

# Gender Differences in Attending Physicians' Feedback to Residents: A Qualitative Analysis

Anna S. Mueller, MA, PhD  
Tania M. Jenkins, MA, PhD  
Melissa Osborne, MA

Arjun Dayal, MD  
Daniel M. O'Connor, MD  
Vineet M. Arora, MD, MAPP

## ABSTRACT

**Background** Prior research has shown a gender gap in the evaluations of emergency medicine (EM) residents' competency on the Accreditation Council for Graduate Medical Education (ACGME) milestones, yet the practical implications of this are not fully understood.

**Objective** To better understand the gender gap in evaluations, we examined qualitative differences in the feedback that male and female residents received from attending physicians.

**Methods** This study used a longitudinal qualitative content analysis of narrative comments by attending physicians during real-time direct observation milestone evaluations of residents. Comments were collected over 2 years from 1 ACGME-accredited EM training program.

**Results** In total, 1317 direct observation evaluations with comments from 67 faculty members were collected for 47 postgraduate year 3 EM residents. Analysis of the comments revealed that the ideal EM resident possesses many stereotypically masculine traits. Additionally, examination of a subset of the residents (those with 15 or more comments,  $n = 35$ ) showed that when male residents struggled, they received consistent feedback from different attending physicians regarding aspects of their performance that needed work. In contrast, when female residents struggled, they received discordant feedback from different attending physicians, particularly regarding issues of autonomy and assertiveness.

**Conclusions** Our study revealed qualitative differences in the kind of feedback that male and female EM residents received from attending physicians. The findings suggest that attending physicians should endeavor to provide male and female residents with consistent feedback and guard against gender bias in their perceptions of residents' capabilities.

## Introduction

Despite achieving parity in medical school graduations,<sup>1</sup> female physicians face barriers to advancement.<sup>2</sup> They hold fewer faculty positions at academic institutions, earn lower adjusted incomes, and are in fewer positions of leadership in medical societies and departments than their male counterparts.<sup>1-7</sup> A recent systematic review suggested that the greatest attrition in commitment to academia occurs during residency, possibly due to implicit gender bias and lack of support in the workplace, among other contributors.<sup>2</sup>

Few studies have examined the status of women in emergency medicine (EM),<sup>4</sup> a specialty where gender parity has yet to be achieved. Women comprise less than 25% of all faculty positions in emergency departments<sup>8</sup> and represent only 38% of EM residents (up from 32% in 2003).<sup>1</sup> It is particularly interesting to investigate gender inequality in EM, as it was one of the first specialties to evaluate residents' competency using the Accreditation Council for Graduate Medical Education (ACGME) Next Accreditation System milestones, a nationally standardized, longitudinal evaluation system.<sup>9</sup> While

assessment of competency is to have a "beneficial effect on learning,"<sup>9</sup> a broad body of literature in the social sciences notes that status characteristics—like gender—matter when evaluating someone's competency, even on theoretically objective standards, and when meritocracy is valued in organizations.<sup>10-12</sup> Research has also found that evaluators struggle to assess competency or merit independent of gender and gendered expectations.<sup>13-16</sup>

A recent study by our group found an attainment gap between male and female EM residents in evaluations of performance on the nationally standardized milestones. In a longitudinal, multicenter study, we found that male residents, on average, had a higher rate of milestone attainment throughout residency, with a widening gap in evaluations that is substantial and statistically significant by postgraduate year 3 (PGY-3).<sup>17</sup> This attainment gap was not dependent on either the gender of the attending physician doing the evaluation, or the gender pairing between the attending physician and the resident.<sup>17</sup> While our prior work demonstrated a gender difference in numerical evaluations, in this study, our aim was to use qualitative data to better understand the lagging performance evaluations of female EM residents in PGY-3.

