# Who Gets the Benefit of the Doubt? Performance Evaluations, Medical Errors, and the Production of Gender Inequality in Emergency Medical Education

Alexandra Brewer,[a] [iD] Melissa Osborne,[b]
Anna S. Mueller,[c] Daniel M. O'Connor,[d] [iD]
Arjun Dayal,[e] and Vineet M. Arora[e]

## Abstract

Why do women continue to face barriers to success in professions, especially male-dominated ones, despite often outperforming men in similar subjects during schooling? With this study, we draw on role expectations theory to understand how inequality in assessment emerges as individuals transition from student to professional roles. To do this, we leverage the case of medical residency so that we can examine how changes in role expectations shape assessment while holding occupation and organization constant. By analyzing a dataset of 2,765 performance evaluations from a three-year emergency medicine training program, we empirically demonstrate that women and men are reviewed as equally capable at the beginning of residency, when the student role dominates; however, in year three, when the colleague role dominates, men are perceived as outperforming women. Furthermore, when we hold resident performance somewhat constant by comparing feedback to medical errors of similar severity, we find that in the third year of residency, but not the first, women receive more harsh criticism and less supportive feedback than men. Ultimately, this study suggests that role expectations, and the implicit biases they can trigger, matter significantly to the production of gender inequality, even when holding organization, occupation, and resident performance constant.

## Keywords

medical sociology, gender, work, education

In recent decades, women have made substantial gains in educational contexts, breaking down barriers into STEM fields and other areas of higher education historically dominated by men (Buchmann and DiPrete 2006; DiPrete and Buchmann 2013; Riegle-Crumb 2010). And yet, research shows that women continue to lag behind their men counterparts when it comes to representation, remuneration,

[a]Wake Forest University
[b]Western Washington University
[c]Indiana University
[d]Harvard University
[e]University of Chicago

**Corresponding Author:**
Alexandra Brewer, Department of Sociology, Wake Forest University, 02A Kirby Hall, Winston-Salem, NC 27109
Email: breweral@wfu.edu

and promotions in the workforce—even when accounting for differences in productivity and training (England 2010; Heilman 2001; Quadlin 2018; Wennerås and Wold 1997). Why are women assessed as less capable than men in the workplace if they outperform them in schools?

To address questions of women's inequality in the workplace, sociological research has spent considerable effort examining how organizational roles and their attached expectations contribute to why women are denied raises and promotions after they are hired (Acker 1990; Gorman and Kmec 2009; Ridgeway 2009; Ridgeway and Smith-Lovin 1999). Indeed, scholars have found ample evidence that men have an easier time fitting the vision of an ideal worker and leader in the workplace than do equally qualified women. However, to date, these insights have not been fully leveraged to understand the barriers women face when transitioning *between* education and work. Given that past research demonstrates the student role is largely "feminized" (DiPrete and Buchmann 2013; Downey and Vogt Yuan 2005; Jacob 2002), whereas the worker role is traditionally "masculinized" in historically male-dominated professions (Acker 1990; Gorman and Kmec 2009; Ridgeway 2009), there is good reason to believe that changes in role expectations inform the emergence of gender inequality in the school-to-work transition. Of course, it is empirically challenging to find a case that allows researchers to isolate processes related only to changes in role expectations, given that the organizational context is usually also changing as individuals transition from schools to workplaces.

With this study, we address this gap in the literature by leveraging the case of medical residency, which is the formal training period that occurs after medical school but before independent medical practice. This case offers a number of key advantages, including that resident physicians (or "residents") are expected to formally transition from the "student" to "professional" role during medical residency (ACGME and ABEM 2015), a

transition that takes place within a single organization in which other salient factors, such as the process of evaluation and the identity of evaluators, are not also changing. Additionally, past research has established that a statistically significant and substantial gender gap in evaluations emerges across this same period that is not dependent on evaluator gender (Dayal et al. 2017). Finally, this case allows us to take advantage of a unique dataset of 2,765 combined textual and numeric performance evaluations of emergency medicine (EM) residents that are composed in-the-moment by attending physicians (or "attendings"; i.e., medical doctors who supervise patient care and the education of residents). We compare 67 attending physicians' (29 women and 38 men) evaluations of the work of 35 first-year residents (8 women and 27 men) to their evaluations of 36 third-year residents (14 women and 22 men) in a three-year[1] accredited EM training program. These two stages of residency, the first and final years, are when role-expectations are most clear and distinct. In the first year, being a student and a learner is the explicitly stated expectation, whereas in the third and final year, residents are explicitly assessed for their ability to perform as independent and reliable medical practitioners after graduation (ACGME and ABEM 2015). We also focus on evaluations referencing serious medical errors (as determined by the physicians on our team), which allows us to examine the language used in the evaluations while holding performance somewhat constant.

## LITERATURE REVIEW

### Gender Inequality in the School-to-Work Transition

One of the most perplexing questions in the sociology of gender, work, and inequality is why women, on average, continue to underperform on multiple metrics in the workplace (Correll, Benard, and Paik 2007; England 2010) despite outperforming men, on average, in schools (Buchmann and DiPrete 2006; DiPrete and Buchmann 2013; Riegle-Crumb

2010). Women have graduated from college at higher rates than men since the 1980s (Snyder, de Brey, and Dillow 2018), long enough to complicate the idea that workplace inequality is purely a "pipeline problem" (Berryman 1983; Hanson, Schaub, and Baker 1996). Indeed, women seem to be thriving within schools: women and girls, on average, earn higher grades and standardized test scores than do men and boys (Buchmann and DiPrete 2006; Buchmann, DiPrete, and McDaniel 2008; Downey and Vogt Yuan 2005) and, as of 2018, women represent 57 percent of all college graduates, and they complete graduate school at higher rates than men (Snyder et al. 2018). These patterns suggest women are assessed as competent and capable in educational contexts. Additionally, women are choosing to pursue degrees in career fields once thought of as masculine, including business, law, and medicine, at higher rates (England et al. 2007; England and Li 2006).

And yet, these achievements have not translated to workplace equality: women continue to earn lower salaries than men and fill fewer high-status positions (Correll et al. 2007; England 2010). This inequality does not simply reflect women's lack of ambition (Correll 2004; Wennerås and Wold 1997), nor can it be fully explained by time taken off for parenting (Budig and England 2001; Correll et al. 2007). Rather, a significant body of sociological research demonstrates that women's underrepresentation in high-status roles in workplaces results from practices of exclusion. For example, women are often left out of informal social networks that can be crucial to professional advancement (O'Meara and Stromquist 2015; Xu and Martin 2011). They are burdened with more service and mentoring work—tasks that are less likely to lead to promotions and raises—compared with men (Babcock et al. 2017; Misra, Lundquist, and Templer 2012). And, perhaps most importantly for workplace inequality, they are consistently assessed as less competent and capable than their men colleagues in both formal and informal evaluations (Bornmann, Mutz, and Daniel 2007; Correll et al.

2018; Reskin 1993; Trix and Psenka 2003; Wennerås and Wold 1997). Even as many workplaces are making efforts to structure evaluation in ways that are objective and gender-neutral (Cecchi-Dimeglio 2017; Heilman 2001; Reskin and McBrier 2000), there is doubt about whether such measures can succeed in reducing bias against women in performance evaluations and other assessments (Acker 1990; Gorman 2005).

Exclusionary practices shape patterns of gender inequality in the workplace because they lead to fewer hires, raises, promotions, and bonuses for women (Castilla 2008; Steinpreis, Anders, and Ritzke 1999), as well as higher rates of attrition from male-dominated fields (Britton 2017; Jacobs 1989). Indeed, women's overrepresentation in historically female-dominated fields of study and occupations, like education and nursing, may stem, in part, from exclusionary practices within male-dominated fields, like STEM, that earn higher salaries and status (Cech et al. 2011; Jacobs 1989).

Taken together, these literatures reveal a perplexing puzzle: women seem to be assessed as more capable than men when in educational settings, but as less capable when in the workplace, even when capability is held constant. To better understand this phenomenon, we first examine the ways in which different organizational roles are linked to gendered expectations. We then use this sociological literature to elaborate our theoretical argument about how role expectations affect the emergence of inequality in the school-to-work transition.

## Role Expectations and Gender Inequality

Role expectations theory describes how certain competencies and character ideals are implicitly linked with either men or women (Fiske et al. 2002; Ridgeway 2009; Ridgeway and Correll 2004). Seemingly gender-neutral social roles are often sex-typed either "male" or "female" based on the status and competencies associated with them, and thus they are seen as more appropriate for one gender

or the other (Reskin 1993). As a result, either men or women may be advantaged when performing different organizational roles that are seen as stereotypically appropriate for someone of their gender.

In workplaces, role expectations often disadvantage women, especially when it comes to career advancement and conventionally masculine work. Traditionally, certain kinds of work (e.g., STEM, law, business, medicine) and high-status roles are implicitly associated with masculine traits (like independence and confidence), whereas subordinate positions are associated with femininity (Acker 1990; Correll et al. 2007; Kanter 1977; Ridgeway 2009; Ridgeway and Correll 2004). These stereotypes, although sometimes intangible, are often embedded in organizational structures, like criteria for evaluation, hiring, and promotions. For example, hiring criteria may contain stereotypically masculine traits like "ambitious" and "independent," and when they do, men are more likely, and women less likely, to be hired (Gorman 2005). Thus, when employees are assessed against abstract ideals that reflect conventional masculinity, men appear to be "naturally" more hirable and promotable regardless of actual qualifications. Women, on the other hand, are viewed as less competent when assessed against implicitly masculine ideals.

And yet in schools, role expectations may actually *promote* women's success (Morris 2011). Stereotypes about men's and boys' superior competence at science and math still permeate educational contexts (Cech et al. 2011; Correll 2001), but the student role is in many ways "feminized." For example, the image of the "ideal student" includes conventionally feminine traits, such as organization and deference to authority, producing an advantage for women and girls (DiPrete and Buchmann 2013; Downey and Vogt Yuan 2005; Jacob 2002). At the same time, some scholars argue that conventionally masculine behaviors are actually *punished* in schools, pointing to higher rates of disciplinary action and learning disability diagnoses among boys (Kleinfeld 1998; Mulvey 2010). Thus, women

and girls, and especially white women and girls (Morris 2007), may appear "naturally" more deserving of educational rewards, such as high grades and other positive evaluations, than boys within educational contexts.

