

Among the 10,488 instances in which narrative comment were provided, we found differences in coded feedback themes by faculty gender. Men faculty compared with women faculty provided more comments about patient load (100 of 182 [54.9%] vs 45 of 95 [47.4%]) and clinical skills (115 of 182 [63.2%] vs 53 of 95 [55.8%]). In contrast, women faculty compared with men faculty provided more comments about team orientation, teaching skills, patient advocacy, attentiveness, professionalism, critical thinking, bedside manner, organization and level of compassion, ranging from 60 of 95 (63.2%) vs 105 of 182 (57.7%) to 44 of 95 (46.3%) and 64 of 182 (35.2%). Figure 3 illustrates the proportion of faculty who provided feedback in specific content areas, highlighting only the instances in which there was greater than 5% difference between men and women faculty.

**Figure 2. Distribution of Narrative Comments Provided by Emergency Medicine Faculty by Gender and Site**

![Figure 2](image)

Boxes indicate the IQR; dots, outliers; line within the box, median; error bars, 95% CI.

**Figure 3. Comparisons of Feedback Themes Provided by Faculty Gender**

![Figure 3](image)
Differences in the Content of Narrative Comments Received by Men vs Women Residents

More men trainees, compared with women trainees, received narrative comments about documentation, differential diagnosis, care planning, patient and clinician communication, critical thinking, bedside manner, trustworthiness, and professionalism, with differences ranging from 140 of 170 (82.4%) vs 86 of 113 (76.1%) to 131 of 170 (77.1%) vs 73 of 113 (64.6%). An example of a comment coded as communication skill included “they treat all patients with unconditional positive regard...and [brings] advanced communication skills to the team.” In contrast, more women residents as compared with men residents received narrative comments about confidence, assertiveness with treatment, and adaptability, with differences ranging from 31 of 113 (27.4%) vs 37 of 170 (21.8%) to 31 of 113 (27.4%) vs 33 of 170 (19.4%). An example of a comment coded as adaptability included: “resident is reliable but at times more independent than I prefer, and it makes me not always trust them—I haven’t seen them do harmful things, but I’ve had conversations with them where they dig their heels in, nearly refuses to act on my suggestion and takes a righteous stand.” Figure 4 illustrates the proportion of residents who received feedback in specific content areas, highlighting only the instances in which there is a greater than 5% difference between men and women residents.

Differences in the Specificity of Narrative Comments Provided Between Men and Women Faculty

There were also differences in the specificity of comments provided by faculty gender. Men faculty provided feedback that was nonspecific more often than women faculty (115 of 182 [63.2%] vs 40 of 95 [42.1%]) (Figure 3). Examples of these nonspecific comments included: “Great work” and “Doing well, keep up the good work.” These nonspecific comments were directed equally to men and women residents and were consistent across study sites.
Differences in the Specificity of Narrative Comments Received Between Men and Women Residents

Men and women residents also differed in the specificity of feedback received. Men residents received feedback on specific cases or behaviors more often than women residents (103 of 170 [60.6%] vs 57 of 113 [50.4%]) (Figure 4). Examples of men receiving feedback on specific cases included “Excellent management of patient with SVT unresponsive to conventional care. Good choice of meds and other management” and “Had a scare where they looked at the wrong dated CT on a patient which was negative when the most recent CT had a head bleed…we discussed the importance of right patient, right date, right time.” These patterns were consistent irrespective of faculty gender and across most of our study sites.

Differences in Men vs Women Faculty Perception of Resident Performance

Among the 10,088 instances in which narrative comments were provided, 2,556 instances (25.3%) included faculty commenting on resident performance. In this subset, more men faculty compared with women faculty assessed that the resident was performing above level (124 of 182 [68.1%] vs 54 of 95 [56.8%]) or at level (109 of 182 [59.9%] vs 48 of 95 [50.5%]) compared with their peers (Figure 3). In contrast, more women faculty compared with men faculty assessed that the resident was performing below level compared with their peers (28 of 95 [29.5%] vs 41 of 182 [22.5%]) (Figure 3).

Differences in Men vs Women Resident Performance Feedback

More men residents were assessed as performing above level (compared with their peers 145 of 170 [85.3%] vs 90 of 113 [79.6%]), while more women residents were assessed as performing below level compared with their peers (36 of 113 [31.9%] vs 43 of 170 [25.3%]) (Figure 4). These patterns were consistent irrespective of faculty gender, and across most of our study sites.