DOI: <http://dx.doi.org/10.4300/JGME-D-17-00126.1>

## Methods

We conducted a qualitative content analysis of comments attached to numerical ACGME milestone-based evaluations of PGY-3 EM residents by attending physicians. We used a post-positivist research paradigm, wherein we acknowledge that perfect objectivity is never fully attainable, but rather a goal toward which we strive, by recognizing the influence of our characteristics, backgrounds, and values on knowledge production.<sup>18</sup> Of the authors, 3 (A.S.M., T.M.J., M.O.) have extensive sociological training in qualitative methods and medical sociology, and 3 (A.D., D.M.O., V.M.A.) are clinicians with qualitative methods experience. Our collective expertise allowed us to approach data analysis rigorously and helped guard against potential disciplinary biases or knowledge gaps. The authors had no contact with the residents or faculty and are not affiliated with the hospital under study. This separation helped prevent subjectivity in the analysis; it also made it impossible to interpret the results with reference to the local institutional context or culture.

### Data Collection

Study data were collected from a single 3-year ACGME-accredited EM training program that we call “University Hospital” (a pseudonym) from July 1, 2013, to July 1, 2015. A total of 1317 direct observation evaluations with comments were collected from 2 cohorts of PGY-3 EM residents, and these included evaluations of 47 PGY-3 residents by 67 faculty members.

Text comments were collected using InstantEval V2.0 (Monte Carlo Software LLC, Annandale, VA), a software application available on faculty’s mobile devices and computers to facilitate real-time, direct observation milestone evaluations. Faculty could choose when to complete evaluations, whom to evaluate, and the number of evaluations to complete, although most training programs encouraged 1 to 3 evaluations per shift. Each evaluation consisted of an ACGME Emergency Medicine Milestone Project-based performance level<sup>19</sup> on 1 of 23 possible individual EM subcompetencies. In certain cases, text comments were provided, which are the focus of this study. All names used in the text are pseudonyms to protect confidentiality.

This study was approved as exempt research by the University of Chicago Institutional Review Board.

### Data Analysis

Our post-positivist approach was guided by a sequential explanatory analytic design,<sup>20</sup> which is

#### What was known and gap

Faculty feedback to male and female emergency residents appears to differ, but the educational and practical implications have not been explored.

#### What is new

A longitudinal qualitative content analysis of narrative comments by attending physicians during real-time direct observation milestone evaluations of emergency residents suggested gender bias.

#### Limitations

Single site, single specialty study limits generalizability.

#### Bottom line

There were qualitative differences in the feedback faculty gave to male and female emergency residents, particularly around the domains of authority and assertiveness.

when qualitative methods are used to better understand previously established quantitative findings. This approach aided our efficient analysis of a large qualitative dataset. The previously established quantitative findings that guide our research are (1) that a gender gap exists in resident evaluations on ACGME standards in EM,<sup>17</sup> and (2) that the gap is most substantial in PGY-3.<sup>17</sup> We focused our analysis on PGY-3 and examined qualitative differences in the feedback male and female residents received in order to shed light on why evaluation gaps may exist.

All qualitative data were coded and analyzed in NVivo 11 (QSR International, Burlington, MA), a qualitative analysis software package. To guard against confirmation bias, we suppressed information about residents’ and attending physicians’ gender during all stages of coding. This process was imperfect, as some comments included gendered pronouns or names.

As a further guard against confirmation bias, we developed a multistage, multianalyst procedure for coding and analyzing the data. In the first stage, 5 team members engaged in simultaneous open coding of all comments to develop themes from the data<sup>21</sup> and to ensure accurate understanding of comments. During this stage, we identified characteristics of residents that are valued in EM (TABLE 1), and identified 4 themes (*strong criticism, praise, possesses, and lacks valued personality traits*) in TABLE 2. In the second stage, 3 team members conducted focused coding for these 4 themes. To limit confirmation bias in coding during this stage, 2 of 3 team members coded every comment for our selected themes. Any discrepancies between codes were discussed by all 3 team members, and consensus was reached in all cases. Third, we analyzed gender differences in the comments residents received. For residents who received at least 15 comments ( $n = 35$ ), we did an in-depth analysis of how different attending