With this study, we posit that changes in role expectations that shift perceptions of men's and women's competency across evaluatory contexts will help us understand the gender disparities that emerge between school and work. We expect to find that as students progress toward workforce entry, new expectations will be activated that reflect professional roles, and because the stereotypes that come with these expectations likely favor men, women who were previously perceived as successful might start to be seen as less talented. It can be challenging, however, to study the nuanced ways in which role expectations produce gender inequality in the student-to-professional transition because work and education typically take place in different organizations with different cultures and bureaucratic structures. This makes it difficult to isolate how much the change in role expectations matters to the broader set of processes that facilitate the emergence of inequality because multiple aspects of role and performance evaluation are shifting simultaneously. An ideal study of gender inequality in the school-to-work transition would hold evaluators and organizational context constant to provide a clearer window onto the salience of role expectations in explaining why women outperform men in schools but fall behind them in the workplace. We are able to do this by leveraging the case of medical residency training.

### The Case of Medical Residency

Medical residency provides a unique opportunity to examine the interaction between gender and role expectations while holding career choice and organizational context constant, as individuals transition between the role of student and the role of professional. Residents occupy a somewhat ambiguous middle space between these two roles. They possess medical degrees (MDs or DOs), but they are only

permitted to practice medicine under the supervision of attending physicians.

Strict institutionalized standards differentiate interns (residents in their first year) from senior residents (those nearing graduation) (ACGME and ABEM 2015). Interns fill a role closer to that of an advanced student: they are still learning basic medical tasks involved in patient care (e.g., how to do a procedure or make a diagnosis), often in the context of a new and unfamiliar hospital, and thus their work is heavily supervised. Senior residents, on the other hand, fill a role closer to that of an attending: they supervise medical students and lower-level residents, and they are responsible for managing patient care, with attendings supervising in the background.

Although the professional hierarchy within medical residency may be different from that in other workplaces, disadvantage in the transition from student to professional is not unique to medicine. Similar patterns appear across a variety of different historically male-dominated career paths, including science, academia, and law (Jacobs 1996; Kay and Gorman 2008; Xie and Shauman 2003). Yet, the fact that the transition from student to professional roles in residency takes place within the context of a single organization, occurs gradually over the course of three years, and is concretely spelled out in institutionalized and theoretically objective performance expectations established by the Accreditation Council for Graduate Medical Education (ACGME) allows us as researchers to examine a set of crucial, but often hard-to-observe, processes.

Within the broader case of medical residency, we focus on emergency medicine (EM) for several important reasons. First, prior research has established that EM values stereotypically masculine traits in its residents, like leadership and assertiveness, as part of their professional role ideals (Mueller et al. 2017). This may be because of the fast-paced and intense nature of EM (Choo 2017); however, EM's emergence out of military field medicine (Goniewicz 2013), with its highly masculinized culture (Braswell and Kushner 2012), likely also contributes. As

might be expected based on these traits, EM has lower levels of gender parity compared with most other medical specialties (Lautenberger et al. 2014): women comprise 48 percent of residents and 38 percent of medical faculty (attendings in academic or university-affiliated hospitals, rather than "community" hospitals) in medicine as a whole, but they represent only 38 percent of all EM residents (DeFazio et al. 2017) and 28 percent of EM faculty (Bennett et al. 2019).[2]

Second, as we have alluded to, EM has extremely institutionalized and formal expectations for resident performance that are tied to the year of residency. EM was an early adopter of the ACGME's Next Accreditation System (NAS), a nationally standardized system for evaluating resident performance (Nasca et al. 2012), which is the latest part of a long-standing effort to standardize graduate medical education by providing objective standards ("milestones") for assessing resident performance and ensuring physicians across the country receive training that is similar in content and rigor (Beck 2004). The clear differences in role expectations between intern and senior residents outlined in the NAS facilitate our tracking changes in expectations and their consequences over time.

Third, prior research, based on the same dataset we use here, has determined that a significant gender gap in resident evaluation scores emerges over the course of residency (Dayal et al. 2017). Specifically, in the third year of residency, women's numerical evaluations lag significantly and substantially behind those of men, so much so that the results suggest women require, on average, an additional three to four months of residency training than men (Dayal et al. 2017). Interestingly, this gendered lag is not present in the first or second years of residency and is robust to controls for the hospital type (e.g., academic versus community), the attending's gender, and the gender-match between residents and attendings. Essentially, both men and women attendings are responsible for producing the gender gap in resident evaluations in the ED (Dayal et al. 2017). One

limitation with prior work, including our own, is that it does little to examine why this gender gap in evaluations emerges.

The final reason for our focus on EM is that, in response to the NAS framework, several EM programs implemented software on medical faculty's mobile devices and computers that allow them to provide numerical and textual feedback to residents about their milestone attainment in real-time. This software generates unique data that provide a window into everyday perceptions of residents' competency, details of which we provide below.

## METHODS
### Data Collection

Our data come from a software application called InstantEval V2.0 (Monte Carlo Software LLC, Annandale, VA), which was designed by Dayal and O'Connor, two of the physician co-authors on this article, to help EM attending physicians efficiently assess residents' performance on ACGME NAS milestones.[3] This software works as an application on attendings' mobile devices and computers. Attendings can choose when to complete evaluations, whom to evaluate, and the number of evaluations to complete, although most training programs using the application encourage one to three evaluations per shift. Each evaluation encourages the assessment of an ACGME Emergency Medicine Milestone Project–based performance level (a numerical score from one to five) on 1 of 23 possible individual EM subcompetencies.[4] Attendings can see descriptors of the individual milestones (and corresponding numerical scores) on the app before assigning a rating, allowing them to reference formal standards of evaluation immediately prior to making an assessment.

Attendings have discretion over when to provide textual comments, which are the focus of this study. These comments can be up to 1,000 characters in length and can be written "publicly" (i.e., visible to both the resident being evaluated and all attendings) or "privately" (i.e., only for attendings to see). We do not have data on how residency programs use these resident evaluations, but many programs use data from similar in-the-moment evaluations of ACGME subcompetencies to make broader assessments about residents' rank, abilities, and preparedness for next career stages. Thus, these evaluations likely represent a permanent record of residents' performances in the ED, including their mistakes and errors.

For this study, we leverage data from a single three-year ACGME-accredited Emergency Medicine residency program that we call "University Hospital" (a pseudonym) over two years (from July 1, 2013, to July 1, 2015). We selected this hospital because it was the largest (in terms of attending and resident population) of the available sites that enabled the textual comments feature of the InstantEval V2.0 app and had two full years of data. Dayal and colleagues (2017) found that the gender gap in evaluations is not present or statistically significant in PGY1 (post-graduate year 1), but is substantial and statistically significant in PGY3 (post-graduate year 3), so comparing PGY1 residents to PGY3 (with the same attendings evaluating both groups) offers an interesting qualitative comparison that can reveal why attendings may be more prone to gender bias in PGY3 compared to PGY1. Thus, our study focuses on 2,765 direct observation evaluations with textual comments that were collected from two cohorts of PGY1 EM residents (*N* evaluations = 1,448) and two cohorts of PGY3 residents (*N* evaluations = 1,317). Specifically, we analyze evaluations of 71 residents (8 women and 27 men in PGY1; 14 women and 22 men in PGY3) by 67 attending physicians (29 women and 38 men). Residents received between 18 and 77 total comments, with an average of 37.9 comments per year that data were collected.

We did not find a statistically significant difference in the number of comments received by men and women. All residents remained in the study sample for the full amount of time expected based on their cohort. Attrition rates from residency are low, and EM residency has the lowest rate of attrition (less than 1 percent)

of all major medical specialties (Lu et al. 2019). Our dataset does not include information about race or other demographic characteristics beyond gender for either attendings or residents, a limitation we discuss in more detail in the Discussion section.

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. Participating hospitals were informed that written and numerical comments may be used for future research purposes, but they were not told that comments would be analyzed specifically with regard to gender. In this sense, our methods of data collection may have provided some protection against social desirability bias (Nederhof 1983).

### Analytic Plan

Our analysis was guided by a sequential explanatory analytic design (Ivankova, Creswell, and Stick 2006), in which we used qualitative methods to better understand previously established quantitative findings. Specifically, we began our analysis with the knowledge that a gender gap exists in evaluations in the third year but not the first year of residency (Dayal et al. 2017). To probe this disparity, we first conducted a qualitative content review of the ACGME's Emergency Medicine Milestones (ACGME and ABEM 2015) to understand the formal expectations for residents and to enable evaluation of how these formal standards are operationalized in daily practice. Once familiar with the ACGME milestones, we compared performance evaluations of two cohorts of PGY1 residents and two different cohorts of PGY3 residents to examine why the gender disparity emerges.

To guard against confirmation bias during data analysis, we developed a multistage, multi-analyst procedure for coding and analyzing the data involving both sociologists and physicians. We also 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. Our analytic procedure began with simultaneous open coding of all comments to develop themes from the data (Lofland et al. 2006) and to ensure accurate understanding of comments. We then undertook an additional order of thematic coding and analysis, focusing on subthemes within these data. Following this coding, we analyzed gender differences in the comments residents received at different stages of training. At every stage of data analysis, at least two sociologist team members coded every comment. All relevant comments that involved a medical error were also coded by at least two of the physicians on our research team. Their coding was reviewed for consistency and clarity by at least one medical sociologist. Any discrepancies between codes and rating of medical errors was discussed collectively by all analysts and consensus was reached in all cases.

### Summary of Analytic Themes

To orient the reader to the concepts we present in our Findings section, we briefly summarize the analytic themes that emerged from our coding process. Italicized words and phrases indicate codes that will be discussed in detail.

