Do Students' and Authors' Genders Affect Evaluations? A Linguistic Analysis of Medical Student Performance Evaluations

Carol Isaac, PT, PhD, Jocelyn Chertoff, MD, MS, Barbara Lee, PhD, and Molly Carnes, MD, MS

Abstract

Purpose

Recent guidelines for the Medical Student Performance Evaluation (MSPE) have standardized the "dean's letter." The authors examined MSPEs for linguistic differences according to student or author gender.

Method

This 2009 study analyzed 297 MSPEs for 227 male and 70 female medical students applying to a diagnostic radiology residency program. Text analysis software identified word counts, categories, frequencies, and contexts; factor analysis detected patterns of word categories in student–author gender pairings.

Results

Analyses showed a main effect for student gender (P=.046) and a group difference for the author–student gender combinations (P=.048). Female authors of male student MSPEs used the fewest "positive emotion" words (P=.006). MSPEs by male authors were shorter than those by females (P=.014). MSPEs for students ranked in the National Resident Matching Program contained more "standout" (P=.002) and "positive emotion" (P=.001) words. There were no differences in the author–gender pairs in the proportion of students ranked, although predominant word categories differed by author and student gender. Factor analysis revealed differences among the author–student groups in patterns of correlations among word categories.

Conclusions

MSPEs differed slightly but significantly by student and author gender. These differences may derive from societal norms for male and female behaviors and the subsequent linguistic interpretation of these behaviors, which itself may be colored by the observer's gender. Although the differences in MSPEs did not seem to influence students' rankings, this work underscores the need for awareness of the complex effects of gender in evaluating students and guiding their specialty choices.

The proportion of women medical students has been over 30% for a quarter of a century.[1] Despite explicit support for gender equity in academic medicine, however, female physicians advance more slowly toward seniority and have not entered the ranks of leadership at rates predicted by their proportion in academic medicine.[2–4] The National Academies of Science[5] concluded that even in the absence of explicit bias, implicit assumptions based on gender stereotypes can subtly but systematically advantage men and disadvantage women in career advancement in academic medicine, science, and engineering. Numerous studies have confirmed that these implicit assumptions interfere with objective assessment of an applicant's qualifications in employment settings,[6] and several studies have documented differences in letters of recommendation for male and female applicants.[7–11]

Letters of recommendation are an important factor in the selection of medical students into residency programs.[12–21] The Association of American Medical Colleges (AAMC) has established guidelines for the "dean's letter" by developing the Medical Student Performance Evaluation (MSPE).[22] Because one goal of the AAMC is to standardize the MSPEs, it is important to determine whether gender has an impact on the words and descriptors used in this document. We undertook this study to determine whether the MSPE is susceptible to differences according to the gender of the student or gender of the author.

Method

The institutional review boards of the University of Wisconsin–Madison and the Dartmouth–Hitchcock Medical Center approved this study.

Sample description

We studied MSPEs of students at U.S. medical schools applying to the Diagnostic Radiology Residency Program at Dartmouth–Hitchcock Medical Center for 2009. Radiology is a competitive specialty wherein approximately 20% of all applicants are selected,[23] and this program is particularly competitive as it accepts less than 5% of applicants. Dean's offices would generally advise only top-ranked medical students to apply. The MSPE is a two- to three-page document with six sections. The sections are "Identifying Information," "Unique Characteristics," "Academic History," "Academic Progress," "Summary," and "Appendices." The MSPE is a mixture of data and prose. The "Academic Progress" section includes original comments from the student's supervisors. The "Summary" section is an assessment of the student's overall performance relative to his or her peers.

Document preparation procedures

After recording the authors' and students' gender, we removed all identifying information and reformatted the documents according to the

Linguistic Inquiry Word Count (LIWC) software manual so that they were accessible to the program.[24] We analyzed only free text, not tables or graphs. LIWC is a validated, word-count-based, text analysis program that compares words in a text document with predefined word categories.[25,26] The intent of LIWC is to assess cognitive processes in written language. To achieve internal and external validity for the default word categories, the software developers employed experts to analyze and iteratively compare hundreds of text files.[24] The program has 80 such word categories, composed of 4,500 words and word stems. It also allows users to construct additional word categories. The LIWC first counts the total number of words in a document and then counts the numbers of words in the 80 default plus any user-designed categories. LIWC presents the results as ratios of identified word categories to total words in the document. Two previous studies have used the LIWC program to analyze letters of recommendation.[7,27]