**TABLE 1**  
Characteristics of Ideal Emergency Medicine Residents

Personality	Practice Style	Management Style
Team leader/takes charge <sup>a</sup>	Aggressive with treatment	Efficient
Independent/autonomous <sup>a</sup>	Performs under pressure	Manages whole department
Decisive <sup>a</sup>	Curious/receptive to learning	Multitasking
Confident <sup>a</sup>	Documents well	Problem solver
Professional	Takes on big enough patient load	Focused
Open to feedback	Evidenced based/well reasoned	Excellent communication skills
Hardworking	Strong sense of ownership	Resourceful
Stoic	Patient advocate	Proactive
Calm	Good planning (anticipates contingencies)	Team player
Thoughtful	Thorough differential diagnoses	Anticipates bottlenecks
Mature	Superlative clinical skills	Prompt
Positive attitude	Thorough	Coordinates with consultants
Compassionate	Acknowledges and learns from errors	Good teaching, oversight
Accountable	Communicates with patients	Maximizes throughput
Reliable	Good prioritization skills	Helpful
Trustworthy		
Dedicated/committed		
Responsible		
Takes initiative <sup>a</sup>		
Diplomatic		
Tireless		
Motivated <sup>a</sup>		
Doesn't complain <sup>a</sup>		
Doesn't get frustrated <sup>a</sup>		

<sup>a</sup> Indicates characteristics identified as stereotypically masculine by prior literature.<sup>22</sup>

Note: Some of these characteristics map directly onto the 23 subcompetencies (eg, multitasking), but many span various competencies.

physicians rated the same resident over the course of PGY-3.

This multistage coding approach produced a robust and thorough inquiry, allowing themes to be (re)interpreted from the data,<sup>18</sup> while also systematically substantiating the validity of findings through rigorous collective analysis.

## Results

### Demographics

Between 2013 and 2015, 31% (22 of 71) of all residents (PGY-1 through PGY-3) at University Hospital were female, and 43% (29 of 67) of attending physicians were female. Of the PGY-3 residents with more than 15 comments, 13 of 35 (37%) were female. This suggests a slight underrepresentation of female residents, an overrepresentation of female attending physicians in our study site compared with national averages,<sup>1,8</sup> and a slight overrepresentation of female residents in our qualitative data compared with University Hospital's resident population (TABLE 2).

### Characteristics Valued During EM Residency

We examined the 1317 comments made by attending physicians about PGY-3 residents for insights into traits valued in the emergency department. Although procedural and diagnostic skills and knowledge were important, the data revealed that residents must also possess certain personality traits (confident, hardworking, calm); practice styles (documents well, performs under pressure); and management styles (communicative, able to multitask) to be considered superlative, even in procedure-specific evaluations. Many of these characteristics have been identified in prior research as stereotypically masculine traits (TABLE 1).<sup>22-24</sup>

While the residents in our sample were repeatedly evaluated on these characteristics, few managed to embody them at all times. One resident at University Hospital who stood out as consistently meeting expectations was Keith, a chief resident who regularly received praise, such as the following example:

**TABLE 2**  
 Characteristics of Comments Received by Emergency Medicine Residents During PGY-3

Pseudonym	Strong Praise	Strong Criticism	Possesses Valued Personality Traits	Lacks Valued Personality Traits	Rank, by End of PGY-3	N Evaluations	N Evaluations With Comments
<b>Female Residents (n = 13)</b>							
Zoey	7 "Performing at level of attending" "Rockstar" "Stellar job as team leader"	0 "Barely functioning at level of a PGY-3" "Should not be a senior resident" "I do not trust her presentations."	0 "Appropriate confidence and autonomy" "Modeled calm/collected professionalism on shift" "Positive attitude"	0 "Be more decisive!" "Leans on consultants when truly not needed" "Be more independent and self-correcting" "You need to take a more assertive leadership role."	1	50	21
Jane	9	0	4	1	8	67	26
Gabrielle	15	0	10	4	12	110	49
Denise	11	1	6	0	17	115	37
Lily	9	2	4	1	20	77	42
Claire	9	1	5	3	21	104	36
Amelia	5	0	1	3	24	67	33
Emma	17	1	7	5	25	112	43
Tamara	6	4	1	0	26	61	28
Nicole	8	0	6	3	28	134	45
Beth	6	4	7	3	29	62	38
Shannon	1	5	2	3	33	107	33
Patricia	7	11	8	5	35	161	77
<b>Median</b>	<b>8.0</b>	<b>1.0</b>	<b>5.0</b>	<b>3.0</b>		<b>104.0</b>	<b>37.0</b>
<b>Male Residents (n = 22)</b>							
James	22	0	7	0	2	106	46
Keith	25	0	6	0	3	103	47
Dylan	14	0	2	0	4	52	29
Philip	7	0	3	0	5	48	23
Brad	15	0	3	0	6	83	41
Carter	14	0	6	0	7	79	35
Ben	16	0	6	0	9	106	40
William	9	0	4	0	10	73	35