*Valence and character ideals.* All comments were coded as either *positive, negative, mixed*, or *neutral* based on whether they contained language that was complimentary of resident performance, critical of resident performance, both, or neither. Several character ideals were identified based on traits/skills that residents would get positive feedback for possessing and negative feedback for not possessing. *Leadership* was defined as the ability to take charge of others, of an individual patient, and of the ED as a whole. *Communication* was defined as the ability to ask for and share information with others.

*Evaluative reference group.* These codes analyze textual comments for the explicitly-stated evaluative reference group

against which resident performance is being compared. Comments that reference the *student role* assess residents for their general ability to fulfill the duties of residency or compare them to other residents in their class. These comments state that residents are either *behind* their expected abilities or their classmates, *on par* with expectations or others in their class, or *ahead* of their expected level of competency or others in their class. Comments that reference the *colleague role* assess residents for their ability to succeed in independent practice after residency. These comments state that residents are either *not ready, almost ready*, or *ready* for their next career stages.

   *Medical errors.* We coded medical errors based on a definition from the Institute of Medicine (2000) as "the failure of a planned action to be completed as intended or the use of a wrong plan to achieve an aim." *Minor errors* were defined as those that violated best practices for physicians but did not pose a serious threat to patients. *Major errors* were those with the potential to harm patients. Within medical errors, we also coded for qualitative differences in the language attendings used when commenting on the error. Supportive feedback includes cases where attendings *reassure* residents of their abilities (e.g., "We were a little slow in instituting EGDT in a patient with severe sepsis. *Otherwise, she managed patients appropriately* during a four-hour G2 shift. *Good to work with her*—competent and trustworthy") and offer them the *benefit of the doubt* (e.g., "Noticed more so today you were distracted while at work. . . . *I know you have it in you*, but today more than in days before I was struck by this"). Critical feedback includes cases where attending feedback is *hurtful*, calling into question residents' basic ability to perform the role of EM physician (e.g., "Overall, seems *clueless and a disaster to work with*"), and when attendings *blame the resident* (e.g., "She is trying hard and definitely improving but still frustrating that *she has not made the personal effort* to learn her

drugs. This is something faculty cannot do for her").

## FINDINGS

In the following section, we present our empirical analysis of why the previously established gender disparity in evaluations emerges between year 1 (PGY1) and year 3 (PGY3) of emergency medical residency. We begin by describing how role-expectations shift between these two years of residency, and we then examine how men and women differentially experience these shifts in expectations. Finally, we examine whether differences in resident performance, rather than differences in *perceptions of* resident performance, may contribute to the observed gender disparities in evaluations. To do this, we hold clinical performance somewhat constant by leveraging a subset of comments about medical errors and comparing attendings' gendered reactions to errors of similar severity.

### Shifting Roles, Shifting Reference Groups

Embedded within the ACGME's Emergency Medicine Milestones (ACGME and ABEM 2015) are differential expectations for resident performance, such that in PGY1 a student-dominant role exists, but by PGY3, residents are expected to be much more like colleagues than students.[5] We find that these formal expectations are mirrored in attendings' comments. Of the 174 comments for PGY1s that contain an explicit reference group, 98.9 percent reference residents' roles as students or learners (this is 11.9 percent of the total comments PGY1s receive [172 out of 1,448]); only 1.1 percent reference attendinghood or independent practice (which is .14 percent of the comments PGY1s receive [2 out of 1,448]). In these comments, attendings compare residents' abilities to those of their fellow PGY1s, writing that residents are "on par with [their] peers" or have performed at a level "appropriate to [their] stage of training." Thus, PGY1 residents are not being assessed

**Table 1.** Role Expectations for Emergency Medicine Residents at University Hospital

| Post-Graduate Year (PGY) 1 | Post-Graduate Year (PGY) 3 |
| --- | --- |
| Learner[a] | Leader[b] |
| Cautious[a] | Confident[b] |
| Seeks assistance and feedback[a] | Independent/autonomous[b] |
| Masters basic skills; recognizes limits | Masters advanced medical knowledge |
| Successfully performs simple procedures | Successfully performs complex procedures |
| Manages a single patient | Oversees the whole ED |
| Reference group = student[a] | Reference group = colleague[b] |

*Note:* Characteristics are derived from a qualitative content analysis of resident performance evaluations at University Hospital. Some map directly onto the ACGME's 23 EM milestones.
[a]Indicates characteristics/roles identified as stereotypically feminine by prior literature (DiPrete and Buchmann 2013; Downey and Vogt Yuan 2005; Jacob 2002; Morris 2011).
[b]Indicates characteristics/roles identified as stereotypically masculine by prior literature (Gorman 2005; Mueller et al. 2017).

for their overall aptitude as *physicians*; rather, they are being assessed relative to a particular and specifically *early* stage of medical training, where learning new things rather than being perfect at them dominates.

That PGY1s are being assessed for their ability to fill the student role is also evident in the characteristics attending physicians value in their residents. Table 1 displays the characteristics attending physicians remark on in PGY1 and PGY3. In PGY1, attendings value residents who seek assistance and feedback, are eager learners, and perform medical tasks cautiously, traits that previous literature connects to both the student role and traditional femininity (DiPrete and Buchmann 2013; Downey and Vogt Yuan 2005; Jacob 2002). This is evident in comments for men and women PGY1s. For example, Lisa, an attending, praised Eugene, a PGY1, for his reliance on his attendings: "Enthusiastic. Asked for assistance when unsure what to do (appropriately so). Actively asked for feedback in order to improve." Similarly, Scott, an attending, wrote about Juliet, a PGY1: "Juliet is off to a great start. Knowledge base is good for her level of training. She has a great attitude and is easy to work with. Readily accepts feedback and works it into her practice." These typical comments communicate a set of character ideals closely linked to the role of student. As we will show, they are starkly different than those used to assess PGY3s.

By PGY3, attendings' expectations of resident performance undergo significant changes, and this is apparent in the comments directed at PGY3s by attendings. In PGY3, 169 comments indicate a reference group explicitly; this is about the same percentage (13 percent) of comments as in PGY1. Of these, 43.2 percent reference the student role (compared to 98.9 percent for PGY1s), and 56.8 percent reference attendinghood or career stages after residency (compared to 1.1 percent for PGY1s). The comments explicitly referencing attendinghood account for 7.4 percent of the comments made to PGY3 residents. This kind of comment contains language indicating PGY3 residents were "ready to work on [their] own" or they would "perform well as junior faculty." The rest of the comments reference the student role. Although the student role is still present in PGY3, the role of professional physician, or independent practitioner, increased in saliency from PGY1.

The fact that PGY3s are assessed for their ability to fill the colleague role is also evident in the characteristics attendings praise in residents at this stage. A summary of these can be found in Table 1. Residents in their final year of training are valued for their independence, confidence, and mastery of advanced medical skills (Mueller et al. 2017), traits that previous literature has linked to traditional masculinity (Gorman 2005). PGY3s receive less praise for deference to their superordinates and are instead leaned on to act as leaders themselves.

This change in character ideals is encapsulated in the following comment written by Brian, an attending, for Lydia, a PGY3: "Leader and role model for her fellow emergency medicine residents. Performs appropriate clinical application of knowledge in practice. Fully participates in the educational, safety, and quality improvement missions of the department. Always a pleasure to work with Lydia." This emphasis on leadership was common in PGY3s' performance evaluations: 230 comments, 17.5 percent of the 1,317 total comments for PGY3s, assessed residents' leadership skills (compared with just 2.6 percent of comments for PGY1s), suggesting this characteristic is considered less important to the PGY1 role. PGY3 residents who master these skills are believed to be well-prepared for their next career stage, as is suggested in the following comment written by Frank, an attending, about Keith, a PGY3: "Keith is an excellent resident, and he works at the level of junior faculty. He manages his patients, assists the junior residents, manages the department, and still takes the time to spend with patients and families." Embodying the ideal characteristics of a PGY3 also means embodying those of an attending physician, and residents at this stage are judged for their ability to succeed in independent practice. This is quite different than the role of PGY1s, who are expected to be good at learning.

## Gender Differences in Perceived Ability to Meet Role Expectations

Having established the shift in reference groups between PGY1 and PGY3, we now analyze how gender shapes residents' abilities to meet expectations under these different circumstances. We begin by examining these patterns in comments where the reference group is explicitly mentioned, and we then examine more general trends in residents' abilities to meet expectations.

*Analyzing comments with explicit reference groups.* In the subset of comments that explicitly assess PGY1 residents

for their ability to fill the student role ($N = 172$), more comments for women than men *positively* compare them to the "student" reference group. Table 2 shows that a larger percentage of comments for women state they are ahead of their classmates (84.3 percent, compared with 62.8 percent for men). Of these positive comparisons to the reference group, a larger subset of comments for women (18.6 percent, compared with 10.5 percent for men) explicitly state they are the best resident in the class (not shown in the table). For instance, John, an attending, wrote about Faith, a PGY1, "Faith continues to make progress and is developing rapidly. She is able to increase her patient load, include sicker[6] patients [patients at risk of dying] and not lose quality. She continues to perform among the best residents in her class." In contrast, more comments for men than for women in PGY1 contain *negative* comparisons to the "student" reference group: a higher percentage of comments received by men than women stated they were either on par with their classmates (14.9 percent, compared to 7.8 percent for women) or behind the curve (19 percent, compared to 7.8 percent for women). A typical negative assessment of men residents' ability to fill the student role appeared in the following comment written by Michael, an attending, for Gavin, a PGY1:

> I feel like we really had some disconnects on this shift and the previous shift we worked: as we discussed in person, I got the feeling you weren't interested in completing a complete exam. When I suggested things like pelvic exams/rectal exams I sensed that you didn't feel these were necessary and I felt you were almost argumentative/defensive. As a brand-new resident, I would suggest that you do your best to make your faculty feel as though you were interested in learning from them.

