Engineering Educators Bringing the World Together
2025 ASEE Annual Conference & Exposition
Palais des congrès de Montréal, Montréal, QC • June 22–25, 2025 ASEE
Paper ID #46516

Are Letters of Recommendation to Engineering PhD Programs Biased? An Initial Examination

Dr. Allyson Flaster, University of Michigan - Ann Arbor, MI
Megan Legault, University of Michigan

n/a

Yifan Li, University of Michigan
Dr. Lisa R. Lattuca, University of Michigan

Lisa Lattuca, Professor of Higher Education and member of the Core Faculty in the Engineering Education Research Program at the University of Michigan.

Are Letters of Recommendation to Engineering PhD Programs Biased? An Initial Examination

Introduction

In this work-in-progress empirical research paper, we present initial findings from a study on whether letters of recommendation submitted as part of applications to engineering Ph.D. programs may contain implicit bias. Specifically, the study is examining whether there are rhetorical differences in the way letters of recommendation (LOR) are written for applicants from groups that are and are not underrepresented in engineering by gender and race/ethnicity.

The weight placed on LOR in making admissions decisions is likely increasing at some institutions as undergraduate and graduate programs drop standardized test score requirements in the hopes of improving equity [1]. In particular, the use of incredibly high quantitative GRE cut-offs for winnowing Ph.D. applicant pools has come under scrutiny as an admissions practice that is incompatible with increasing representation of marginalized students in STEM [2]. For example, at the institution providing data for our study, GRE scores were no longer considered for admission to Ph.D. programs as of 2022-2023 in an effort to remove barriers to inclusivity.

However, discontinuing use of the GRE will only result in an equitable graduate admissions landscape if the remaining application materials that admissions committees evaluate, such as letters of recommendation, are free of the racial and gender biases that standardized tests arguably possess. This is unlikely, as students from less resourced institutions and from less advantaged backgrounds have higher barriers to developing high-quality relationships with potential recommenders than their peers [3]. Even for students who do develop such relationships, there is no guarantee that their accomplishments will be viewed equally. Social psychologists warn implicit bias is ubiquitous, even among individuals who aim to treat others without prejudice, and especially in circumstances that involve high-stakes decisions [4].

Previous research on LORs, conducted primarily on small samples from medical residency and faculty searches, suggests that the language used in LORs for qualified applicants from groups underrepresented in STEM can differ from groups than aren't underrepresented in STEM. For example, using dictionaries of words and phrases with positive and negative associations, some find that letters written for men contain more 'standout' words (e.g., "one of a kind") than those written for women [5]-[9]. These studies find that letters for men are also more specific. However, a different study finds women are *more* likely than men to be described using standout words [10]. Another finds no evidence that men and women are described in systematically different ways but does find evidence that applicants from racially minoritized groups are described in less positive ways than White applicants [11]. Small samples, use of domain irrelevant term dictionaries, and a variety of field-specific contexts (e.g., urology, ophthalmology) may contribute to these inconsistent results. This paper builds on previous research by asking the following questions:

> **RQ1**: When deciding who to admit to PhD programs, what kinds of LOR do engineering faculty find most credible as sources of information about applicants' potential for long-term success as engineering researchers?

RQ2: Are engineering PhD program applicants from underrepresented backgrounds less likely than other applicants to have the types of LOR that faculty find most credible?

Framework

Our study is informed by two strands of literature. The first comes from communication studies and linguistics and emphasizes that LORs are a specific genre of rhetorical discourse [12], [13]. Rhetoricians attempt to convince their readers by establishing credibility (ethos), appealing to reason through logical arguments and evidence (logos), and appealing to emotion or feeling (pathos). As a genre, LORs have an expected content and form [12]. The expected form of a LOR generally comprises three segments [8]. In the first segment, the writer reinforces their ethos by describing the ways, contexts, and length of time in which they have known the applicant. In the second segment, the writer describes in detail what they have observed of the applicant's traits and actions, thereby providing the needed logical appeal. The third segment provides a summary recommendation for or against the applicant, and perhaps ends with another appeal to ethos by comparing the applicant to other candidates the writer has known. Although direct emotional appeals (pathos) are not a favored tactic in academic discourse, we suspect that writers may still exert influence over the reader's emotions throughout the letter by using words that invoke implicit positive or negative feelings about the applicant. The previous research on LORs discussed above has primarily examined this aspect of discourse.