**TABLE 2**  
 Characteristics of Comments Received by Emergency Medicine Residents During PGY-3 (continued)

Pseudonym	Strong Praise	Strong Criticism	Possesses Valued Personality Traits	Lacks Valued Personality Traits	Rank, by End of PGY-3	N Evaluations	N Evaluations With Comments
Henry	10	1	3	3	11	66	31
Logan	17	0	6	0	13	94	43
Julian	13	1	2	2	14	59	32
Jackson	2	0	0	0	15	55	19
Isaac	13	0	4	0	16	81	38
Ethan	9	2	6	2	18	119	39
Matt	4	4	7	5	19	65	37
Oliver	5	0	10	1	22	88	31
Liam	6	3	4	1	23	138	45
Tyler	9	4	5	3	27	91	45
Aaron	11	1	7	1	30	111	36
Noah	5	1	3	4	31	111	44
Owen	4	9	1	4	32	73	37
Caleb	3	4	0	3	34	46	19
<b>Median</b>	<b>9.5</b>	<b>0.0</b>	<b>4.0</b>	<b>0.5</b>		<b>82.0</b>	<b>37.0</b>

Note: TABLE 2 only includes data from the 35 residents who received at least 15 comments.

“Keith continues to perform at the top of his class with regard to managing the entire [department] and being able to direct and teach interns [without] losing speed. He is also able to work [with] nursing to find ways to increase throughput. His greatest advantage over his peers is his ability to problem solve in difficult practice environments [without] losing patience or getting frustrated.” (John, attending)

In addition to illustrating what high praise sounds like, Keith’s evaluations indicated that this ideal was attainable.

### Gender Differences in Feedback

To better understand the gender gap in milestone attainment, we examined how different attending physicians evaluated residents’ performance on the milestones throughout PGY-3. For this analysis, we focused on residents with at least 15 comments (TABLE 2). Our primary finding was that female residents received less consistent feedback from attending physicians than male residents, particularly regarding personality traits valued in EM. A total of 62% (8 of 13) of female residents received both strong praise for their performance and strong criticism, compared with 45% (10 of 22) of male residents (TABLE 2). Half of male residents (11 of 22) received no negative comments regarding their possession of ideal EM personality traits from any attending physician, compared with 23% (3 of 13) of female residents. Only 1 female resident (Zoey) received no strong criticism and no negative personality comments, compared with 50% (11 of 22) of male residents. Sixty-two percent of female residents (8 of 13) were criticized multiple times for lacking valued EM personality traits compared with 36% of male residents (8 of 22).

The substantive content of the feedback also illustrated this gender gap. Male residents received consistent messages about what they needed to improve. As illustrated in the quotes that follow, multiple attending physicians noted the same concern about Matt’s diagnostic skills and autonomy:

- “I enjoy working with Matt ... However, as a third year, I continue to be

disappointed . . . I think he is capable of doing the job. However, *I still find his clinical decisions are limited to asking what the staff wants to do (as opposed to even offering a suggestion and then discussion [sic] differences).*” (Harrison, attending; emphasis added)

- “You [Matt] need to *take a more assertive leadership role*. You have a lot of military experience, but at this point *you should not defer so much to the attending. Confidently craft a plan and lay out to the attending how you will execute it*. Not all people need a million [dollar] work up.” (David, attending; emphasis added)
- “Overall good job, *but keep working on expanding that differential as you head into attend- inghood.*” (Ken, attending; emphasis added).

Matt, along with other poor(er) performing male residents (like Owen and Tyler; see TABLE 2) tended to receive reasonably consistent messaging from multiple attending physicians about what they needed to improve. Matt was even praised for his “self-awareness about . . . [his] limits,” though he never received praise for confidence or autonomy during PGY-3.