Michael directly linked Gavin's apparent lack of interest in learning to his failing to appropriately meet the role expectations of a "brand new resident" or student. This comment is part of a pattern wherein attendings find that

**Table 2**. Gender and Perceived Ability to Meet Role Expectations

| Code | Men | | Women | |
|---|---|---|---|---|
| | *N* | % | *N* | % |
| *PGY1* | | | | |
| Reference to Student Role | 121 | | 51 | |
| Behind | 23 | 19.0 | 4 | 7.8 |
| On par | 18 | 14.9 | 4 | 7.8 |
| Ahead | 76 | 62.8 | 43 | 84.3 |
| Total Comments | 1,061 | | 387 | |
| Residents | 27 | | 8 | |
| *PGY3* | | | | |
| Reference to Colleague Role | 65 | | 31 | |
| Not ready | 4 | 6.2 | 13 | 41.9 |
| Almost Ready | 9 | 13.8 | 5 | 16.1 |
| Ready | 53 | 81.5 | 13 | 41.9 |
| Total Comments | 804 | | 513 | |
| Residents | 22 | | 14 | |

men PGY1s' performances in the ER compare negatively to the traits expected of them at this stage. In general, women residents appear to better exemplify the ideals of the PGY1 learner role.

The trend of an apparent advantage for women and disadvantage for men in PGY1 reverses in PGY3. In their third and final year of training before independent practice, men residents received more positive comparisons to the reference group than did women residents in comments with explicit reference groups. Of the comments evaluating residents' abilities to perform the attending role, only 58 percent of comments directed at women residents contained a positive assessment (that they were either ready or almost ready for independent practice), compared to 95.3 percent of comments directed at men. A larger percentage of these comments for men (81.5 percent) than for women (41.9 percent) stated the resident was ready for independent practice or post-residency career stages. The following evaluation written by Howard, an attending, for Aaron, a PGY3, shows a typical example of a positive comparison to the "colleague" reference for men:

Aaron did an excellent job as usual with succinct presentations, appropriate workups and thorough differentials. He correctly

immediately stained the cornea on a young female patient with Bell's [palsy] and oral herpes, and helped out an off service intern with a ketamine sedation on a 3yo with a cheek lac[eration]. [The critical care area] was quite busy during the entire shift and the [lower acuity] patients never suffered. *I'm sure I've said this before but Aaron impresses me with his attitude and professionalism every shift and I'd be happy to have him as a partner. He's ready for solo practice without a doubt.* [emphasis added]

Howard took Aaron's successful performance of medical procedures, in addition to his team leadership and personality, as evidence that he was not only ready to practice EM independently, but that he would make a good colleague at University Hospital, an academic medical center. Several other men (5 of 22 [22.7 percent] men PGY3s) received similar comments, suggesting they would be good fits for careers in academic medicine, which is typically more prestigious than other types of medical practice. This is particularly noteworthy when compared to comments made to PGY3 women.

Some women were told they were ready for independent practice (5 out of 14 women in PGY3, or 35.7 percent), but *none* of the

women received a comment suggesting they should try for a career in academic medicine (whether at University Hospital or elsewhere). Additionally, more comments for women than for men contained negative comparisons to the "colleague" reference group. Of the subset of comments that assessed PGY3s' ability to fill the role of attending physician, 41.9 percent suggested women residents were presently unable to do so, compared with 6.2 percent of similar assessments of men residents. Within these negative comparisons to the role of colleague, attendings often commented on deficiencies in knowledge, skills, or personality that would prevent women from being solid EM doctors after residency. For example, Peter, an attending, told Nicole, a PGY3, that she was overly meticulous:

> Did a good job trying to tease out potentially sick patients who were otherwise well-appearing. Pays attention to detail. Hard working. Clearly cares about her profession, and it shows. The *danger when being so meticulous* is possibly over-evaluating patients who can easily be managed in the outpatient setting. You should *work hard on your efficiency this year.* In the community, you will be expected to see about 1.75 [patients]/hr minimum, and expected to keep your head afloat during the periods that balloon to 3 or 4 [patients]/hour. Challenge yourself now, I think you'll find you can manage it easily once you push yourself. [emphasis added]

In earlier stages of residency, "attention to detail" is often praised as an attribute. But when considering Nicole's post-residency career, Peter instead construed this trait as a potential impediment to her success. Whereas comments for men residents focused on the constellation of skills that would make them excellent independent practitioners, comments for women residents, like this comment for Nicole, instead focused on the work they still had to do in order to be successful colleagues. Evaluations for PGY3s more often construed men as ready for attendinghood and women as not yet ready for this superordinate role.

*Analyzing more general trends.* Our analysis of comments with explicit mentions of reference groups suggests these reference groups matter to understanding gender inequality in PGY3, but they account for a small percentage of the total comments residents receive. Do the gendered trends we observed in the previous section appear more generally in the comments residents received? To examine this question, we now expand our analysis to all comments residents received in PGY1 and PGY3. Indeed, we see similar trends to those revealed by our analysis of comments with explicit reference groups. Table 3 presents these findings.

In PGY1, compared to men residents, women residents received more feedback that was purely positive in tone and content and less feedback that was purely negative. Specifically, 67.4 percent of feedback to women PGY1 residents was positive, compared to 57.5 percent for men residents, and only 5.9 percent of feedback for women residents was negative, compared to 13.3 percent for men residents. Additionally, 28.9 percent of comments directed at women residents indicated they performed exceptionally well (i.e., a very strongly positive comment [see Table 3]), compared to only 18.9 percent of comments directed at men. As mentioned earlier, performance as leaders was not heavily commented on in PGY1, but interestingly, when it was, men and women received similar percentages of positive comments. The same is the case for residents' communication skills (with patients, nurses, consultants, and family members): in PGY1, men and women received relatively similar rates of positive feedback. Thus, not everything is skewed to advantage women in PGY1. This makes sense, as past research has not found a statistically significant advantage for women in PGY1 (Dayal et al. 2017).

By PGY3, these patterns shift, revealing an advantage for men residents. Interestingly, in PGY3, men and women received fairly equal percentages of positive (73.5 and 69 percent, respectively) and negative (6 and 6.4 percent, respectively) comments; however, women received fewer comments indicating

**Table 3**. Gender and General Trends

| | PGY1 | | | | PGY3 | | | |
| | Men | | Women | | Men | | Women | |
| Code | N | % | N | % | N | % | N | % |
|---|---|---|---|---|---|---|---|---|
| Positive | 610 | 57.5 | 261 | 67.4 | 591 | 73.5 | 354 | 69.0 |
| Exceptionally Positive | 201 | 18.9 | 112 | 28.9 | 290 | 36.1 | 141 | 27.5 |
| Negative | 141 | 13.3 | 23 | 5.9 | 48 | 6.0 | 33 | 6.4 |
| Exceptionally Negative | 77 | 7.3 | 21 | 5.4 | 34 | 4.2 | 30 | 5.8 |
| Leadership + Positive | 19 | 1.8 | 3 | .8 | 128 | 15.9 | 50 | 9.7 |
| Communication + Positive | 132 | 12.4 | 57 | 14.7 | 208 | 25.9 | 97 | 18.9 |
| Total Comments | 1,061 | | 387 | | 804 | | 513 | |

that they performed exceptionally well (i.e., extremely positive) than men (27.5 versus 36.1 percent). Additionally, in PGY3, men received more positive feedback about their ability to communicate with others in the ED and their ability to lead the ED than did women. Specifically, of comments directed at men, 25.9 percent noted a good job communicating and 15.9 percent noted a good job with leadership, compared to 18.9 and 9.7 percent, respectively, of comments directed at women. These trends differ from what we saw in PGY1 and thus provide further evidence that women have a harder time than men fulfilling the role expectations for PGY3s than for PGY1s.

## Analyzing Gender Differences in Reactions to Medical Errors

In the previous sections, we provided evidence that men have an advantage over women as the reference group for performance evaluations shifts from student to colleague. However, thus far, we have not been able to account for the possibility of *real* differences in men's and women's performance. To address this, we conclude our analysis with an examination of gender differences in evaluations of 410 comments that concern medical errors, while holding performance somewhat constant by comparing gender differences within PGY and within error severity categories.

In PGY1, a relatively similar percentage of comments directed at men (15.9 percent) and women (11.4 percent) concerned medical errors (see Table 4). Additionally, of the errors men and women residents made, there was a relatively equal breakdown across genders by the severity of the error. The majority of errors (about 84 percent of errors for both genders in PGY1) were minor, where a patient's well-being was not in danger; about 16 percent of errors for both genders were major, where a patient's well-being had the potential to be seriously harmed. By analyzing qualitative differences in how attending physicians reacted to errors, we are able to see more nuance in how gender matters to evaluations. Table 4 shows that in PGY1, feedback for women who made major errors contained less reassurance than did feedback for their men peers (0 compared to 10.7 percent), more language blaming them for the error (57.1 percent compared to 39.3 percent), and more harsh or hurtful language around the error (28.6 percent compared to 17.9 percent). To illustrate this difference, we compare the cases of Spencer and Megan, the two PGY1 residents who made the largest number of major medical errors in PGY1. Lisa, an attending, wrote the following evaluation of Spencer's work:

> *Quick to see patients. Got consultant on board quickly*, but remember to also look at patient carefully to determine severity of

**Table 4**. Attendings' Reactions to Medical Errors by Error Severity

| | PGY1 | | | | PGY3 | | | |
| | Men | | Women | | Men | | Women | |
| Code | N | % | N | % | N | % | N | % |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| *N* Medical Errors | 169 | 15.9 | 44 | 11.4 | 108 | 13.4 | 89 | 17.3 |
| Minor Errors | 141 | 83.4 | 37 | 84.1 | 89 | 82.4 | 65 | 73.0 |
|   Benefit of the doubt | 29 | 20.6 | 2 | 5.4 | 7 | 7.9 | 6 | 9.2 |
|   Reassured | 40 | 28.4 | 13 | 35.1 | 11 | 12.4 | 11 | 16.9 |
|   Hurtful | 8 | 5.7 | 0 | 0 | 3 | 3.4 | 4 | 6.2 |
|   Blamed | 32 | 22.7 | 5 | 13.5 | 22 | 24.7 | 26 | 40.0 |
| Major Errors | 28 | 16.6 | 7 | 15.9 | 19 | 17.6 | 24 | 27.0 |
|   Benefit of the doubt | 3 | 10.7 | 1 | 14.3 | 3 | 15.8 | 1 | 4.2 |
|   Reassured | 3 | 10.7 | 0 | 0 | 3 | 15.8 | 2 | 8.3 |
|   Hurtful | 5 | 17.9 | 2 | 28.6 | 5 | 26.3 | 11 | 45.8 |
|   Blamed | 11 | 39.3 | 4 | 57.1 | 11 | 57.9 | 17 | 70.8 |
| Total Comments | 1,061 | | 387 | | 804 | | 513 | |
| Total Residents | 27 | | 8 | | 22 | | 14 | |

illness & need for resuscitation (e.g., transfer patient with mesenteric ischemia & perforated bowel who was cool & clammy needed [intravenous fluids])—[you] want to expedite entering these orders & getting resuscitation started (probably even before calling consultant—or at least simultaneously). . . . Don't hesitate to ask the upper level resident or attending if you have problems (e.g., trying to order [antibiotics]). [emphasis added]

Here, Spencer failed to recognize that a patient needed urgent fluid resuscitation, a potentially life-saving procedure whose delay could have seriously harmed the patient. Despite pointing out this major error in medical judgment, Lisa reassured Spencer of his abilities, complimenting his enthusiasm to see new patients and the speed at which he called in a specialist on this patient's complex case. Men received slightly more comments like this one than did women in PGY1.

