

Gender Differences in Language of Standardized Letter of Evaluation Narratives for Emergency Medicine Residency Applicants

Simiao Li, MD, MS, Abra L. Fant, MD, MS, Danielle M. McCarthy, MD, MS, Danielle Miller, MD, Jill Craig, BA, and Amy Kontrick, MD

ABSTRACT

Objective: While gender differences in language for letters of recommendation have been identified in other fields, no prior studies have evaluated the narrative portion of the emergency medicine (EM) standardized letter of evaluation (SLOE). We aim to examine the differences in language used to describe male and female applicants within the SLOE narrative.

Methods: Invited applicants to a 4-year academic EM residency program within a single application year with a SLOE were included in the sample. Exclusion criteria were SLOE of applicants from non-Liaison Committee on Medical Education (LCME) schools or first rotation SLOE not available for download. Data were collected on applicant gender, age, rotation grade, Alpha Omega Alpha designation, and medical school rank. The previously validated Linguistic Inquiry and Word Count (LIWC) program was used to analyze frequency of words within categories relevant to letters of recommendation. Descriptive statistics, t-tests, and chi-square tests were employed in analysis.

Results: Of 1,025 applicants within a single application year, 265 were invited to interview; 237 applicants had a first rotation SLOE available for analysis. There were no differences between male and female applicants for baseline characteristics. The median word count per SLOE narrative was 199; within the LIWC dictionary and user-defined categories, words within the categories of affiliation and ability appeared more frequently for female applicants.

Conclusions: Our results with respect to the SLOE narrative reinforce prior research that letters of recommendation for female applicants highlight communal characteristics of teamwork, helpfulness, and compassion. Contrary to prior research, ability words highlighting intelligence and skill appeared with greater frequency for female applicants. No pervasive differences were found in other word categories. In this sample, the standardized format of the SLOE resulted in letters that were relatively free of gender bias.

As early as two decades ago, it was noted that disproportionately few women were applying to the specialty of emergency medicine (EM) and pursuing careers within academic EM.¹ One facet of advancement within academics is the letter of recommendation, and a significant body of research has been devoted to gender differences in language used in these letters.^{2–5} Research has found that women tend to be described using grindstone characteristics such as capacity for hard work while men are

From the Department of Emergency Medicine, Northwestern McGaw Medical Center (SL, ALF, DMM, DM, JC, AK), Chicago, IL. Received June 23, 2017; revision received July 26, 2017; accepted August 8, 2017.

Presented at the Central Group on Education Affairs Regional Spring Meeting, Chicago, IL, March 30, 2017; and the Council of Emergency Medicine Residency Directors Academic Assembly, Ft. Lauderdale, FL, April 28, 2017.

The authors have no relevant financial information or potential conflicts to disclose.

Supervising Editor: Sebastian J. Cico, MD, MEd.

Address for correspondence and reprints: Simiao Li, MD, MS; e-mail: simiao.li@gmail.com

AEM EDUCATION AND TRAINING 2017;1:334–339.

described in terms of innate ability with more stand-out adjectives.^{4,5} Letters for female applicants contain fewer status terms and raise more doubt, and emphasize communal qualities such as helping others and maintaining relationships, which in turn have a negative association with hiring.^{3,6}

In the application process to residency, EM educators have sought to standardize letters of recommendations to decrease the utilization of “code words” and the time-consuming task of deciphering these letters.^{7–9} Since adopting the standardized letter of recommendation (SLOR) and now standardized letter of evaluation (SLOE), it has become the most important factor in selecting applicants for interview.¹⁰ Several iterations of the SLOR and SLOE have been used. The current SLOE simplifies the description of an applicant into two sections, a quantitative evaluation of applicant characteristics (commitment to EM, work ethic, clinical abilities, prediction of success) and a short narrative detailing noncognitive attributes.¹¹ The intent of the narrative portion of the SLOE is a succinct, focused description of the applicant with respect to his or her noncognitive attributes felt by educators to be essential for successful EM residency training.