Five user-defined categories were previously created by Schmader et al[7]; these were derived from a linguistic study by Trix and Psenka[8] of gender differences in letters of recommendation for faculty in an academic medical center. We chose to add these established categories because of their pertinence to our study. These five word categories encompassed grindstone traits (e.g., hardworking, conscientious), ability traits (e.g., adept, competent), standout adjectives (e.g., exceptional, outstanding), research terms (e.g., discover, journal), and teaching terms (e.g., student, instruct).[28] We also entered the text into NVivo,[29] a program for qualitative analysis of text. NVivo enabled us to examine the occurrence of individual words within each LIWC category in context with surrounding text.

Data analysis

Of the 298 applicants to the radiology residency for 2009 (227 men and 71 women), 297 MSPEs were available for analysis: 227 for male and 70 for female students from 104 different schools. Of these students, 35 were ranked for the National Resident Matching Program (NRMP) (26 men and 9 women). The gender of the author was missing for 6 of the 297 letters, for a final sample of 291: 151 written by men and 140 written by women. All MSPEs were authored by a senior administrator. We performed a MANOVA with 85 LIWC dependent variables (the 80 default and 5 user-defined categories) and two independent variables (author and student gender) using SPSS version 17.0 software (SPSS Inc., Chicago, Illinois). We repeated this analysis for the text that is most directly under the control of the author: "Unique Characteristics," "Academic History," and "Summary." These sections together accounted for 32% of the total word count. With NVivo, we identified the 1,000 most frequent words and cross-referenced these with the words in the LIWC categories. For example, the LIWC category named "space" includes words that are relative to each other such as "above," "below," and "across." With NVivo, we found that the most frequent words in this category were "high," "level," "above," "where," and "over." For descriptive purposes, we renamed two categories to reflect actual word usage because we thought the LIWC name was misleading. Thus, the only word from the LIWC "inhibition" category was "responsible," so we renamed this category "responsible," and the word used almost exclusively from the LIWC "home" category was "family," so we refer to this category as "family." NVivo allowed us to examine individual words in context, assisting us in factor naming. Table 1 presents the final set of 18 word categories selected as dependent variables, and the discrete words identified within each category.

To probe for differences in the patterns of word categories, we performed a principal-components factor analysis with variance maximizing (varimax) rotation.[30,31] We retained factors with eigenvalues ≥1.0 and LIWC categories within factors with correlations ≥0.45.[31]

We tested for differences in NRMP-ranked and unranked students in the four author–student gender pairs (female author/male student; female author/female student; male author/male student; male author/female student) with chi-square analysis. To see whether the ranking of students for each of the four author–student pairs could be predicted based on a linear combination of LIWC categories, we performed a discriminant function analysis with student ranking (yes, no) and all LIWC categories.[30,31] We did not have information on the ranked order.

Results

The MANOVA indicated a main effect for student gender (Wilkes lambda, $P = .046$). The interaction of author gender with student gender ($P = .077$) and the main effect of author gender ($P = .071$) were not significant. We coded author and student gender into a single independent variable (sexmix) with four author–student gender combinations (with LIWC categories as the dependent variables) and confirmed overall sexmix group differences (Wilkes lambda, $P = .048$). Eleven variables were significantly related to sexmix by univariate F tests ($P < .05$) and 17 other categories of interest. Word count was significantly related to sexmix ($P = .015$), with male authors writing letters on average 209 words shorter than female authors ($P = .014$). Analysis of text from the "Unique Characteristics," "Academic History," and "Summary" sections combined showed no significant sexmix group effect, although male authors had the lowest word counts ($P = .075$). MANOVA for summary statements alone (11% of the total text) did show a significant sexmix effect (Wilkes lambda, $P = .042$).