Women in PGY1 received less reassurance and instead were given a larger number of evaluations that assigned them blame for errors or that contained hurtful commentary about their ability to perform in the ED. The following comment written by Brian, an attending, for Megan, a PGY1, is characteristic of these kinds of comments:

Needs to demonstrate substantial improvement in efficiency to be considered prepared for 2nd year. Knowledge deficits below the level of her peers, resulting in extraneous "shotgun" lab testing requiring significant supervision. Knowledge deficits for management of some simple ambulatory complaints requiring more than the average level of supervision, with some difficulty utilizing practice improvement resources for on-shift education to augment education and prevent errors (e.g., difficulty differentiating cervicitis from [pelvic inflammatory disease], difficulty identifying correct antibiotics [therapy], despite multiple prompts to check online reference material). *Does not currently appear to be on track for the level of independence and capability necessary for 2nd year.* [emphasis added]

Here, Brian described several medical errors made by Megan, with the most concerning being her inability to correctly recognize a serious illness and identify the appropriate

treatment. Brian communicated that Megan was at fault for these issues when he cited her knowledge deficits as the root cause of the errors and noted that despite the availability of educational resources that might help her improve, she failed to use these effectively. The comment was also hurtful in that Brian called into question Megan's ability to move on to her next year of residency.

Comments about major errors for women PGY1s were less supportive and more critical than those for men PGY1s, but this pattern does not hold for minor errors. There is no clear pattern of advantage for either men or women: men received more benefit of the doubt (present in 20.6 percent of comments about men's minor errors and in 5.4 percent of women's), more hurtful comments (5.7 percent of minor error comments compared with 0 for women), *and* more blame for the error (22.7 percent of minor error comments for men compared with 13.5 percent for women). Women, on the other hand, received slightly more reassurance than men (35.1 percent compared with 28.4 percent).

Together, our analyses of medical errors in PGY1 do not suggest that the role expectations for PGY1s clearly advantage either men or women. The one exception is when PGY1s make major mistakes: attendings react more harshly to women residents than to men residents. However, attendings still tend to contextualize these major mistakes within the student-doctor role for both men and women residents, as can be seen in the earlier comments for Spencer and Megan. Every PGY1 resident received at least one comment about a medical error, and most received multiple. If we look at the data by resident, men received a slightly higher number of comments about errors (median of six, compared with a median of four for women), and 12 of 26 (46.2 percent) men PGY1s, compared with 2 of 8 (25 percent) women PGY1s, had at least one comment about a major error. Most residents were still told they were doing a great job in their PGY1 role despite these errors, and 19.2 percent of men (5 of 26 men PGY1s) and 50 percent of women (4 of 8

women PGY1s) were told they were the best resident in their class despite having made errors. For two women and one man, this included major errors. These patterns do not suggest a clear advantage for either men or women.

In PGY3, we see important gender differences in both the prevalence of errors by gender and in attendings' reactions to errors. First, as Table 4 reveals, in PGY3, women received slightly more comments about errors (17.3 percent) than did men (13.4 percent), and of the errors noted, a slightly larger proportion of comments directed to women concerned major errors (27 percent compared to 17.6 percent for men). This may suggest women were performing somewhat worse than men during PGY3 in a way we did not observe in PGY1. For that reason, it is important to analyze how attendings responded to equally serious errors to determine whether these comments are most likely indicative of real skill deficits or whether they may suggest implicit bias on the part of attendings.

First, Table 4 shows qualitative differences in the feedback men and women received for minor and major errors. Attendings reacted to women's *minor* medical errors slightly more harshly (e.g., 40.0 percent of minor error comments received by women blamed the resident, compared with 24.7 percent of similar comments for men), but they also reassured men and women residents or made hurtful comments at relatively equal rates. However, the gender difference in reactions to men's and women's *major* medical errors is more striking. Comments for men residents contained more supportive feedback when they made major medical errors than did comments for their women peers: 15.8 percent of major error comments for men contained language extending the benefit of the doubt to the resident, compared with 4.2 percent of comments for women residents, and 15.8 percent contained reassurance, compared with 8.3 percent for women residents. Men residents also received less hurtful feedback for major errors than did their women peers: 26.3 percent of comments on major errors for

men contained hurtful feedback, compared with 45.8 percent for women residents, and 57.9 percent of comments for men's major errors communicated that blame for the error rested with the resident (and not the attending or both of them), compared with 70.8 percent of similar comments for women.

We illustrate the gender difference in attendings' reactions to major errors by comparing the cases of Patricia and Graham, two residents who made the largest number of major errors in PGY3 and were among the lowest numerically ranked residents in their cohorts. Both residents received comments from their attendings about major errors that had the potential to seriously harm the patients under their care. First, Victoria, an attending, wrote the following comment for Patricia:

> Had a febrile, tachycardic anxious spina bifida teenager recently who was getting a septic workup recently. [Patricia] had taken out the [lumbar puncture] tray and was preparing to the procedure on an agitated patient in whom she had not looked for another source. He smelled like urosepsis so I told her to get cultures and give [antibiotics], but get the [urinalysis] back first before we "discuss" the need for [lumbar puncture]. He ended up having urosepsis and his mental status cleared when his fever went down and he got hydrated. *She should not be a senior resident.* [emphasis added]

Here, Patricia failed to recognize that her patient had urosepsis, a dangerous complication of a urinary tract infection that requires urgent medical attention. When Victoria pointed out symptoms of urosepsis to Patricia and gave her explicit instructions on how to proceed with patient care, Patricia disregarded her comments and planned to perform a lumbar puncture, a more complex and riskier procedure than was called for, and one that would not help cure the patient's urosepsis. Victoria concluded by declaring that Patricia was incapable of performing her occupational role.

In contrast, Graham, who received a comment about a medical error of similar severity

to Patricia's (if not more severe, as it had the potential to result in a fatality) got much more reassurance from his attending, Harrison:

> I enjoy working with Graham, he is very social and easy to get along with. However, as a third year, I continue to be disappointed. *I think he has the fundamental knowledge and basic clinical skills.* He relates well to patient and staff. However, I do not find him clearly focused on the patient care, or clearly focused on effective teaching, or clearly focused on departmental management. *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 discussing differences). I did not find him, after requested, clearly helping a medicine intern suture a large superficial wound (he did get it started, but was absent for most of the procedure), I did express my concern, he had in his hand an EKG [electrocardiogram] with reading of 'STEMI' which I did not receive for an hour and a half after the time on the EKG (I do not know how long he had it). [emphasis added]

Harrison listed a number of problems he had noticed with Graham, including a failure to follow explicit instructions from his attending, something Victoria also noted about Patricia in the earlier comment. The most serious error, however, was that Graham missed an electrocardiogram (EKG) reading that indicated the patient was having a STEMI, a severe type of heart attack, which should be recognized and acted upon as fast as possible on arrival in the emergency department. Yet, Harrison reaffirmed Graham's capabilities as an EM physician.

These errors were similarly serious, and yet, the attendings' reactions were substantially different. When Patricia made a major error, Victoria called into question her basic competence as an emergency room physician and explicitly linked Patricia's mistake with her inability to fill her role as a senior resident. In contrast, when Graham made a major

error, Harrison stated that he still believed in Graham's overall abilities as a physician.

These comments about Graham and Patricia are part of a broader pattern in which attendings gave men, but not women, the benefit of the doubt in PGY3. We found that men residents who received a large number of comments about errors also received comments telling them they would make excellent EM doctors; women, however, did not. Instead, women with multiple comments about errors received comments that indicated they would struggle in independent practice. Specifically, of the five PGY3 men who received 10 or more critical comments about medical errors, four were told they were ready for independent practice or would make good EM doctors. Only one was told he might struggle in independent practice. This suggests attendings were able to see men as competent EM physicians despite their medical errors.

Of the three women who received 10 or more critical comments about medical errors, *none* were told they were ready for independent practice or would make a good doctor. Two of these three women were instead told by their attendings that they would struggle with independent practice. Furthermore, *none* of the 13 women from the PGY3 cohort, even those who made few or no errors, were told they would be a good fit in academic medicine, but four men who made 10 or more errors were told this.

To summarize, we present evidence that as the ACGME milestones for residents shift from a student-dominant role to a colleague-dominant role, attendings' perceptions of residents' performance also shift. In the first year, where the student role dominates, women and men generally experience similar feedback from their attendings, even when they make serious medical errors (which they make at relatively equal rates). Additionally, more women residents than men residents are praised for being ahead of their peers. However, by residents' third and final year, these patterns shift such that women residents receive significantly fewer comments indicating their readiness for independent practice or

that they are talented enough to make it in academic medicine, they are noted to be lagging behind their men counterparts, and most saliently, they receive much harsher reactions to medical errors than do men residents. When women make mistakes at this stage, it is seen as discrediting their ability to practice emergency medicine, whereas men can have many life-threatening medical errors on their record and still be seen as ready to launch their careers in academic medicine.