Despite the standardization of the SLOE, variability continues to exist in the composition of the narrative. Some of this variability is based on writer experience or unnecessarily long commentary.^{12–14} Within the EM literature, ethical guidelines for writing letters exist but there are limited data on gender and its relationship to language use in the letters.¹⁵ A review of the EM literature revealed only one study evaluating gender that examined association of the superlative “guaranteed match” SLOE ranking within writer/applicant gender dyads (e.g., male/female, male/male). This study found that the female writer/female applicant combination was twice as likely to receive a “guaranteed match” ranking compared to all others.¹⁶

To our knowledge, no prior studies have evaluated the narrative portion of the EM SLOE with respect to gender. The aim of this study is to compare differences in language used to describe male and female applicants in the SLOE narrative.

METHODS

Study Setting

Northwestern University’s EM residency (NUEM) is a 4-year, academic residency program with 58 total residents who train in an urban setting at Northwestern

McGaw Medical Center. Applications to NUEM residency program are accepted through the Electronic Residency Application Service (ERAS), which transmits applications, letters of recommendation, medical student performance evaluations, and transcripts to residency programs. Applicants must participate in the National Resident Matching Program (NRMP) to be eligible for selection to NUEM. This study was approved by the Institutional Review Board at Northwestern University.

Sample

Invited applicants to the NUEM residency for application cycle 2015–2016 were included in the sample. The narrative portion of the SLOE was included in analysis. It is limited to 250 words and asks the writer to “Please concisely summarize this applicant’s candidacy including . . . (1) Areas that will require attention, (2) Any low rankings from the SLOE, and (3) Any relevant noncognitive attributes such as leadership, compassion, positive attitude, professionalism, maturity, self-motivation, likelihood to go above and beyond, altruism, recognition of limits, conscientiousness, etc.”¹¹ If applicants submitted more than one SLOE, the SLOE from the first clinical rotation in EM was included in analysis. Exclusion criteria were SLOE of applicants from non-Liaison Committee on Medical Education (LCME) schools (those that confer DO, MBBS, MBCh degrees), as well as applicants with a first-rotation SLOE that was not available for download. SLOE analysis began after all NRMP decisions had been made and finalized.

Measurements

Measures obtained from the ERAS application for use in describing the sample include: age at time of application, gender, grade on the rotation corresponding to the SLOE, Alpha Omega Alpha designation at the time of application, if the letter was attributed to a “group” of writers or to a single author, and gender of the letter writer if it was written by a single author. The national rank of applicants’ medical schools were obtained from the application years’ *US News and World Report* rankings in the research category.

A linguistic approach was taken to measure the language use in the letters. The Linguistic Inquiry and Word Count (LIWC) program is a validated text analysis software program that analyzes text files and compares words within a text document to predefined word dictionaries.^{17,18} LIWC reports the percentage of

words that fall into 90 possible word categories composed of over 6,400 words and word stems that reflect emotions, thinking styles, social concerns, and parts of speech. LIWC output reports the ratio of the words in each category relative to the total word count of the queried text file; this ratio is then converted into a word count per category. LIWC software has been used by three prior studies to analyze letters of recommendation.²⁻⁴