After removing variables that were strongly related to word count or to each other, MANOVA on the final set of 18 dependent variables (Table 1) continued to show a significant effect for sexmix ($P = .026$). Univariate F tests for each of the 18 word categories found significant differences for the "positive emotion" ($P = .028$), "motion" ($P = .027$), and "space" ($P = .030$) categories. In pairwise comparisons, male students with female authors had the fewest "positive emotion" words ($P = .006$), female students with female authors had more "motion" words than male students with male authors ($P = .027$), and male students with female authors had more "space" words than male students with male authors ($P = .007$).

Factor analysis

Factor analysis identified four factors for each author–student gender group (Table 2). After close examination of the words within each category, their groupings within each factor, and the context in which they appeared, we labeled the

Table 1

Final 18 Word Categories From the Linguistic Inquiry Word Count (LIWC) Software, Examined as Dependent Variables in a Linguistic Analysis of Medical Student Performance Evaluations, Dartmouth–Hitchcock Medical Center, 2009

Word category name used by LIWC	Most frequent words in the LIWC category found in text by NVivo software
Ability	skills, ability, bright, able, analytical
Achieve	organized, performed, earned
Adverb	well, very
Cause	motivated, make, based, solving
Certain	always, complete,* extremely, clearly
Expectation	outstanding, expected, expectations, would, needs
Family[†]	family
Grindstone	hardworking, conscientious
Insight	knowledge, learn,* motivated, questions, solving
Motion	pass,* received, following, took, step, attending, advanced
Positive emotion	good, excellent, honors, eager, enthusiastic
Research	research, fund, science,* data
Responsible[†]	responsible
Space	high, level, above, where, over
Standout	excellent, outstanding, unique
Teaching	teaching, education, training
Tentative	overall, questions, most
Work	student,* excellent, work,* team

* Includes any endings with this root word.
† The original LIWC category name was misleading; these names are more representative of the actual words from the category that appeared in the text.

factors and synthesized these into general descriptions of male and female students (Figure 1). Negative correlations indicate that MSPEs with words in these categories were less likely to contain words in the other categories in the factor.

Description of male students. The synthesis of factors identified for male students was, "Works eagerly, responsibly, and above expectations toward becoming an outstanding, insightful specialist." Although "work," "responsible," "teaching," "achieve," and "standout" appeared frequently, this grouping occurred only in MSPEs for male students in the first (male authors) and second (female authors) factors. We labeled this factor "responsible exceller" because these students were typically described as "responsible, hardworking." The male author/male student group additionally had "standout" words that prompted us to assign the label "unique responsible exceller," with descriptions such as "hard worker, excellent student, very responsible" and "motivated, organized, and responsible."

The three categories of "positive emotion," "adverb," and "grindstone" were present for male students in factor 1 (female authors) and factor 2 (male authors). Because these connote enthusiasm for the students' hard work, we named this factor "eager beaver." Authors depicted these students as "hardworking, intelligent, eager to learn," "enthusiastic, well prepared, and hardworking," and "always eager for feedback." The female author–male student factor structure added "certain" words, making this the "clearly an eager beaver" factor.

The first word category in factor 3 is "expectation." Male authors paired "expectation" with "achieve," "insight," and "cause," so we assigned the label "performance above expectation," a general affirmation of intellectual motivation. One male author reported that a student "earned superior marks for his interest in learning, and ability to organize," and another indicated that "his fund of knowledge and clinical problem solving skills are above the level expected." Although similar to the

"responsible exceller" and "eager beaver" factors, the presence of "expectation" adds a dimension of potential. Female authors paired "expectation" with "standout," which we labeled "outstander" because both categories include the word "outstanding."

Factor 4 for male students contains "ability." The negative correlation with "space" words for male authors may be due to the "above average" descriptor captured in this category. We refer to this factor as "more than above average." Examples include, "[he] went above and beyond" and "[he] exceeded the expected high level." Factor 4 for female authors has a negative correlation of "family" with "ability" and "insight." For all factors, the category "family" represents the word "family" used as "family medicine" or "family practice" over 80% of the time (with the remaining 10% pertaining to the students' current families and 10% to their patients' families). We labeled this factor "insightful specialist" not only because of this negative correlation but also because some text implied that male students with "ability" and "insight" are less likely to go into family medicine. As one female author noted, "[he] really surprised us! [He] is an exceptional student [in family medicine]." Another stated, "although [he] received highest honors on [his] family medicine rotation, surely [his] finest performance was on surgery … [he] was outstanding—spoke with families, got consent forms signed, was extremely aggressive."