## DISCUSSION

Why do women struggle to be seen as capable in the workplace despite outperforming men during school? This article engages with role expectations theory to investigate the production of gender inequality in the school-to-work transition. To do this, we leverage the case of medical residency, the training period during which aspiring physicians transition out of the role of student and into the role of independent practitioner, which allows us to hold organization and occupation constant as we examine the emergence of inequality. Drawing on a dataset of in-the-moment performance evaluations for emergency medicine residents, we find that the reference group for resident performance shifts explicitly from "student" in year one (PGY1) to "colleague" in year three (PGY3), and this shift appears to trigger bias against women. This pattern is particularly obvious in a subset of comments about medical errors: although there is little difference in how all errors are assessed in PGY1 and minor errors are assessed in PGY3, women who make major errors in PGY3 receive feedback that is less supportive and more critical of their ability to fill the role of EM doctor than do men who make equally severe errors.

Our findings suggest the attending bias against women residents is not constant throughout residency, but rather becomes salient depending on organizational roles. Specifically, bias against women and in favor of men appears most extreme when residents are being assessed for their ability to perform

as colleagues and when they exhibit behaviors (major medical errors) that are especially disconfirming of their ability to succeed in this role. Based on this, we argue that the structure and content of interaction can shape the activation of gender bias.

Our study has three primary implications for gender inequality in work and schools. First, we contribute to the sociological literature on gender inequality in the school-to-work transition by demonstrating how changes in role expectations that come with this transition may trigger implicit bias against women. We build on previous literature that demonstrates role expectations benefit women in school because the student role reflects many aspects of traditional femininity (DiPrete and Buchmann 2013; Downey and Vogt Yuan 2005; Jacob 2002; Morris 2011), but they serve as a barrier to women's advancement in the workplace because higher-status occupational roles typically reflect aspects of traditional masculinity (Acker 1990; Kanter 1977; Ridgeway 2009; Ridgeway and Smith-Lovin 1999). This literature suggests the emergence of new role expectations disadvantages women as they transition from school to work, but it is challenging to find an empirical case that allows processes related only to changes in role expectations to be isolated, given that the organizational context is usually also changing as individuals transition from schools to workplaces. The multitude of changes that take place as individuals move from purely educational contexts to purely workplace contexts may obscure the extent to which role expectations inform bias against women.

By focusing on medical residency, we are able to isolate role expectations and compare assessments of men and women within a single organization and a single occupation as they occupy different organizational roles at two distinct career stages. We show that attendings do in fact assess women as poorer fits for emergency medicine when their organizational role is that of a colleague, which occurs in PGY3, compared to when their organization role is that of a student, which occurs in PGY1. This suggests role expectations, net of performance, organizational context, or occupation, shape gender inequalities in assessment. This finding may be relevant to understanding the production of gender inequality in the school-to-work transition across a number of fields, especially male-dominated ones like science, academia, and law, where these inequalities are most prevalent (Jacobs 1996; Kay and Gorman 2008; Xie and Shauman 2003). Women who prove their competency at the requisite skills for these fields through success in career training may nonetheless be assessed as less capable in the workplace in part because of gendered role expectations. This may occur even when the school and work contexts require the exact same set of skills, as is the case in a number of professions with long credentialing processes (Cech et al. 2011).

Second, our findings also shed some light on women's attrition from male-dominated fields. Across many of these fields, including higher-status career tracks in medicine, engineering, and law, the highest rates of women's attrition take place immediately after the acquisition of credentials that allow individuals to be full-practitioners of their chosen profession (i.e., after residency in medicine, undergraduate education in engineering, and law school in law) (Cech et al. 2011; Dayal et al. 2017; Kay and Gorman 2008). This pattern points to dynamics taking place within educational contexts that encourage women to persist through lengthy credential-acquisition processes but discourage them from pursuing high-status careers after graduation. Our study provides some insight into this phenomenon. When women were in the learner role, they were often praised for their abilities and told they were among the best residents in the class. Based on attending feedback, women at this stage might be optimistic about their career success. But even though high-performing PGY1 women were often told they were among the best residents in the class, none of the women PGY3s, even the best performers, were told they should pursue academic medicine. At the same time, men,

even lower performers, were told they should pursue these higher-status careers. These differences in feedback from attendings might inform women's underrepresentation in academic medicine (Bennett et al. 2019), especially because residents typically make plans for their next career stages during their final year of training. Moving forward, there is need for longitudinal research that directly connects the feedback aspiring physicians receive in residency to these gender disparities in career trajectories after residency in order to understand the long-term effect of biased evaluations. This research would be useful in shedding light on the relationship between biased feedback and unequal outcomes in the workforce more generally.

Finally, this article also contributes to an ongoing discussion within sociology of work and gender on how to make performance evaluations and similar assessment tools "fair." Some research indicates performance evaluations may contain less gender bias if objective standards—that are formal, clearly-articulated, and detailed—are in place (Cecchi-Dimeglio 2017; Heilman 2001) and if evaluations are written closer to the time an employee's performance is observed (Cecchi-Dimeglio 2017). Formal, clearly-articulated performance standards might counteract discrimination that comes from "shifting standards" of evaluation, whereby men and women are assessed based on different criteria (Fuegen 2007). However, other studies call this notion into question. Seemingly-neutral formal evaluative standards may actually *reinforce* bias against women if the standards themselves reflect conventional masculinity in some way (Acker 1990; Gorman 2005). Additionally, evaluations written in real-time may be rushed, especially in fast-paced and high-stakes workplaces like medicine. Social psychological research on implicit bias indicates stereotypes are more likely to be revealed under conditions like these in which individuals have less time to reflect on their decisions (Fiske 1998).

Our findings provide support for this second school of thought. We show that in the case of emergency medical residency, evaluators bring gender biases into their implementation of formal standards of evaluation. As residents approach graduation, attendings increasingly evaluate women as less capable than men of filling the role of independent practitioner, net of actual performance. Thus, our study suggests organizations must be cautious in how they interpret performance evaluation data. Supposedly-objective evaluative criteria do not necessarily produce fair and egalitarian assessments, and relying on them to do so may exacerbate gender inequalities in the workplace. For example, the ACGME is currently considering moving toward a competency-based system for graduation, wherein a resident's ability to graduate is not based on progressing though a multi-year program and scoring adequately on exams (which is the current system), but rather on attaining ACGME milestones (Iobst et al. 2010; Ten Cate 2017). Even though women receive lower scores on milestone attainment (Dayal et al. 2017), our study provides evidence that women are not performing worse than men: when we hold performance constant via the severity of medical errors, we see pretty significant differences in men's and women's feedback and evaluations that are more suggestive of bias on the part of attendings than real performance differentials between men and women. Thus, biased assessments by attendings could mean the competency-based graduation system will serve as a barrier to women's ability to become fully-licensed practitioners, exacerbating gender inequality in the medical profession.

## Limitations and Future Directions

There are several limitations to this study. To start, the empirical generalizability of our findings is limited by the fact that University Hospital is just one among several hundred emergency medicine residency programs. As with any single organization, it may not represent the rest. That said, previously published quantitative findings using the same data (although with eight hospitals [including University] rather

than just University Hospital) (Dayal et al. 2017) allow us to be more confident that the gender differences in evaluations we observe are robust, as they appeared in the quantitative analysis across all the hospitals. This suggests the local culture of the hospital likely does not explain our observations.

Furthermore, we are uncertain of the empirical generalizability of our findings outside of the field of medicine. Medicine is different from other careers in that it has an especially long training process, a unique professional hierarchy, and may be higher-stakes and faster-paced than other fields. Even so, we believe this study can inform how we think about the production of gender inequality in the student-to-professional transition, especially in male-dominated fields where the colleague role is tied to stereotypically masculine expectations. Career aspirants in law, science, and academia go through similar transitions between role expectations and similar gender inequalities emerge as they do so (Jacobs 1996; Kay and Gorman 2008; Xie and Shauman 2003).

We are unable to account for additional ways, beyond the numerical scores and content of textual comments, that gender may have shaped attending feedback for residents. For example, attendings may have opted to give men negative feedback in-person rather than in textual comments to "shield" them from the potential negative ramifications of having this kind of feedback in their permanent record. That said, we found no statistically significant difference in the number of comments written for men and women residents, suggesting attendings did not prefer to give residents of one gender feedback in person and the other using the app. Additionally, although the numerical scores of evaluations with textual comments were generally lower than those without comments, we did not find that this differed by gender. It therefore appears unlikely that attendings shielded either men or women preferentially from the consequences of having negative feedback in written comments (that could be seen by other

attendings, program directors, and superordinates) by commenting on mistakes in person. Moreover, comments written using InstantEval V2.0 may hold more weight than in-person feedback in terms of formal resident performance evaluations, because while the content of in-person feedback might remain confidential, written feedback for residents can be read by the residency program director and other attendings involved in evaluating resident performance. Written comments are attached to residents' permanent records and have the potential to shape their reputation among the medical faculty as a whole.

Finally, our dataset does not contain information about race and other potentially relevant demographic characteristics that may have shaped how attendings evaluated their residents. Organizational structures often function in ways that reproduce racial inequality (Ray 2019), and medicine is likely no different. Thus, while our study points to the importance of organizational culture and interactional context contributing to gender bias in the evaluation of trainees and employees, further research is needed to elaborate the salience of these findings across organizational contexts and to explore the importance of gender as it intersects with other primary social identities.