The comments directed toward female residents, however, often contained discordant feedback from different attending physicians, which generally focused on their lack of a specific personality characteristic, particularly autonomy and assertiveness. Consider the following comments (presented in chronological order) received by Emma, the female resident with the most instances of strong praise:

- “[Emma is] progressing well, very thoughtful, reliable, *appropriate confidence and autonomy.*” (Harrison, attending; emphasis added)
- “I would encourage Emma to *be more assertive*. During critical resuscitations, she should let those working around her *know that she is the team leader.*” (Adam, attending; emphasis added)
- “[Emma] *argues a lot with the attending, is very confident in her diagnosis, and has a hard time entertaining other possibilities.*” (Hillary, attending; emphasis added)

As is evident in Emma’s case, the discordance in feedback for female residents generally resulted from different attending physicians having dissimilar opinions, and not from the same attending physician providing a different commentary on the same topic on different occasions. Attending physicians were often consistent in their praise or sanctioning of resident personalities. For example, Harrison (quoted

earlier) praised Emma twice for her “appropriate autonomy.”

Emma also received mixed reactions with regard to her openness to feedback (which appeared in Hillary’s comment in the previous list) and also below (again, in chronological order):

- “*Receptive with feedback and asks good questions* on differential management. Aims at improving herself even though already a great performer.” (Sofia, attending; emphasis added)
- “[Emma] can *improve by being more open to feedback and constructive criticism* and recognizing that others may be able to contribute to her learning and provide another perspective on practice and patient care.” (Frank, attending; emphasis added)

This tension between demonstrating autonomy while still being open to attending physicians’ directions was a theme that appeared in feedback for multiple female residents but never appeared for male residents. Gabrielle, Tamara, Jane, and Beth were all both praised and criticized for their receptiveness to feedback and their autonomy. Beth, for example, received the following comments about her handling of constructive feedback (in chronological order):

- “I think you [Beth] have a *natural ability to communicate well with regard to difficult information/constructive feedback*. You develop a very good report [sic] with students and your colleagues because you treat them with respect. It’s a wonderful thing to work around. *Thanks for the role-modeling of this intangible.*” (Steven, attending; emphasis added)
- “[Beth is] *interested in learning* despite of being ready for graduation.” (Martin, attending, emphasis added)
- “[Beth] *seemed to respond negatively to my input on her plans last shift . . . I know she’s late in the third year and needs progressive autonomy, but she seemed to have a negative attitude toward supervision.*” (Richard, attending; emphasis added)

By comparison, only 1 male resident (Owen) was criticized for being “sensitive to feedback” by the attending physician (Brian), and 1 male resident was actually praised for the same thing: “Doing well. [Oliver is] sometimes argumentative, but he is trying to assert his confidence” (Michael, attending). This further illustrates the likely salience of gender to

interpreting residents' abilities to exhibit appropriate confidence and assertiveness.

## Discussion

Our goal for this study was to examine qualitative differences in the feedback that male and female EM residents received in order to better understand the emergence of a substantial gender gap in ACGME milestone attainment in PGY-3 established by our prior quantitative work.<sup>17</sup> To do this, we analyzed text comments written by attending physicians for residents at the time of their evaluation for milestone attainment and identified 2 findings that likely are relevant to females' lagging performance and gender equality in EM graduate medical education.

First, the data revealed that the ideal EM resident must possess many stereotypically masculine traits: he or she must be a calm, decisive, confident leader who efficiently manages scarce resources in order to achieve the best outcomes for patients.<sup>22–25</sup> This ideal may be related to the specialty's historical roots in the military,<sup>26</sup> which also has highly masculinized norms.<sup>27</sup> Of course, it is also the case that in EM, hesitation may mean the difference between life and death and that some of these stereotypically masculine traits may serve EM physicians well.

Second, we found that when males struggled, they received consistent feedback from different attending physicians regarding the aspect of their performance that needed work. In contrast, different attending physicians had discordant views about what female residents needed to improve.

This inconsistent feedback for women was particularly apparent around issues of autonomy and leadership, which are personality characteristics that may be more challenging to improve than procedural skills. Female residents were frequently praised for their performance as an autonomous leader by 1 attending, only to be criticized by another for being argumentative.