Of the 90 default categories within the LIWC 2015 dictionary, we identified categories of particular interest based on prior research utilizing LIWC in the context of letters of recommendation and consensus of the study team.⁴ These categories included positive emotion (e.g., nice), negative emotion (e.g., nasty), tentative words (e.g., maybe), certainty words (e.g., always), affiliation words (e.g. social), achievement words (e.g., success), power words (e.g., superior), reward words (e.g., benefit), and risk words (e.g., doubt). Additionally, LIWC software allows for user-defined dictionaries to be created, and we employed an additional seven user-defined dictionaries based on prior studies analyzing gender differenced in letters of recommendation. Schmader and colleagues⁴ created five user-defined LIWC dictionaries based on the qualitative work of Trix and Penska.⁵ These categories include grindstone traits (e.g., diligent), ability traits (e.g., talented), standout adjectives (e.g., exceptional), research terms (e.g., project), and teaching terms (e.g., teach). Madera and colleagues³ created an additional two dictionaries based on the social role theory and the work of Eagly⁶ in communal and agentic characteristics. Communal work behaviors include helping others or maintaining relationships, and the adjectives in the user-defined dictionary reflect this (e.g., kind, caring); whereas agency comprises descriptions of independence and self-confidence (e.g., ambitious, confident).

Procedure

SLOE narratives were downloaded from ERAS by NUEM program coordinators and converted to Microsoft Word format. Stock sentences not related to applicant characteristics, such as descriptions of the rotation, were deleted prior to analysis.

Analysis

We used descriptive statistics to report applicants' characteristics and assessed for differences by gender using t-tests and chi-square tests as appropriate. Median word counts for the identified 16 categories of

interest (nine LIWC default categories, seven user-defined dictionaries) were reported and differences by gender were assessed using Wilcoxon rank sum test. No prior studies have been performed on SLOE narratives or letters of similar length (word limit = 250); therefore, estimating word counts for a power analysis was challenging. However, we estimated that with our sample size of 237 (allocation 1.5 male/female) that we would have 80% power to detect a difference of 0.37 mean words within a single word category with a 5% Type I error (based on estimated baseline word frequency per category of three words). All analyses were performed using Stata 13.1 (StataCorp).

RESULTS

There were 1,025 applicants to NUEM residency in the single application cycle of study (2015–2016). Of these, 265 applicants were invited to interview. 237 applicants had a first-rotation SLOE available for analysis. First-rotation SLOEs which were excluded ($n = 28$) were those from applicants from non-LCME schools or those for whom a first-rotation SLOE was not available or unable to be downloaded (Figure 1). The male-to-female ratio of applicants was approximately 1.5:1. Comparing male and female applicants, there were no differences between groups for baseline characteristics (Table 1).

The analysis of the letters with LIWC revealed a median word count per SLOE narrative of 199 (male applicant median [IQR] = 184.5 [129–255] words; female applicant median [IQR] = 217 [144–258] words; $p = 0.22$). Within LIWC dictionary and user-defined categories, the frequency of affiliation (male median [IQR] = 5 [3–7] words; female median [IQR] = 6 [4–8]; $p = 0.03$) and ability (male median [IQR] = 2 [1–3]; female median [IQR] = 2 [1–4]; $p = 0.04$) words were significantly different between male and female applicants (Table 2). There were no gender-based differences for the remaining 14 categories evaluated.

DISCUSSION

In this study, we found a difference in frequency of certain descriptors used to characterize male and female applicants. One predefined LIWC dictionary category, affiliation, stood out as a group of words that appeared in greater frequency for female applicants.