## Conclusion

Women's continued inequality in the professional workforce is particularly puzzling given their high levels of success within educational contexts. We contribute an explanation of how role expectations inform the emergence of gender inequality in the school-to-work transition through a case study of medical residency. Drawing on a dataset of numerical and textual performance evaluations for emergency medicine residents, we empirically demonstrate that as residents go from being evaluated as students to colleagues, men come to be seen as better fits for their role. By focusing on a subset of comments about medical errors in which we

control for error severity, we demonstrate that attendings' implicit bias against women residents in their final year of training shapes the content of evaluations. Although there is little difference in how attendings treat men and women who make errors in their first year of residency, major differences emerge in the third and final year, especially when it comes to the most serious medical errors. These findings show how role expectations produce emergent gender bias as individuals progress from student to professional, providing an explanation for why women are evaluated as less capable than men in workplaces despite outperforming them in school.

## ORCID iDs

Alexandra Brewer (iD) https://orcid.org/0000-0003 -1910-2739
Daniel M. O'Connor (iD) https://orcid.org/0000-0001 -5464-2031

## Notes

1. Medical residency training programs typically range from three to five years.
2. Women's attrition from EM does not typically represent attrition from the medical profession itself (Ginde, Sullivan, and Camargo 2010), but rather their segregation into community medicine, which is typically lower status, resulting in a disproportionate number of men EM MDs in higher-status positions within academic medicine.

3. An additional function of the InstantEval V2.0 tool is that it enables residency programs to collect standardized data on residents' performances to report to the ACGME. The ACGME requires milestone data for each resident to be reported every six months.
4. ACGME emergency medical subcompetencies include emergency stabilization, focused history and physician examination, diagnostic studies, diagnosis, pharmacotherapy, observation and reassessment, disposition, multitasking/task-switching, general approach to procedures, airway management, anesthesia and acute pain management, goal-directed focused ultrasonography, wound management, vascular access, medical knowledge, patient safety, systems-based management, technology, practice-based performance improvement, professional values, accountability, patient-centered communication, and team management.
5. We focus on a comparison of PGY1 and PGY3 because these are the stages when role expectations for resident performance—and gender inequalities in evaluation—are most distinct, according to the quantitative analysis in Dayal and colleagues (2017), as well as our qualitative analysis. PGY2 represents a middle space between a student-dominant and colleague-dominant reference group. For example, we find that out of 218 comments for PGY2s that contain an explicit reference group, 208 (90.4 percent) reference the student role (compared to 98.9 percent in PGY1 and 43.2 percent in PGY3) and 10 (4.7 percent) reference the colleague role (compared to 1.1 percent in PGY1 and 56.8 percent in PGY3). Traits like leadership that are valued in PGY3, but not PGY1, become more salient in PGY2: 5.9 percent of PGY2s' total comments evaluate leadership skills, compared with 1.4 percent of comments for PGY1s and 9 percent of comments for PGY3s. Similarly, PGY2 represents an in-between stage for gender inequality. For example, a higher percentage of comments assessing women's role-fit contain negative commentary in PGY2 than in PGY1, but not as many as in PGY3.
6. In the context of the ED, "sick" patients indicate those who are dying, may be dying, or are experiencing significant threats to their lives. "Not sick" refers to patients who may need emergent care but whose lives are not immediately at risk.

## References

Accreditation Council for Graduate Medical Education and American Board of Emergency Medicine (ACGME and ABEM). 2015. "The Emergency Medicine Milestone Project." Retrieved January 25, 2018 (https://www.acgme.org/Portals/0/PDFs/Milestones/ EmergencyMedicineMilestones.pdf).

Acker, Joan. 1990. "Hierarchies, Jobs, Bodies: A Theory of Gendered Organizations." *Gender & Society* 4(2):139–58.

Babcock, Linda, Maria P. Recalde, Lise Vesterlund, and Laurie Weingart. 2017. "Gender Differences in Accepting and Receiving Requests for Tasks with Low Promotability." *American Economic Review* 107(3):714–47.

Beck, Andrew H. 2004. "The Flexner Report and the Standardization of American Medical Education." *Journal of the American Medical Association* 291(17):2139–40.

Bennett, Christopher L., Ali S. Raja, Neena Kapoor, Dara Kass, Daniel M. Blumenthal, Nate Gross, and Angela M. Mills. 2019. "Gender Differences in Faculty among Academic Emergency Physicians in the United States." *Academic Emergency Medicine* 26(3):281–5.

Berryman, Sue E. 1983. "Who Will Do Science? Minority and Female Attainment of Science and Mathematics Degrees: Trends and Causes." New York: Rockefeller Foundation.

Bornmann, Lutz, Rüdiger Mutz, and Hans-Dieter Daniel. 2007. "Gender Differences in Grant Peer Review: A Meta-Analysis." *Journal of Informetrics* 1(3):226–38.

Braswell, Harold, and Harold I. Kushner. 2012. "Suicide, Social Integration, and Masculinity in the US Military." *Social Science & Medicine* 74(4):530–6.

Britton, Dana M. 2017. "Beyond the Chilly Climate: The Salience of Gender in Women's Academic Careers." *Gender & Society* 31(1):5–27.

Buchmann, Claudia, and Thomas A. DiPrete. 2006. "The Growing Female Advantage in College Completion: The Role of Family Background and Academic Achievement." *American Sociological Review* 71(4):515–41.

Buchmann, Claudia, Thomas DiPrete, and Anne McDaniel. 2008. "Gender Inequalities in Education." *Annual Review of Sociology* 34:319–37.

Budig, Michelle, and Paula England. 2001. "The Wage Penalty for Motherhood." *American Sociological Review* 66(2):204–25.

Castilla, Emilio J. 2008. "Gender, Race, and Meritocracy in Organizational Careers." *American Journal of Sociology* 113(6):1479–526.

Cecchi-Dimeglio, Paola. 2017. "How Gender Bias Corrupts Performance Reviews, and What to Do About It." *Harvard Business Review*. Retrieved December 12, 2018 (https://hbr.org/2017/04/how-gender-bias-corrupts-performance-reviews-and-what-to-do-about-it).

Cech, Erin, Brian Rubineau, Susan Sibley, and Carroll Seron. 2011. "Professional Role Confidence and Gendered Persistence in Engineering." *American Sociological Review* 76(5):641–66.

Choo, Esther K. 2017. "Damned if You Do, Damned if You Don't: Bias in Evaluations of Female Resident Physicians." *Journal of Graduate Medical Education* 9(5):586–7.

Correll, Shelley J. 2001. "Gender and the Career Choice Process: The Role of Biased Self-Assessment." *American Journal of Sociology* 106(6):1691–730.

Correll, Shelley J. 2004. "Constraints into Preferences: Gender, Status, and Emerging Career Aspirations." *American Sociological Review* 69(1):93–113.

Correll, Shelly J., Stephen Benard, and In Paik. 2007. "Getting a Job: Is There a Motherhood Penalty?" *American Journal of Sociology* 112(5):1297–338.

Correll, Shelly J., Kate Weisshaar, Alison T. Wynn, and JoAnne Wehner. 2018. "Inside the Black Box of Organizational Life: The Gendered Language of Performance Assessment." Working paper.

Dayal, Arjun, Daniel M. O'Connor, Usama Qadri, and Vineet Arora. 2017. "Comparison of Male vs Female Resident Milestone Evaluations by Faculty during Emergency Medicine Residency Training." *JAMA Internal Medicine* 117(5):651–7.

DeFazio, Christian R., Samuel D. Cloud, Christine M. Verni, Jessica M. Strauss, Karen M. Yun, Paul R. May, and Heather A. Lindstrom. 2017. "Women in Emergency Medicine Residency Programs: An Analysis of Data from Accreditation Council for Graduate Medical Education-Approved Residency Programs." *Academic Emergency Medicine* 1(3):175–8.

DiPrete, Thomas A., and Claudia Buchmann. 2013. *The Rise of Women: The Growing Gender Gap in Education and What It Means for American Schools*. New York: Russell Sage Foundation.

Downey, Douglas B., and Anastasia S. Vogt Yuan. 2005. "Sex Differences in School Performance during High School: Puzzling Patterns and Possible Explanations." *Sociological Quarterly* 46(2):299–321.

England, Paula. 2010. "The Gender Revolution: Uneven and Stalled." *Gender & Society* 24(2):149–66.

England, Paula, Paul Allison, Su Li, Noah Mark, Jennifer Thompson, Michelle Budig, and Han Sun. 2007. "Why Are Some Academic Fields Tipping Toward Female? The Sex Composition of U.S. Fields of Doctoral Degree Receipt, 1971–2002." *Sociology of Education* 80(1):23–42.

England, Paula, and Su Li. 2006. "Desegregation Stalled: The Changing Gender Composition of College Majors, 1971–2002." *Gender & Society* 20(5):657–77.

Fiske, Susan T. 1998. "Stereotyping, Prejudice, and Discrimination." Pp. 357–411 in *Handbook of Social Psychology*, Vol. 2. New York: McGraw-Hill.

Fiske, Susan T., Amy J. Cuddy, Peter Glick, and Jun Xu. 2002. "A Model of (Often Mixed) Stereotype Content: Competence and Warmth Respectively Follow from Perceived Status and Competence." *Journal of Personality and Social Psychology* 82(6):878–902.

Fuegen, Kathleen. 2007. "The Effects of Gender Stereotypes on Judgments and Decisions in Organizations." Pp. 79–98 in *The Social Psychology of Gender*. San Diego, CA: Elsevier.

Ginde, Adit A., Ashley F. Sullivan, and Carlos A. Camargo. 2010. "Attrition from Emergency Medicine Clinical Practice in the United States." *Annals of Emergency Medicine* 56(2):166–71.

Goniewicz, Mariusz. 2013. "Effect of Military Conflicts on the Formation of Emergency Medical Services

Systems Worldwide." *Academic Emergency Medicine* 20(5):507–13.

Gorman, Elizabeth H. 2005. "Gender Stereotypes, Same-Gender Preferences, and Organizational Variation in the Hiring of Women: Evidence from Law Firms." *American Sociological Review* 70(4):702–28.

Gorman, Elizabeth H., and Julie A. Kmec. 2009. "Hierarchical Rank and Women's Organizational Mobility: Glass Ceilings in Corporate Law Firms." *American Journal of Sociology* 114(5):1428–74.