Inconsistent feedback is worthy of attention for 2 reasons. First, because inconsistent feedback disproportionately appeared among female residents, and rarely among male residents, female residents may be receiving poorer-quality mentoring and instruction. This could have consequences for their ability to progress, particularly if that feedback is less related to their actual performance than it is for male residents. Prior research has shown that consistent and clear feedback is important for the improvement of performance.<sup>28</sup> Future research should examine the consequences this gender gap in consistent feedback may have for actual learning.

Second, these critiques are highly gendered. "Assertiveness" and "autonomy" are more associated with masculinity than femininity; this is potentially an instance when male residents have an easier time meeting expectations than their female colleagues.<sup>11</sup> This is consistent with a much broader scholarship showing that women have a harder time being evaluated as competent on traits that are traditionally considered masculine,<sup>12,29</sup> particularly in organizations that have been dominated generally by men.<sup>11</sup> The emergency department fits both of these criteria, given the underrepresentation of women in EM<sup>1,8</sup> and the masculine traits valued in this setting. Although our data did not allow evaluation of whether attending physicians adhere to biased beliefs about female residents, past research suggests that this may be the case. When evaluating grant proposals,<sup>14</sup> job applicants and tenure candidates,<sup>15</sup> graduate student applicants,<sup>30</sup> teaching,<sup>31</sup> and mentoring ability,<sup>13</sup> women, on average, are deemed less competent or ideal than their male counterparts.

It also is possible that female residents' performance is less consistent than that of male residents, and the comments received may be attending physicians' attempts to encourage improvement. Past research has shown that female residents reported feeling stressed or uncomfortable with having to violate stereotypically feminine behavioral norms, like using directive language and a loud voice while leading a code.<sup>32</sup> This discomfort could trigger stereotype threat and harm their performance.<sup>33</sup> Future research should continue to examine the complex processes that may undergird the gender gap in competency evaluations that research has found in the emergency department.<sup>17</sup>

There are limitations to our qualitative study. First, we lack the context in which the comments were made, although both male and female residents were evaluated against the same ACGME milestone criteria. Second, our data were drawn from a single hospital. Finally, it is unfortunate that we lack information in our data about residents' or attending physicians' race or ethnicity, which likely matters, given the underrepresentation of people of color in medicine.<sup>34</sup>

There are advantages to our analysis. First, faculty members were not being observed by researchers when making their evaluations of residents, diminishing a Hawthorne effect or social desirability bias. Second, previously published findings allow us to be more confident that factors, such as the skill being evaluated (procedural or otherwise), the gender of the attending physician, or the local culture of the hospital, likely do not fully explain the gender gap

in evaluations, as those factors were controlled in our statistical models and found to be insignificant.<sup>17</sup>

To further explore this important area, real-time observational research and interviews with EM residents and attending physicians would contribute important knowledge and complement this study. Examining other EM residency programs would also be informative, particularly if studies considered the role of the local emergency department culture.

## Conclusion

Our research contributes to understandings of gender inequality in graduate medical education. Namely, initiatives to reduce gender bias<sup>35–38</sup> should sensitize faculty to potential biases in their perceptions of female residents' capabilities and to the importance of providing residents' consistent and clear feedback.