Description of female students. The synthesis of factors for female students was, "Works hard and enthusiastically; asks insightful questions befitting a specialist but would be exceptional in family medicine which requires less initiative and responsibility." Factor 1 (male authors) and factor 2 (female authors) contained "grindstone," "work," "positive emotion," and "teaching" variables that we labeled "enthusiastic worker bee," attesting to students' conscientious hard work. The only variation between male and female authors was that the latter included "research," and thus we named this factor "enthusiastic worker bee with research experience" (e.g., "[she was] very self-motivated; would ask questions and do

Table 2

Factor Analysis of Medical Student Performance Evaluations, by Gender of Author and Student, From a Linguistic Analysis of Medical Student Performance Evaluations, Dartmouth–Hitchcock Medical Center, 2009

Factor structure	Male authors					Female authors				
	Total	Factor 1	Factor 2	Factor 3	Factor 4	Total	Factor 1	Factor 2	Factor 3	Factor 4
Male students										
% Variance explained	44.7	13.92	12.14	10.65	7.99	48.8	15.96	11.65	9.21	8.98
Factor label		Unique responsible exceller	Eager beaver	Performance above expectation	More than above average		Clearly an eager beaver	Responsible exceller	Outstander	Insightful specialist
Categories (correlations)		Work (.859)	Positive emotion (.830)	Expectation (.725)	Ability (−.805)		Positive emotion (.850)	Work (.761)	Expectation (.889)	Family (−.773)
		Responsible (.730)	Adverb (.806)	Insight (.707)	Space (.471)		Adverb (.741)	Teaching (.662)	Standout (.662)	Ability (.729)
		Teaching (.668)	Grindstone (.675)	Cause (.563)			Grindstone (.715)	Achieve (.576)		Insight (.485)
		Achieve (.511)		Achieve (.539)			Space (−.601)	Responsible (.521)		
		Standout (.457)					Certain (.498)			
Female students										
% Variance explained	47.5	12.95	12.68	12.11	9.71	54.0	15.48	14.96	13.10	10.47
Factor label		Enthusiastic worker bee	Questioning insightful specialist	Outstanding performance as expected	Does very well but has less responsibility		Very motivated questioner	Enthusiastic worker bee with research experience	Exceptional in family medicine which requires less initiative and responsibility	Will make capable MD
Categories (correlations)		Grindstone (.787)	Tentative (.769)	Standout (.857)	Adverb (.849)		Adverb (.776)	Work (.797)	Family (.800)	Ability (.837)
		Work (.720)	Teaching (.697)	Achieve (.755)	Responsible (−.754)		Insight (.743)	Research (.792)	Motion (−.697)	Cause (.734)
		Positive emotion (.544)	Insight (.687)	Expectation (.667)			Tentative (.663)	Teaching (.623)	Standout (.629)	
		Teaching (.479)	Ability (.454)	Certain (.452)			Positive emotion (.569)	Positive emotion (.600)	Responsible (−.609)	
			Family (−.457)				Space (−.493)	Grindstone (.493)		
							Expectation (.452)			

independent research; very organized; systematic").

Factor 2 for female students with male authors is similar to the female author/ male student factor 4 "insightful specialist," with the addition of the categories "tentative" and "teaching." The variable "tentative" appeared only in MSPEs for female students (factor 1 for female authors and factor 2 for male authors), often because of the word "questions." One male author noted that a female student "asked intelligent and insightful questions." Another reported, "[she] asked challenging questions." Because "questions" is in two categories, we labeled this factor "questioning insightful specialist." In addition to "tentative," "insight," and "ability," the categories "adverb," "positive emotion," and "expectation" occur in factor 1 for female students with female authors, earning the label "very motivated questioner." Examples included "very enjoyable … asked good questions and knowledgeable"

and "[Her] preparation goes well beyond what would be expected .… [She] was affable, enthusiastic, and bright … was motivated."