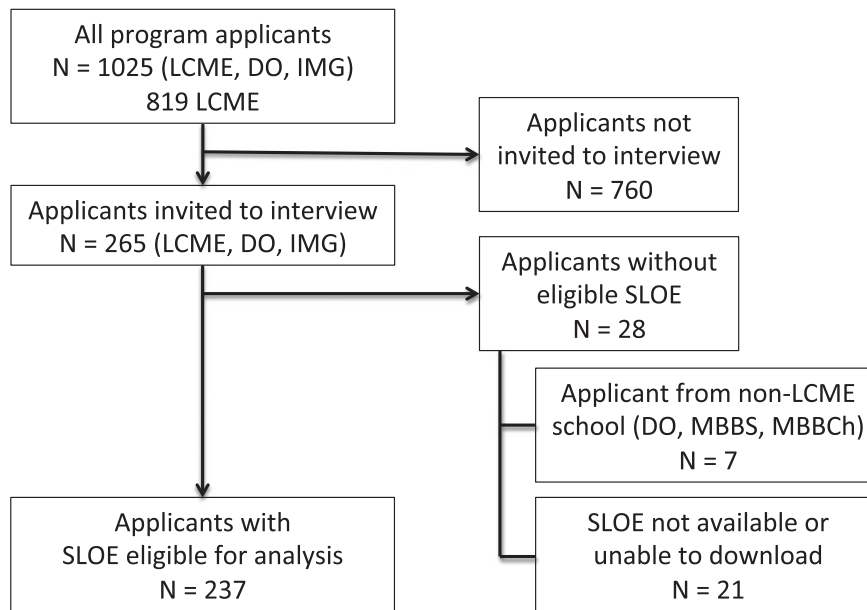


Figure 1. Selection of SLOE for inclusion in analysis. LCME = Liaison Committee on Medical Education; DO = Doctor of Osteopathic Medicine; IMG = International Medical Graduate; SLOE = Standardized Letter of Evaluation; MBBS, MBBCh = Bachelor of Medicine, Bachelor of Surgery.

Table 1
Applicant Information

Variable	Total (N = 237)	Male (n = 142)	Female (n = 95)	p-value
Age (y)	27 (SD ± 2.5)	27	27	0.45
Top 25	34%	36%	32%	0.49
Alpha Omega Alpha	39%	44%	32%	0.06
Rotation honors grade	66%	64%	68%	0.60

Table 2
Select LIWC Output Variables

Variable	Total (N = 237)	Male (n = 142)	Female (n = 95)	p-value
Word count	199	184.5	217	0.22
Words per sentence	15.6	15.4	16	0.37
Positive emotion	12 (8–7)	11 (8–17)	13 (9–18)	0.24
Negative emotion	1 (0–2)	1 (0–2)	1 (0–2)	0.09
Social	24 (16–33)	23 (16–33)	26 (17–36)	0.10
Cognitive processes	15 (10–23)	15 (10–22)	16 (11–23)	0.22
Affiliation	5 (3–7)	5 (3–7)	6 (4–8)	0.03
Achievement	10 (6–12)	10 (6–12)	10 (7–13)	0.33
Power	7 (5–10)	7 (5–10)	8 (5–11)	0.09
Reward	4 (2–6)	4 (2–6)	4 (2–6)	0.44
Risk	0 (0–1)	0 (0–1)	0 (0–1)	0.82
Standout	2 (1–3)	2 (1–3)	2 (1–4)	0.94
Ability	2 (1–3)	2 (1–3)	2 (1–4)	0.04
Grindstone	2 (1–4)	2 (1–4)	2 (1–4)	0.61
Teaching	3 (2–4)	3 (2–4)	3 (1–4)	0.63
Research	0 (0–2)	0 (0–2)	0 (0–1)	0.51
Communal	1 (0–2)	1 (1–2)	1 (0–2)	0.65
Agency	2 (1–3)	2 (1–3)	2 (1–3)	0.16

Data are reported as median (IQR).
IQR = interquartile range.

Affiliation words as defined by the LIWC dictionary include those emphasizing teamwork, helpfulness, communication, compassion, and empathy. These words are in line with the framework established by Eagly,⁶ a psychologist well known for research on sex differences in the workplace. Eagly posits that expected and valued behaviors by women are communal (helping others, maintaining relationships), while men embody behaviors that demonstrate individual agency (aggressiveness, assertiveness, independence, confidence). The use of affiliation words on SLOEs may be intended to highlight applicants' dedication toward patient care; however, these data demonstrate that such topics are mentioned more frequently in the letters of female applicants. This difference may result from the female applicants having displayed more communal behaviors than their male counterparts during the rotation; alternatively, this difference may also be the result of gender stereotypes.