## References

- Lautenberger DM, Dandar VM, Raezer CL et al. The state of women in academic medicine: the pipeline and pathways to leadership, 2013–2014. <https://members.aamc.org/eweb/upload/The%20State%20of%20Women%20in%20Academic%20Medicine%202013-2014%20FINAL.pdf>. Accessed August 3, 2017.
- Edmunds LD, Ovseiko PV, Shepperd S, et al. Why do women choose or reject careers in academic medicine? A narrative review of empirical evidence. *Lancet*. 2016;388(10062):2948–2958.
- Wehner MR, Nead KT, Linos K, et al. Plenty of moustaches but not enough women: cross sectional study of medical leaders. *BMJ*. 2015;351:h6311.
- Kuhn GJ, Abbuhl SB, Clem KJ, et al. Recommendations from the Society for Academic Emergency Medicine (SAEM) Taskforce on women in academic emergency medicine. *Acad Emerg Med*. 2008;15(8):762–767.
- Cydulka RK, D'Onofrio G, Schneider S, et al. Women in academic emergency medicine. *Acad Emerg Med*. 2000;7(9):999–1007.
- Freund KM, Raj A, Kaplan SE, et al. Inequities in academic compensation by gender: a follow-up to the national faculty survey cohort study. *Acad Med*. 2016;91(8):1068–1073.
- Jena AB, Olenski AR, Blumenthal DM. Sex differences in physician salary in US public medical schools. *JAMA Intern Med*. 2016;176(9):1294–1304.
- Cheng D, Promes S, Clem K, et al. Chairperson and faculty gender in academic emergency medicine departments. *Acad Emerg Med*. 2006;13(8):904–906.
- Nasca TJ, Philibert I, Brigham T, et al. The next GME accreditation system—rationale and benefits. *N Engl J Med*. 2012;366(11):1051–1056.
- Castilla EJ, Benard S. The paradox of meritocracy in organizations. *Admin Sci Q*. 2010;55(4):543–676.
- Correll SJ, Ridgeway CL, Zuckerman EW, et al. It's the conventional thought that counts: how third-order inference produces status advantage. *Am Soc Rev*. 2017;82(2):297–327.
- Foschi M. Double standards for competence: theory and research. *Ann Rev Soc*. 2000;26:21–42.
- Hirshfield LE. “She’s not good with crying”: the effect of gender expectations on graduate students’ assessments of their principal investigators. *Gender Educ*. 2014;26(6):601–617.
- Wenneras C, Wold A. Nepotism and sexism in peer-review. *Nature*. 1997;387(6631):341–343.
- Steinpreis RE, Anders KA, Ritzke D. The impact of gender on the review of the curricula vitae of job applicants and tenure candidates: a national empirical study. *Sex Roles*. 1999;41(7):509–528.
- Sprague J, Massoni K. Student evaluations and gendered expectations: what we can’t count can hurt us. *Sex Roles*. 2005;53(11):779–793.
- Dayal A, O’Connor DM, Qadri U, et al. Comparison of male vs female resident milestone evaluations by faculty during emergency medicine residency training. *JAMA Intern Med*. 2017;177(5):651–657.
- Varpio L, Ajjawi R, Monrouxe LV, et al. Shedding the cobra effect: problematising thematic emergence, triangulation, saturation and member checking. *Med Educ*. 2017;51(1):40–50.
- Accreditation Council for Graduate Medical Education; American Board of Emergency Medicine. The Emergency Medicine Milestone Project. July 2015. <https://www.acgme.org/Portals/0/PDFs/Milestones/EmergencyMedicineMilestones.pdf>. Accessed August 3, 2017.
- Ivankova NV, Creswell JW, Stick SL. Using mixed-methods sequential explanatory design: from theory to practice. *Field Methods*. 2006;18(1):3–20. [http://wtgrantmixedmethods.com/sites/default/files/literature/Ivankova%20etal\\_2006\\_mixed%20methods%20sequential%20design.pdf](http://wtgrantmixedmethods.com/sites/default/files/literature/Ivankova%20etal_2006_mixed%20methods%20sequential%20design.pdf). Accessed August 3, 2017.
- Lofland J, Snow DA, Anderson L, et al. *Analyzing Social Settings: A Guide to Qualitative Observation and Analysis*. 4th ed. Belmont, CA: Thomson/Wadsworth; 2006.
- Gorman EH. Gender stereotypes, same-gender preferences, and organizational variation in the hiring of women: evidence from law firms. *Am Soc Rev*. 2005;70(4):702–728.
- Koenig AM, Eagly AH, Mitchell AA, et al. Are leader stereotypes masculine? A meta-analysis of three research paradigms. *Psychol Bull*. 2011;137(4):616–642.