Factor 3 for female students with male authors, which we labeled "outstanding performance as expected," is similar to "performance above expectation" and "outstander" in that it acknowledges the student's potential intellectual capacity. Also included in this factor are "certain" words; for example, "always completed
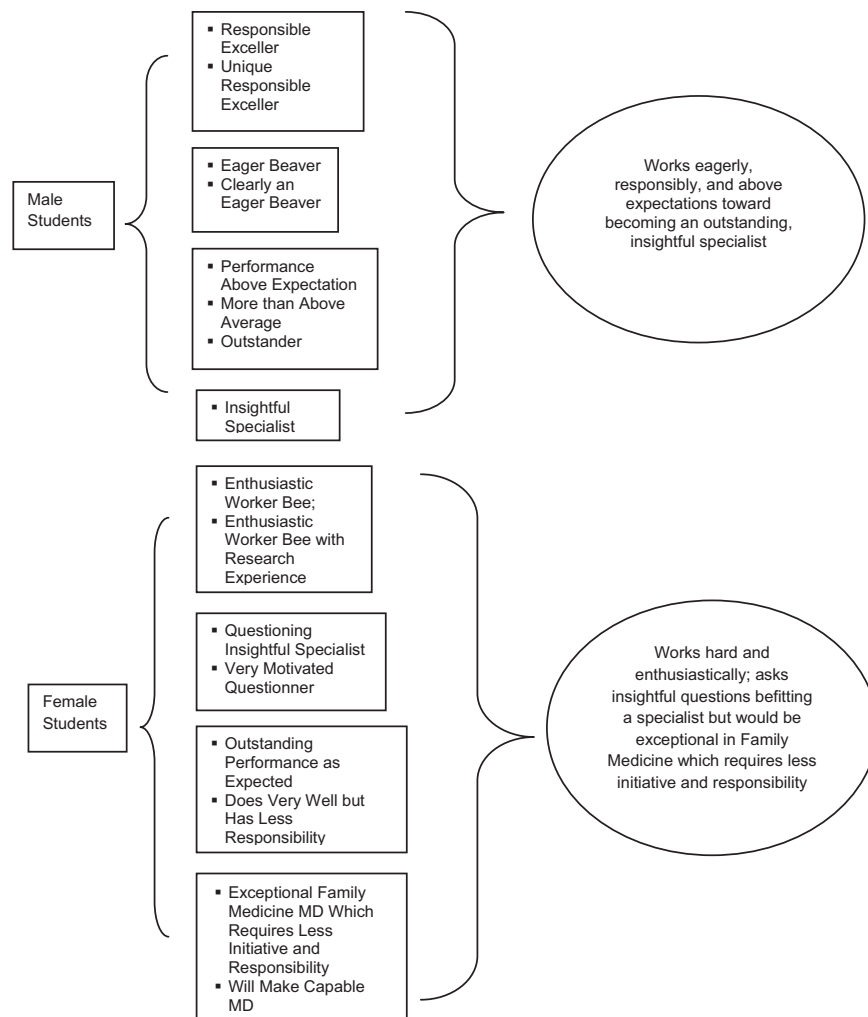
Figure 1 Factor synthesis for descriptions of male and female students, from a linguistic analysis of medical student performance evaluations, 2009.

[assignments] on time," "always well prepared," and "an extremely hard worker, always willing and interested in helping out."

Factor 4 for female students with female authors was labeled "will make capable MD" and included "ability" and "cause" words. This factor stresses ability with phrases such as "able to get the confidence of patients and staff" and "efficient, no nonsense, hardworking, appropriately confident." One female author included, "In summary, [she] has the abilities to 'play this game' of internal medicine at a very high level."

Female students had negative correlations with "responsible" (factor 4 for male authors and factor 3 for female authors), which had only positive correlations for male students. In the text, "responsible" is positive; for example, "[she was] professional, well groomed, and very

responsible and reliable" and "[she was] incredibly mature, and always responsible." However, the negative correlation indicates that female author/ female student MSPEs that referred to family medicine or included "standout" words were less likely to include "responsible." We named this factor "exceptional in family medicine which requires less initiative and responsibility." Adverbs in factor 4 for female students with male authors included words in such phrases as "very well liked and respected," and "very well organized." Because of the negative correlation with "responsibility," we named this factor "does very well but has less responsibility."