Hanson, Sandra L., Maryellen Schaub, and David P. Baker. 1996. "Gender Stratification in the Science Pipeline: A Comparative Analysis of Seven Countries." *Gender & Society* 10(3):271–90.

Heilman, Madeline E. 2001. "Description and Prescription: How Gender Stereotypes Prevent Women's Ascent Up the Organizational Ladder." *Journal of Social Issues* 57(4):657–74.

Institute of Medicine. 2000. *To Err is Human: Building a Safer Health System*. Washington, DC: National Academies Press.

Iobst, William F., Jonathan Sherbino, Olle Ten Cate, Denyse L. Richardson, Deepak Dath, Susan R. Swing, Peter Harris, Rani Mungroo, Eric S. Holmboe, and Jason R. Frank. 2010. "Competency-Based Medical Education in Postgraduate Medical Education." *Medical Teacher* 32(8):651–56.

Ivankova, Natalita V., John W. Creswell, and Sheldon L. Stick. 2006. "Using Mixed-Methods Sequential Explanatory Design: From Theory to Practice." *Field Methods* 18(1):3–20.

Jacob, Brian A. 2002. "Where the Boys Aren't': Noncognitive Skills, Returns to School and the Gender Gap in Higher Education." *Economics of Education Review* 21(6):589–98.

Jacobs, Jerry A. 1989. *Revolving Doors: Sex Segregation and Women's Careers*. Stanford, CA: Stanford University Press.

Jacobs, Jerry A. 1996. "Gender Inequality and Higher Education." *Annual Review of Sociology* 22:153–85.

Kanter, Rosabeth Moss. 1977. *Men and Women of the Corporation*. New York: Basic Books.

Kay, Fiona M., and Elizabeth H. Gorman. 2008. "Women in the Legal Profession." *Annual Review of Law and Social Science* 4:299–332.

Kleinfeld, Judith. 1998. "The Myth That Schools Shortchange Girls: Social Science in the Service of Deception." Women's Freedom Network, Washington, DC. ERIC (Education Research Information Clearinghouse) document number ED 423 210.

Lautenberger, Diana M., Valerie M. Dandar, Claudia L. Raezer, and Rae Ann Sloane. 2014. "The State of Women in Academic Medicine: The Pipeline and Pathways to Leadership, 2013–2014." Association of American Medical Colleges. Retrieved August 3, 2017 (https://store.aamc.org/the-state-of-women-in-academic-medicine-the-pipeline-and-pathways-to-leadership-2013-2014.html).

Lofland, John, David Snow, Leon Anderson, and Lyn H. Lofland. 2006. *Analyzing Social Settings: A Guide to Qualitative Observation and Analysis*, 4th ed. Belmont, CA: Thomson/ Wadsworth.

Lu, Dave W., Nicholas D. Hartman, Jeffrey Druck, Jennifer Mitzman, and Tania D. Strout. 2019. "Why Residents Quit: National Rates and Reasons for Attrition among Emergency Medicine Physicians in Training." *Western Journal of Emergency Medicine* 20(2).

Misra, Joya, Jennifer Hickes Lundquist, and Abby Templer. 2012. "Gender, Work Time, and Care Responsibilities among Faculty." *Sociological Forum* 27(2):300–323.

Morris, Edward W. 2007. "'Ladies' or 'Loudies'? Perceptions and Experiences of Black Girls in Classrooms." *Youth & Society* 38(4):490–515.

Morris, Edward W. 2011. "Bridging the Gap: 'Doing Gender,' 'Hegemonic Masculinity,' and the Educational Troubles of Boys." *Sociology Compass* 5(1):92–103.

Mueller, Anna S., Tania M. Jenkins, Melissa Osborne, Arjun Dayal, Daniel M. O'Connor, and Vineet M. Arora. 2017. "Gender Differences in Attending Physicians' Feedback to Residents: A Qualitative Analysis." *Journal of Graduate Medical Education* 9(5):577–85.

Mulvey, Janet. 2010. "The Feminization of Schools: Young Boys Are Being Left Behind: What Targeted Teaching Strategies Can Help Them Reach Their Potential?" *Educational Digest* 75(8):35–8.

Nasca, Thomas J., Ingrid Philibert, Timothy Brigham, and Timothy C. Flynn. 2012. "The Next GME Accreditation System—Rationale and Benefits." *New England Journal of Medicine* 366(11):1051–6.

Nederhof, Anton J. 1983. "Methods of Coping with Social Desirability Bias: A Review." *European Journal of Social Psychology* 15(3):263–80.

O'Meara, KerryAnn, and Nelly P. Stromquist. 2015. "Faculty Peer Networks: Role and Relevance in Advancing Agency and Gender Equity." *Gender and Education* 27(3):336–58.

Quadlin, Natasha. 2018. "The Mark of a Woman's Record: Gender and Academic Performance in Hiring." *American Sociological Review* 83(2):331–60.

Ray, Victor. 2019. "A Theory of Racialized Organizations." *American Sociological Review* 84(1):26–53.

Reskin, Barbara. 1993. "Sex Segregation in the Workplace." *Annual Review of Sociology* 19:241–70.

Reskin, Barbara, and Debra McBrier. 2000. "Why Not Ascription? Organizations' Employment of Male and Female Managers." *American Sociological Review* 65(2):210–33.

Ridgeway, Cecilia L. 2009. "Framed Before We Know It: How Gender Shapes Social Relations." *Gender & Society* 23:145–60.

Ridgeway, Cecilia L., and Shelley J. Correll. 2004. "Unpacking the Gender System: A Theoretical Perspective on Gender Beliefs and Social Relations." *Gender & Society* 18(4):510–31.

Ridgeway, Cecilia L., and Lynn Smith-Lovin. 1999. "The Gender System and Interaction." *Annual Review of Sociology* 25:191–216.

Riegle-Crumb, Catherine. 2010. "More Girls Go to College: Exploring the Social and Academic Factors Behind the Female Postsecondary Advantage among Hispanic and White Students." *Research in Higher Education* 51(6):573–93.

Snyder, Thomas D., Cristobal de Brey, and Sally A. Dillow. 2018. *Digest of Education Statistics 2016*. Washington, DC: National Center for Educational Statistics.

Steinpreis, Rhea E., Katie A. Anders, and Dawn Ritzke. 1999. "The Impact of Gender on the Review of the Curricula Vitae of Job Applicants and Tenure Candidates: A National Empirical Study." *Sex Roles* 41(7/8):509–28.

Ten Cate, Olle. 2017. "Competency-Based Postgraduate Medical Education: Past, Present and Future." *German Medical Science Journal for Medical Education* 34(5):1–13.

Trix, Frances, and Caroline Psenka. 2003. "Exploring the Color of Glass: Letters of Recommendation for Female and Male Faculty." *Discourse and Society* 14(2):191–220.

Wennerås, Christine, and Agnes Wold. 1997. "Nepotism and Sexism in Peer-Review." *Nature* 387:341–43.

Xie, Yu, and Kimberlee A. Shauman. 2003. *Women in Science: Career Process and Outcomes*. Cambridge, MA: Harvard University Press.

Xu, Yonghong J., and Cynthia L. Martin. 2011. "Gender Differences in STEM Disciplines: From the Aspects of Informal Professional Networking and Faculty Career Development." *Gender Issues* 28:134–54.

**Alexandra Brewer** is an Assistant Professor in the Department of Sociology at Wake Forest University. She received her PhD in sociology at the University of Chicago in 2020. Her primary research interests are health, work, race, and gender. She is currently working on projects that examine how organizational processes reproduce social inequalities in the healthcare system and in the medical profession. Her work has won awards from the American Sociological Association and the Society for the Study of Social Problems and has been published in *Social Science & Medicine*.

**Melissa Osborne** is an Assistant Professor in the Sociology Department at Western Washington University. She received her PhD in sociology from the University of Chicago in 2019. Her research examines how "people changing" organizations—like schools and social services—shape life course trajectories, identities, and processes of social mobility, especially among marginalized populations. Her current project explores how first-generation students across the United States navigate college while contending with the unexpected complexities of social mobility. Osborne's publications can be read in *Society and Mental Health, The Journal of Contemporary Ethnography*, and *Contexts* among other journals.

**Anna S. Mueller** is an Associate Professor in the Department of Sociology at Indiana University. While the primary strand of Mueller's research examines youth suicide, she also investigates the production of inequality in medicine and education, with a focus on gender and emergency medicine. To do this, she integrates insights from sociology of gender, education, and work, with medical sociology. Her research has won several awards for its contributions to knowledge and has been published in journals such as the *American Sociological Review, Sociological Theory*, and *Journal of Health and Social Behavior*. For more about Mueller visit http://www.annasmueller.com.

**Daniel M. O'Connor** is a resident physician at the Harvard Combined Dermatology Residency Training Program sponsored by the hospitals affiliated with Harvard Medical School and a graduate of the University of Pennsylvania Perelman School of Medicine. His current research focuses on cutaneous oncology and how gender influences the evaluation of medical trainees.

**Arjun Dayal** is a resident physician at the University of Chicago Medicine Dermatology Residency Training Program and a graduate of the University of Chicago Pritzker School of Medicine, where he was awarded the Dean's Scholarship for Outstanding Promise in Medicine. His current research focuses on novel applications of technology in dermatology, and how gender influences the evaluation of medical trainees. His recent work has been published in *JAMA Internal Medicine, Journal of Graduate Medical Education*, and *The Journal of Neuroscience*.

**Vineet M. Arora** is the Herbert T. Abelson Professor of Medicine, Assistant Dean for Scholarship and Discovery, and Associate Chief Medical Officer-Clinical Learning Environment at University of Chicago Medicine. Her scholarship on improving the learning environment and care delivered to patients in teaching hospitals has been cited over 10,000 times. She is a founding member of two organizations dedicated to advancing gender equity in healthcare, TIME'S UP Healthcare and Women of Impact. She is an elected member of the National Academy of Medicine and published in journals such as *JAMA, Annals of Internal Medicine*, and *Academic Medicine*.