Within user-defined categories, ability words (intelligence, instinct, skill, competence, expertise) also appeared with greater frequency for female applicants. This finding is in contrast to prior research examining letters of recommendation for medical school faculty positions that determined that while women were described in recommendation letters by their capacity for hard work (e.g. "grindstone" characteristics), men were described in terms of innate ability.⁵ The standardized questions posed to SLOE authors may have contributed to this finding.

In contrast to prior research which found that letters for female applicants raised more doubt (i.e., negative language, hedges, and faint praise),^{2,5} we found no difference in negative language between SLOEs for male and female applicants within our analysis. Three other word categories in which there was no statistical significance between groups were "standout," "teaching," and "research." Prior research using LIWC to analyze letters of recommendation for chemistry and biochemistry faculty applicants at a research university determined that male candidates' letters contained more adjectives that made the candidate "standout" (e.g., exceptional, unique).⁴ This difference was not captured within our analysis. Finally, within the work of Trix and Psenka,⁵ letters of recommendation reinforced the preexisting schema of women as teachers and students and men as researchers and professionals. Our data did not reflect a difference in the categories of "teaching" and "research" with respect to male and female applicants. It is possible that within

this sample, SLOE authors did not highlight teaching and research characteristics as the role of the SLOE narrative primarily serves to describe the applicant's performance during a clinical rotation. Additionally, in contrast to prior studies wherein letters were for faculty positions, the role of applicants in our analysis was that of the junior learner.

While some differences exist in language used to describe male and female applicants, the SLOE narratives in this sample are, as a whole, free of biased descriptors. The standardization of the SLOE, with specific prompts for the writer, strict word count guidelines, and a focus on qualities deemed important by a group of educators may have resulted in fewer gender-based differences between applicants.^{8,12,19} While these results demonstrate a lack of bias in one format of evaluation, recent studies suggest that implicit bias persists within the process of evaluation of trainees, including during residency: a recent multicenter study by Dayal et al.²⁰ highlights lower scores for female residents in EM milestones ratings compared to male peers as they progress through residency.

LIMITATIONS

Only SLOE narratives from applicants invited for interview at a single academic EM residency program were included in analysis. While limiting analysis to narratives of invited applicants minimizes differences due to confounders such as applicant competitiveness, this also limits generalizability. It is possible that greater differences in language may be revealed when considering all applicants rather than only those invited to interview. Similarly, the SLOEs of seven applicants (2.6% of those invited) from non-LCME programs were excluded from analysis. This may decrease the variability of language within SLOE. Widening the sample to all applicants (rather than only "applicants invited to interview") may show different language use patterns for a more diverse range of applicants. ERAS does not allow an individual program to access SLOEs for applicants who have not selected that program; therefore, a national sample is not feasible using this methodology of data collection. Validated word lists for "gendered" words are limited. The paucity of words within these categories compared to the LIWC dictionary may limit the ability to capture biased language within the EM vocabulary. Word count limitations of the SLOE format may have limited the statistical power to detect differences between

genders within our sample size. Finally, as the majority of letters were composed by a group of authors, we were not able to evaluate relationships between author gender and applicant gender with respect to language used in the SLOE.

CONCLUSIONS

In a sample of SLOEs from a single residency program, this linguistic analysis found differences in word frequency between narratives of male and female applicants for the categories of affiliation and ability; of the 16 total word categories investigated, no differences were found for the other 14 categories. SLOE authors are, as a whole, describing male and female applicants to residency using similar language. A standardized format similar to the SLOE could be considered for other specialties within the field of medicine, particularly for those in which gender imbalance may be a concern. While further study is necessary to determine whether the dictionaries used in our research reflect the language and values of emergency physicians, in this sample the standardized format of the SLOE provided a framework in which language used was relatively free of gender bias.

The authors acknowledge Juan M. Madera, PhD, for sharing his communal and agentic word dictionaries.

