As part of a long overdue reckoning on diversity and inclusion in academic medicine, recent research has identified significant and persistent race-, gender-, and ethnicity-based disparities in pulmonary and critical care medicine (PCCM) training programs (1). A crucial first step in addressing these disparities is to understand the mechanisms that create and perpetuate biases in the selection process. Gender-based language differences in letters of recommendation (LORs) have been an early target for faculty development efforts, but many questions remain about how biases manifest in these letters, particularly in fields in which women are underrepresented, such as PCCM.

To explore these important questions, in this issue of ATS Scholar, Viglianti and colleagues (2) report on a single-center retrospective cohort study using natural language processing techniques (3) to assess for gender differences in LORs for PCCM applicants. The authors evaluated associations of applicant gender with LOR word count and a composite outcome of supportive word categories (including grindstone terms, agentic terms, research terms, ability terms, and standout terms). The model adjusted for letter writer gender, applicant race and ethnicity, and multiple factors associated with applicant achievement (publications, presentations, activities, chief resident status, and honor society status). It is worth noting that no applicants identified as nonbinary or gender diverse in the sample, which is an important limitation.

The authors found that LORs for women applicants were longer and contained more terms of support compared with letters for men. These findings represent a significant shift from early literature in which LORs for women were more likely to include references to communal traits (which includes words such as “compassionate,” “approachable,” and “empathetic”) and less likely to include standout terms and positive comments about their abilities (4, 5). Similar findings have been observed in the fields of radiation oncology (6) and surgery (7), both fields in which women are underrepresented, similar to PCCM. The authors propose that these findings suggest that “progress is being made in the letters of...
recommendation written on behalf of [women] applicants,” which may reflect faculty efforts to mitigate gender-based biases when writing LORs.

Despite this signal toward progress for women applicants, LORs for applicants underrepresented in medicine and Asian applicants were shorter and contained fewer supportive terms. This is a disheartening but not unexpected finding, given prior work demonstrating racial biases in written evaluations for trainees underrepresented in medicine (8).

Importantly, in addition to gender, implicit biases related to race and ethnicity (as well as other characteristics, such as disability, religion, and sexual orientation) may influence perceptions of trainee behavior and performance and ultimately can manifest as important differences in LORs. Although faculty development efforts to mitigate biases in LORs often target gender-based differences, this study highlights the need to extend this focus beyond this narrow lens.

We were struck by several secondary findings that warrant further exploration. There was a trend, although not statistically significant, toward higher use of communal terms in LORs for men. One potential explanation is that letter writers, through well-meaning efforts to minimize potential gender bias, are intentionally censoring communal terms from LORs for women while allowing them to remain in LORs for men. Alternatively, faculty may find these characteristics noteworthy, or praiseworthy, in men, yet expect these traits in women (and therefore not comment on them as a positive characteristic in LORs of women).

Regardless of the underlying reason, there may be unintended negative downstream effects on fellowship recruitment. Throughout the screening, interview, and selection process, reviewers of LORs may perceive the inclusion and/or omission of supportive and communal traits differently for men and women, affecting their impressions of applicants. In other words, whereas the absence of communal terms from a man’s LOR might go unnoticed, the omission could be perceived as a “red flag” for a woman applicant. This complicates one of the assumptions underpinning this study and research on LORs more generally: that the inclusion of more supportive terms reflects progress. Although it is certainly crucial that LORs comment on skills and abilities, communal traits such as empathy and kindness, or the ability to project those traits to others, seem essential as well.

Interestingly, the study also revealed that women faculty wrote letters that were longer and included more supportive terms compared with letters written by men and that this association was more pronounced in letters for women applicants. The authors should be applauded for evaluating the role of the letter writer in these findings, particularly as prior research has shown that faculty idiosyncrasies can drive evaluation differences and writing style (9).

Overall, these findings raise many additional questions. The increased word count may suggest that women faculty spend more time writing letters, especially for women applicants. This could reflect a positive effect of representation in medicine and/or a deliberate effort to promote gender equity. However, this may also hint at an additional burden for women faculty, as more time spent writing LORs may mean less time spent completing other tasks that align with traditional metrics of promotion. This would be a compelling area for further investigation.

On the basis of this study and others, we agree with the authors that the impact of
LORs in the screening, interview, and selection process must be carefully examined, with adoption of mitigation strategies, such as redaction of LORs for interviewers. The vulnerability of LORs to multiple layers of implicit bias, and the persistence of differences based on gender, race, and ethnicity, leaves us questioning the appropriateness of their continued use overall. It is critical to define the roles that LORs play in the selection process and consider who should write and read them. We should explore whether there are more reliable, less biased ways to capture the information that LORs aim to provide or whether there are processes we can implement to counteract these known biases. Overall, although it prompts many questions, this study is an important addition to the literature, describing continued linguistic differences in LORs for applicants to PCCM fellowships on the basis of gender, race, and/or ethnicity. We temper the authors’ conclusions that the findings of the study represent progress toward gender equity, given that the impact of these new differences on the selection process is not yet elucidated.

That said, the course of progress is rarely linear, and the authors have taken an important step in exploring the current state. This study should serve as a challenge to our pulmonary and critical care community to critically examine the complex ways in which implicit bias is embedded in our recruitment, interview, and selection processes, so we can continue to strive toward true equity in our field.

Author disclosures are available with the text of this article at www.atsjournals.org.

REFERENCES

1. Santhosh L, Babik JM. Trends in racial and ethnic diversity in internal medicine subspecialty fellowships from 2006 to 2018. JAMA Netw Open 3:e1920482.

2. Viglianti EM, Sheffield V, Choudhuri I, Goldberg R, Vita A, Byrd KM, et al. Gendered language in letters of recommendation for applicants to pulmonary critical care fellowships. ATS Scholar 2022;3:413–424.

3. Pennebaker JW, Booth RJ, Boyd RL, Francis ME. Linguistic inquiry and word count: LIWC2015. Austin, TX: Pennebaker Conglomerates; 2015. Available from: www.LIWC.net.

4. Madera JM, Hebl MR, Martin RC. Gender and letters of recommendation for academia: agentic and communal differences. J Appl Psychol 2009;94:1591–1599.

5. Trix F, Psenka C. Exploring the color of glass: letters of recommendation for female and male medical faculty. Discourse Soc 2003;14:191–220.

6. Grimm LJ, Redmond RA, Campbell JC, Rosette AS. Gender and racial bias in radiology residency letters of recommendation. J Am Coll Radiol 2020;17:64–71.

7. Horwood C, McDermott S, Gennell T, Pawlik TM, Grignon VP, Hughes TM. Letters of recommendation for surgical fellowships—does gender make a difference? Am J Surg 2021;221:90-94.

8. Ross DA, Boatright D, Nunez-Smith M, Jordan A, Chekroud A, Moore EZ. Differences in words used to describe racial and gender groups in medical student performance evaluations. PLoS One 2017;12:e0181659.

9. Ginsburg S, Gingerich A, Kogan JR, Watling CJ, Eva KW. Idiosyncrasy in assessment comments: do faculty have distinct writing styles when completing in-training evaluation reports? Acad Med 2020;95:S81–S88.