Below Level Comments to EM Trainees

Among the subset of residents with faculty assessment of below level performance, we observed that more than 5% of women trainees received comments that their procedural skills were below level compared with men (7 of 113 [6.19%] vs 2 of 170 [1.18%]). Furthermore, we identified that below-level procedural skill comments directed to women trainees were frequently co-coded with comments about confidence (or lack thereof). Examples included “Resident lacks some confidence also when it comes to procedures—given that they are at the end of their [postgrad year 2], they need to find more of them to build up their confidence” and “Resident appears to have solid clinical skills but has gotten flustered when procedures are required.”

In contrast, below-level procedural skill comments directed to men residents included specific actionable items without mention of their confidence with procedures. Examples included “But the last shift we worked together you missed a few things including...getting an opening pressure on a transplant patient in whom you were performing a lumbar puncture” and “Advised using direct laryngoscopy for very bloody airway.”

Discussion

To our knowledge, this is the first multisite study to examine gender differences in narrative comments to EM residents. We analyzed 10,488 narrative comments submitted from 2015 to 2018 across 5 training programs and found gender differences in multiple domains, including whether narrative comments were provided to trainees, the content and specificity of the comments to trainees, and the perceptions of trainee performance. Among faculty, more men compared with women provided either no narrative comment feedback or provided nonspecific comments. Among trainees, women were more likely to receive narrative comments from both men and women faculty;
however, those comments were less specific, and they were assessed as performing below level more often than men residents.

Prior studies have revealed that men achieve competency ratings faster than women residents. This study builds upon prior work to reveal what modifiable factors may be contributing to this disparity. Our qualitative analysis identified differences in perceived competency related to procedural skills by gender, with women trainees often noted to lack confidence with procedures, as a leading driver of the below level narrative feedback received by women trainees. This finding suggests that, among women residents, attending physician perception of resident confidence may be conflated with ratings of procedural competency. These findings, while identified in EM training programs, are translatable across graduate medical education.

Achieving mastery of procedural skills is an important component of clinical competency. Residents have a minimum number of procedures they need to complete in order to progress to independent practice. Prior studies have revealed no gender differences in procedural outcomes among EM physicians including first pass-intubation success and accuracy of cricothyroid membrane identification. However, both the Daya et al and Santen et al studies revealed that among EM trainees, women residents lag behind men in achieving procedural subcompetency scores. Our study findings suggest that among women residents who are assessed as performing below level in procedural skills, competency with procedure may be conflated with confidence in executing the procedure. It is unclear whether women express less confidence in procedures or whether faculty perception of trainee confidence in procedures is leading to poorer ratings of procedural skills or overinflation of ratings among men. However, none of the comments included discussion of procedural outcomes.

Etiologies for this disparity are likely multifactorial but may be at least partially explained by gender and leadership norms in medicine. Historically, emergency medicine has roots in surgery and in the army, fields that have until recently been composed primarily of men with agentic or action-oriented and assertive leadership styles being lauded without accounting for the strengths of leadership styles that are not in line with this norm. Not surprisingly, among surgical training programs a similar phenomenon exists with women residents’ self-reporting lower autonomy in the operating room. In this context, it is possible that our collective definitions of clinical skills or procedural skills are so entrenched in a gendered model of competency that we do not have the appropriate tools to determine resident abilities that do not fit this model. One means to break out of this outdated mold would be to assess residents with more objective parameters. For example, a recent study identified that using checklist based procedural evaluations may mitigate gendered assessments. Alternatively, assessment of procedural competency could include reference to procedural outcomes, or clinical based competency could include discussion of patient outcomes or care patterns. It is notable that while our frameworks for assessment, such as the EM milestones, may not be inherently biased, the lens by which we interpret these models and apply them at the level of the CCC can lead to biased assessment. Given this reality, we may need to reconsider how we use these frameworks, particularly when using high-stakes advancement decisions.