The second strand of literature that informs our study focuses on the organizational context of academic decision making and is influenced by theories about epistemic cultures. This literature highlights that the graduate admissions process is highly decentralized and local, meaning admissions decisions are made by representatives of the academic department or program, almost always faculty [2], [14]. A major goal of doctoral admissions is to admit the candidates who have the best chance of developing into successful researchers over the long term [2]. In doing so, faculty draw from ideals of success that are firmly rooted in the epistemologies and cultures of their disciplines [15], [16]. This literature points to the importance of seeking faculty members' perspectives on what constitutes potential for long-term success in their specific engineering discipline, which informs our study's methodology.

Methodology

In this project, we employ elements of qualitative research, quantitative content analysis, natural language processing techniques, and descriptive statistics to produce preliminary analyses of letters of recommendation for applicants to engineering PhD programs.[1] We focus our analysis on domestic applicants to four engineering programs (aerospace, chemical, electrical, and mechanical) at a large public research university in the U.S. between 2010 and 2022. We focus on domestic applicants because previous research has found substantial cross-national differences in norms for writing LORs and we wanted to minimize the influence of those differences [5]. We selected four programs to sample from to provide variation in program size, selectivity, and diversity.

[1] This research was initially funded by the National Science Foundation under Grant No. 2225209, awarded in 2022. Funding was discontinued by the NSF in April 2025 because the project's focus on broadening participation in STEM "no longer effectuate[s] the program goals or agency priorities."

The data for this study comes from three sources: PhD program applications provided by the university's graduate school, student records from the university's data warehouse, and interviews conducted by the research team of 11 engineering faculty across the four programs. Interview questions asked about the credibility of different types of LORs (the ethos element of LORs) and what evidence faculty looked for in the letters to indicate that applicants would be successful in the program (the logos elements of LORs). Two team members worked together to code the interview transcripts, looking for common concepts mentioned across faculty.

The PhD program applications were provided as PDF and TIFF files. To parse the letters of recommendation from the applications and turn them into text files for analyses, a member of the research team used the pdfminer package for PDF files and the pytesseract package for TIFF files in Python. The student records contain demographic information about applicants as well as the outcome of their application (denied or admitted). To conduct our content analysis, we randomly selected a sub-sample of 70 applicants from each of the four PhD programs to import the text files of their LORs into MAXQDA software for manual coding. We considered each sampled applicant to be a case, and all of the case's letters were included in the analysis. Only the body of the letters were retained for analysis. As an initial step, two members of the research team read all of the letters and coded the relationship(s) between the letter writer and the applicant. A full description of the relationship codes is provided in Table A1 in the Appendix.

For this paper, we present only the initial results of our examination of the credibility (ethos) of LORs. Future analyses will use data from many more applicants, examine differences across the four programs, and expand the focus to other rhetorical elements.

Findings

RQ1 asked, *what kinds of LOR do engineering faculty find most credible as sources of information about applicants' potential for long-term success as engineering researchers?* The interview data indicates that, by far, engineering faculty find letters of recommendation from research supervisors to be the most credible source of information about applicants' potential for success. Additionally, they value information from recommenders that provide substantial detail about the applicant's traits and performance. Based on this information, we created document-level variables to measure the credibility of the letter; they are described in Table 1.

Table 1. Constructs Used to Measure Letter Credibility

Construct	Operationalization/Variable Description	Rationale
Written by research supervisor	Takes a value of 1 if the writer has supervised the applicant in conducting research, and 0 otherwise	Engineering faculty find letters from research supervisors to be more credible than letters from other types of recommenders when gauging applicants' potential for success.
Level of detail	Word count of the letter. Values range from 0 (this occurs if the recommender declines to submit a letter) to the maximum observed value of 1,778.	The more detail writers provide about the applicant, the longer the letter is [8].