24. Fischbach A, Lichtenthaler PW, Horstmann N. Leadership and gender stereotyping of emotions. *J Personnel Psychol*. 2015;14:153–162.
25. American Association of University Women. Barriers and bias: the status of women in leadership. 2016. <http://www.aauw.org/resource/barriers-and-bias>. Accessed August 3, 2017.
26. Goniewicz M. Effect of military conflicts on the formation of emergency medical services systems worldwide. *Acad Emerg Med*. 2013;20(5):507–513.
27. Braswell H, Kushner HI. Suicide, social integration, and masculinity in the US military. *Soc Sci Med*. 2012;74(4):530–536.
28. Correll S, Simard C. Research: vague feedback is holding women back. *Harvard Business Review*. April 29, 2016. <https://hbr.org/2016/04/research-vague-feedback-is-holding-women-back>. Accessed August 3, 2017.
29. Ridgeway CL. Framed before we know it: how gender shapes social relations. *Gender Soc*. 2009;23(2):145–160.
30. Moss-Racusin CA, Dovidio JF, Brescoll VL, et al. Science faculty's subtle gender biases favor male students. *Proc Natl Acad Sci USA*. 2012;109(41):16474–16479.
31. Laube H, Massoni K, Sprague J, et al. The impact of gender on the evaluation of teaching: what we know and what we can do. *NWSA J*. 2007;19(3):87–104.
32. Kolehmainen C, Brennan M, Filut A, et al. Afraid of being “witchy with a ‘b’”: a qualitative study of how gender influences residents’ experiences leading cardiopulmonary resuscitation. *Acad Med*. 2014;89(9):1276–1281.
33. Burgess DJ, Joseph A, Van Ryn M, et al. Does stereotype threat affect women in academic medicine? *Acad Med*. 2012;87(4):506–512.
34. Deville C, Hwang WT, Burgos R, et al. Diversity in graduate medical education in the United States by race, ethnicity, and sex, 2012. *JAMA Intern Med*. 2015;175(10):1706–1708.
35. Krupat E, Pololi L, Schnell ER, et al. Changing the culture of academic medicine: the C-Change learning action network and its impact at participating medical schools. *Acad Med*. 2013;88(9):1252–1258.
36. Carnes M, Devine PG, Baier Manwell L, et al. Effect of an intervention to break the gender bias habit for faculty at one institution: a cluster randomized, controlled trial. *Acad Med*. 2015;90(2):221–230.
37. Powell D, Scott JL, Rosenblatt M, et al. Commentary: a call for culture change in academic medicine. *Acad Med*. 2010;85(4):586–587.
38. Westring AF, Speck RM, Dupuis Sammel M, et al. Culture matters: the pivotal role of culture for women’s careers in academic medicine. *Acad Med*. 2014;89(4):658–663.



**Anna S. Mueller, MA, PhD**, is Assistant Professor, Department of Comparative Human Development and the College, University of Chicago; **Tania M. Jenkins, MA, PhD**, is Assistant Professor, Department of Sociology, Temple University; **Melissa Osborne, MA**, is Fourth-Year PhD Student, Department of Sociology, University of Chicago; **Arjun Dayal, MD**, is Resident Physician (PGY-1), University of Chicago/NorthShore University HealthSystem; **Daniel M. O’Connor, MD**, is Resident Physician (PGY-1), Crozer-Chester Medical Center; and **Vineet M. Arora, MD, MAPP**, is Associate Professor of Medicine, Assistant Dean for Scholarship and Discovery, and Director of Graduate Medical Education Clinical Learning Environment Innovation, University of Chicago.

Funding: This project was supported by a grant from the National Center for Advancing Translational Sciences. Additional funding was provided by a University of Chicago Diversity Small Grant. Dr Jenkins received salary support from the Canadian Institutes of Health Research (No. 140811).

Conflict of interest: Drs O’Connor and Dayal co-developed InstantEval, which was used to collect the evaluation data in this study, and have a financial interest in this product. Dr Arora receives honoraria from the American Board of Internal Medicine.

A prior version of this study was presented at the Sociologists for Women in Society Winter Meeting, Albuquerque, New Mexico, February 9–12, 2017.

The authors would like to thank Keith Mausner, MD, FAAEM, for his insightful feedback on earlier drafts of this paper.

Corresponding author: Anna S. Mueller, MA, PhD, University of Chicago, Department of Comparative Human Development, 1126 East 59th Street, Suite 103, Chicago, IL 60637-1580, 773.834.1179, [amueller@uchicago.edu](mailto:amueller@uchicago.edu)

Received February 17, 2017; revision received May 4, 2017; accepted May 22, 2017.