Our study also suggests that the focus and volume of faculty feedback differs by resident gender. Women, compared with men, provide more narrative comments to residents. Providing feedback is integral to the growth and development of learners, and yet not recognized in the tenure and promotion process, consistent with prior literature that women engage in work that is valuable but often not rewarded by academic institutions. In addition, while women residents receive concentrated feedback on personal attributes, such as adaptability, assertiveness, and confidence, men residents receive feedback in multiple domains. This finding is consistent with prior studies across specialties indicating that women receive more comments about their personal attributes and disposition, compared with men who receive feedback in competency-related behaviors. Studies have shown that gender biases are prevalent, exist in both men and women, and that it is notoriously hard to train out biases. However, a recent study by Peterson et al examining student assessments of teachers revealed that identifying that biases exist prior to filling out assessment can
mitigate the disparities seen. Training programs and institutions could consider including an opportunity for individuals engaging in assessment to understand and acknowledge their biases prior to providing narrative assessment comments.

Lastly, our data suggest that among faculty, men may be more likely to provide no comment or nonspecific comments, and among residents, men may be more likely to receive feedback on specific cases. While prior studies have revealed gender differences in the consistency of feedback to EM trainees, to our knowledge, this is first study to identify gender differences in specificity of feedback to EM residents. Nonspecific feedback may represent missed opportunities to assess skills and deficiencies. As women residents are more likely to be assessed as performing below level but less likely to receive feedback on specific cases or behaviors, compared with men, this may limit opportunities for growth and development throughout their training. These disparities can be propagated beyond residency training, given that assessments often inform letters of recommendations for employment or fellowships. As providing specific, actionable feedback is integral to resident growth and development, it is imperative that efforts are made to improve the specificity of feedback provided by EM faculty. Interventions such as real-time point of care nudges while completing narrative comments and faculty training sessions should be explored to address this critical need.

Limitations

Our study had several limitations. While we identified gender differences in the narrative comments provided to EM trainees, demographic factors, such as age, race and ethnicity, sexual orientation, were not available in the data set, so we were unable to assess the multifactorial effects of these intersecting identities. In addition, as we had no participants who self-identified as nonbinary during the study period, we do not account for findings that may be relevant to those who identify on the gender continuum. Furthermore, while we have identified that our major findings are consistent across study sites, there may be additional site-specific variations influenced by local culture that should be explored in future studies. Lastly, as we chose to highlight gender differences that were greater than or equal to 5% in the narrative comments provided to trainees, we might have missed additional emerging themes that were notable in the data set for which the differences were less than 5%.

Conclusions

In this qualitative study of narrative comments provided by EM faculty, we have identified multiple modifiable contributors that are associated with gender disparities in narrative comments to EM residents, including the presence, content, and specificity of the comments as well as identifying that among women residents procedural confidence may be conflated with procedural competency. The results of this study could inform future interventions to mitigate these disparities across graduate medical education.
Author Contributions: Drs Mamtani and Shofer had full access to all of the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.

Concept and design: Mamtani, Scott, Gallahue, Yarris, Abbuhl, Aysola.

Acquisition, analysis, or interpretation of data: Mamtani, Shofer, Scott, Kaminstein, Eriksen, Takacs, Hall, Weiss, Walter, Gallahue, Yarris, Abbuhl.

Drafting of the manuscript: Mamtani, Shofer, Yarris, Aysola.

Critical revision of the manuscript for important intellectual content: All authors.

Statistical analysis: Mamtani, Shofer, Scott, Eriksen, Aysola.

Obtained funding: Mamtani, Scott, Abbuhl.

Administrative, technical, or material support: Mamtani, Scott, Takacs, Gallahue, Yarris, Abbuhl.

Supervision: Mamtani, Hall, Aysola.

Qualitative research expertise: Kaminstein.

Conflict of Interest Disclosures: Dr Hall reported being a contract employee of the Royal College of Physicians and Surgeons of Canada outside the submitted work. Dr Walter reported receiving grants from Substance Abuse and Mental Health Services Administration, Gilead FOCUS, and the Alabama Department of Public Health outside the submitted work. No other disclosures were reported.

Funding/Support: This study was funded by the Council of Residency Directors and Emergency Medicine Foundation Education Research Grant.

Role of the Funder/Sponsor: The funders had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Meeting Presentation: Part of this article was presented as an oral presentation at the American College of Emergency Physicians conference on October 2019.

Additional Contributions: We thank the Penn Medicine program FOCUS on Health and Leadership for Women for their guidance and support.

REFERENCES


SUPPLEMENT.
eAppendix. Codebook With Description