To validate these measures, we calculate summary statistics with a two-sample t-test for applicants that were admitted and applicants that were denied. The statistics in Table 2 suggest that almost all applicants who are admitted have at least one letter of recommendation from a

research supervisor (95%), compared to only 3/4 of applicants who are denied (74%), a difference that is highly significant. Similarly, applicants who are admitted have more detailed letters than applicants who are denied (1352 words across all three letters vs. 1121 words across all three letters). This suggests that our measures serve as acceptable proxies for letter credibility.

Table 2. LOR Credibility Measures, by Applicant Admission Status

	Admitted (n=47)	Denied (n=195)	
Variable	**Mean/ Percent**	**Mean/ Percent**	**Sig. Diff.**
At least one LOR written by research supervisor	95.74	73.85	***
Total word count across all LORs	1352	1121	***

Notes: Statistics were calculated at the applicant level. Since programs require 3 LORs for admission, we drop incomplete applications (those with less than 3 LORs). The resulting applicant sample size is n=242. Legend: ŧ<.10; * p<.05; ** p<.01; *** p<.001

To answer RQ2 (*Are engineering PhD program applicants from underrepresented backgrounds less likely than other applicants to have the types of LOR that faculty find most credible?),* we conduct preliminary analyses to see if there are differences in LORs by underrepresented status. Table 3 presents the results of those two-sample t-tests. We find no significant differences in the credibility measures between women and men. However, there are differences between applicants who do and do not identify as African American/Black, Latinx/Hispanic, and/or Native American/Alaskan Native (underrepresented students of color [USC]). In comparison to students who identify as White and/or Asian (which are racial groups that are not underrepresented in engineering), USC applicants are 10 percentage points less likely to have a letter written by a research supervisor and the total number of words written about them across all three LORs is 148 words fewer. We do not currently have enough data coded to examine whether there are significant differences by engineering programs, but plan to do so.

Table 3. LOR Credibility Measures, by Underrepresented Gender and Race/Ethnicity Status

	Women (n=46)	Men (n=196)		USC (n=39)	White/ Asian (n=203)	
Variable	**Mean/ Percent**	**Mean/ Percent**	**Sig. Diff**	**Mean/ Percent**	**Mean/ Percent**	**Sig. Diff**
At least one LOR written by research sup.	84.78	76.53		69.23	79.80	ŧ
Total word count across all LORs	1217	1154		1042	1190	*

Notes: Statistics were calculated at the applicant level. Only applicants with complete applications were included (n=242). Complete applications are those containing at least 3 LORs. Legend: ŧ<.10; * p<.05; ** p<.01; *** p<.001

Implications

How faculty decide who to admit to graduate programs is an opaque process with substantial consequences for equity in STEM. The findings from this study are a first step in providing policy-relevant knowledge about the role that LORs play in structuring access to engineering research and teaching careers. Our initial results suggest that eliminating the threat of bias against groups that are underrepresented in engineering by race and ethnicity will require more action than just discontinuing or downplaying the use of standardized tests. There is initial evidence that, at the university represented in this study, underrepresented students of color may be disadvantaged in the admissions process by trends in the level of detail provided by their recommenders and in their ability to garner recommendations from research supervisors.

References

[1] K. Langin, "A wave of graduate programs drops the GRE application requirement," *Science*, May 2019.

[2] J. R. Posselt, *Inside Graduate Admissions: Merit, Diversity, and Faculty Gatekeeping*. Cambridge, MA: Harvard University Press, 2016.

[3] S. Pierszalowski, J. Bouwma-Gearhart, and L. Marlow, "A systematic review of barriers to accessing undergraduate research for STEM students: Problematizing under-researched factors for students of color," *Soc. Sci. (Basel)*, vol. 10, no 328, Sep. 2021. https://doi.org/10.3390/socsci10090328

[4] J. L. Eberhardt, *Biased: Uncovering the Hidden Prejudice that Shapes What We See, Think, and Do*. New York, NY: Penguin, 2019.

[5] K. Dutt, D. L. Pfaff, A. F. Bernstein, J. S. Dillard, and C. J. Block, "Gender differences in recommendation letters for postdoctoral fellowships in geoscience," *Nat. Geosci.*, vol. 9, no. 11, pp. 805–808, Nov. 2016.