References

1. Martin ML. Applicant pool for emergency medicine residency programs: information on minority and female applicants. *Ann Emerg Med* 1996;27:331–8.
2. Isaac C, Chertoff J, Lee B, Carnes M. Do students' and authors' genders affect evaluations? A linguistic analysis of medical student performance evaluations. *Acad Med* 2011;86:59–66.
3. Madera JM, Hebl MR, Martin RC. Gender and letters of recommendation for academia: agentic and communal differences. *J Appl Psychol* 2009;94:1591–9.
4. Schmader T, Whitehead J, Wysocki VH. A linguistic comparison of letters of recommendation for male and female chemistry and biochemistry job applicants. *Sex Roles* 2007;57:509–14.
5. Trix F, Psenka C. Exploring the color of glass: letters of recommendation for female and male medical faculty. *Discourse Society* 2003;14:191–220.
6. Eagly AH, Wood W, Diekmann AB. Social role theory of sex differences and similarities: a current appraisal. In: Eckes T, Trautner HM, editors. *The Developmental Social Psychology of Gender*. Mahwah, NJ: Erlbaum, 2000. pp. 123–74.
7. Balentine J, Gaeta T, Spevack T. Evaluating applicants to emergency medicine residency programs. *J Emerg Med* 1999;17:131–4.
8. Girzadas DV Jr, Harwood RC, Dearie J, Garrett S. A comparison of standardized and narrative letters of recommendation. *Acad Emerg Med* 1998;5:1101–4.
9. Keim SM, Rein JA, Chisholm C, et al. A standardized letter of recommendation for residency application. *Acad Emerg Med* 1999;6:1141–6.
10. Love JN, Smith J, Weizberg M, et al. Council of Emergency Medicine Residency Directors' standardized letter of recommendation: the program director's perspective. *Acad Emerg Med* 2014;21:680–7.
11. Official CORD Standardized Letter of Evaluation (SLOE) 2015-2016 application season. 2015. Available at: <https://www.cordem.org/files/DOCUMENTLIBRARY/SLOR/SLOE%20Standard%20Letter%20of%20Evaluation%202015.pdf>. Accessed May 29, 2017.
12. Beskind DL, Hiller KM, Stolz U, et al. Does the experience of the writer affect the evaluative components on the standardized letter of recommendation in emergency medicine? *J Emerg Med* 2014;46:544–50.
13. Hegarty CB, Lane DR, Love JN, et al. Council of Emergency Medicine Residency Directors standardized letter of recommendation writers' questionnaire. *J Grad Med Educ* 2014;6:301–6.
14. Love JN, Deiorio NM, Ronan-Bentle S, et al. Characterization of the Council of Emergency Medicine Residency Directors' standardized letter of recommendation in 2011-2012. *Acad Emerg Med* 2013;20:926–32.
15. Larkin GL, Marco CA. Ethics seminars: beyond authorship requirements—ethical considerations in writing letters of recommendation. *Acad Emerg Med* 2001;8:70–3.
16. Girzadas DV Jr, Harwood RC, Davis N, Schulze L. Gender and the Council of Emergency Medicine Residency Directors standardized letter of recommendation. *Acad Emerg Med* 2004;11:988–91.
17. Pennebaker JW, Boyd RL, Jordan K, Blackburn K. *The development and psychometric properties of LIWC2015*. Austin, TX: University of Texas at Austin, 2015.
18. Mehl MR. Quantitative text analysis. In: Eid M, Diener E, editors. *Handbook of Multimethod Measurement in Psychology*. Washington, DC: American Psychological Association, 2005. pp. 141–56.
19. Keim SM, Rein JA. Standardized vs narrative letters for residency applicant evaluation. *Acad Emerg Med* 1999;6:768–70.
20. Dayal A, O'Connor DM, Qadri U, Arora VM. Comparison of male vs female resident milestone evaluations by faculty during emergency medicine residency training. *JAMA Intern Med* 2017;177:651–7.