[6] J. M. Madera, M. R. Hebl, and R. C. Martin, "Gender and letters of recommendation for academia: agentic and communal differences," *J. Appl. Psychol.*, vol. 94, no. 6, pp. 1591–1599, Nov. 2009.

[7] T. Schmader, J. Whitehead, and V. H. Wysocki, "A linguistic comparison of letters of recommendation for male and female chemistry and biochemistry job applicants," *Sex Roles*, vol. 57, no. 7–8, pp. 509–514, 2007.

[8] F. Trix and C. Psenka, "Exploring the color of glass: Letters of recommendation for female and male medical faculty," *Discourse & Society*, vol. 14, no. 2, pp. 191–220, Mar. 2003.

[9] F. E. Turrentine, C. N. Dreisbach, A. R. St Ivany, J. B. Hanks, and A. T. Schroen, "Influence of gender on surgical residency applicants' recommendation letters," *J. Am. Coll. Surg.*, vol. 228, no. 4, pp. 356-365.e3, Apr. 2019.

[10] A. Powers, K. M. Gerull, R. Rothman, S. A. Klein, R. W. Wright, and C. J. Dy, "Race- and gender-based differences in descriptions of applicants in the letters of recommendation for orthopaedic surgery residency," *JB JS Open Access*, vol. 5, no. 3, Jul. 2020.

[11] B. V. Chapman *et al.*, "Linguistic biases in letters of recommendation for radiation oncology residency applicants from 2015 to 2019," *J. Cancer Educ.*, Oct. 2020.

[12] M. Albakry, "Telling by omission," in *Studies in Corpus Linguistics*, Amsterdam: John Benjamins Publishing Company, 2015, pp. 79–98.

[13] J. Yates and W. J. Orlikowski, “Genres of organizational communication: A structurational approach to studying communication and media,” *AMRO*, vol. 17, no. 2, pp. 299–326, Apr. 1992.

[14] R. S. Michel, V. Belur, B. Naemi, and H. J. Kell, “Graduate admissions practices: A targeted review of the literature,” *ETS Res. Rep. Ser.*, vol. 2019, no. 1, pp. 1–18, Dec. 2019.

[15] T. Becher and P. Trowler, *Academic Tribes and Territories*, 2nd ed. Buckingham, England: Open University Press, 2001.

[16] K. Knorr-Cetina, *Epistemic cultures: How the Sciences Make Knowledge*. London, England: Harvard University Press, 2009.

Appendix

Disclaimer: The findings and conclusions presented herein are solely those of the study team in their capacity as researchers and do not necessarily reflect positions or policies of the University of Michigan Rackham Graduate School.

Table A1: LOR Relationship Codebook

For each LOR, consider the type of relationship(s) the letter writer has with the applicant they are recommending. For each type of relationship that appears in the letter, enter a 1 for the appropriate variable(s).

More than one type of relationship may be applied.

LETTER WRITER RELATIONSHIP	Type of Variable	Description
Research Supervisor	integer (0/1)	Letter writer supervised the applicant's work on a research project, in a lab, and/or another research setting
Instructor	integer (0/1)	Letter writer instructed the applicant in a class (e.g., professor, graduate student instructor)
Industry Supervisor	integer (0/1)	Letter writer oversaw the applicant's work in an industry setting; the applicant could have been an intern, coop student, or an employee
Advisor	integer (0/1)	Letter writer served as an advisor for the applicant in an academic or program capacity (e.g., academic advisor, program advisor, undergraduate major advisor)
Teaching Role Supervisor	integer (0/1)	Letter writer oversaw the applicant's work in a teaching role (e.g., the student was a teaching assistant, graduate student instructor, or tutor for the instructor's class)
Other	text	If the type of letter writer relationship is not one of the listed categories, briefly describe the type of relationship
No letter provided/only form	integer (0/1)	Mark as a 1 if the recommender only completed the recommendation form and did not submit a letter with their recommendation
OTHER RELATIONSHIPS		**Description**
Thesis advisor		If letter writer says they were the applicant's thesis advisor, (1) mark the type of relationship as a **Research Supervisor** and (2) write in **Thesis advisor** in the **Other** relationship category
Undeterminable		If the relationship between the letter writer and the applicant is unclear from the letter, write **Undeterminable** in the **Other** relationship category