Comparison of NRMP-ranked and nonranked students

There was no difference in the proportion of those ranked for each of

the four author–student gender groups. Two categories of words appeared significantly more often in MSPEs from ranked students: "standout" ($P = .002$) and "positive emotion" ($P = .001$). Discriminant function analysis indicated that "standout" words predicted 88.5% of the 26/227 male students who were ranked ($P = .011$). Five word categories ("expectation," "tentative," "responsible," "achieve," and "positive emotion") predicted 90.0% of the 9/70 female students who were ranked ($P < .001$). "Standout" and "positive emotion" words predicted 85.4% of the 21/151 male-authored MSPEs for ranked students ($P = .004$). "Cause" words predicted 90% of the 14/140 female-authored MSPEs for ranked students ($P = .007$).

Discussion

Our detailed analysis of MSPEs for applicants to a competitive diagnostic radiology residency found differences in word usage and patterns of descriptors by the gender of the author as well as the student. Major differences (List 1) were shorter MSPEs from male authors, fewer "positive emotion" words used by female authors for male students, more "motion" words used by female authors for female students, and more "space" words used by female authors for male students. MSPEs for NRMP-ranked students contained more "standout" and "positive emotion" words. The proportion of ranked students compared with unranked students in any of the four author–student gender groups was not different. However, in the discriminant function analysis (which begins with the word categories and essentially performs a MANOVA backwards), "standout" words predicted the MSPEs only for ranked male students. For ranked female students, five other categories ("expectation," "tentative," "responsible," "achieve," and "positive emotion") predicted the MSPEs. Via factor analysis, we identified different structures for MSPEs in the four author– student gender pairings. We discuss our findings in the context of research on gender and previous studies of letters of recommendation.

Research on gender

Societal norms exist for male and female behavior across multiple domains.[32,33]

List 1

Summary of Significant Differences in Medical Student Performance Evaluations (MSPEs) Written by Male and Female Authors for 291 Students, From a Linguistic Analysis of MSPEs, Dartmouth–Hitchcock Medical Center, 2009

- Male authors wrote shorter letters than female authors.
- Male students with female authors had the fewest "positive emotion" words.
- Female students with female authors had more "motion" words than did male students with male authors.
- Male students with female authors had more "space" words than those with male authors.
- Factor structure* of word categories differed among the four author–student gender groups:
  - Only female students had "tentative" words.
  - Female students had a negative correlation with "responsible," whereas male students had a positive correlation.
  - Male students with male authors had an absence of "family" words.
  - Only female students with female authors had a positive correlation with "family" words (usually family medicine or family practice).
  - Female authors for male students and male authors for female students had strong emphasis ("certain") words, which were absent from the other two groups.
- MSPEs from ranked students had more "standout" and "positive emotion" words than those from unranked students, but there was no difference in the proportion ranked from any author–student gender group.
- Discriminant function analysis of ranked versus unranked MSPEs found that
  - "Standout" words predicted male MSPEs.
  - "Expectation," "tentative," "responsible," "achieve," and "positive emotion" words predicted female MSPEs.
  - "Standout" and "positive emotion" words predicted MSPEs by male authors.
  - "Cause" words predicted MSPEs by female authors.

* Factors with eigenvalues ≥1.0 and Linguistic Inquiry Word Count software categories within factors with correlations ≥0.45.

These gender norms influence the way men and women act and how their actions are interpreted by others,[34,35] creating several ways in which gender could influence MSPEs.

Male and female medical students behave differently. In keeping with societal norms, male and female medical students comport themselves differently (e.g., in dress, communication styles, and interpersonal interactions). Those who violate these norms perhaps risk lower evaluations.[32,36,37] The MSPEs may reflect these differences, as male students generally were "responsible excellers" and "eager beavers" who "perform above expectation," and female students generally were "enthusiastic worker bees" and "motivated questioners." Carli et al[38] and others[37,39] have found that women are most influential when they deliver their messages in a manner that is not overly directive or assertive. Framing content as a question or using tentative language are ways to accomplish this linguistically.[40] Top female medical students have likely learned to balance assertiveness with tentativeness to positively influence the opinions of their

performance, as "tentative" words appeared only in the factor structure for female student MSPEs.

The negative correlation of "responsible" in two factors for female students is difficult to interpret. The specific text of the various MSPEs we evaluated does not suggest that female students were irresponsible, but this negative correlation was not seen for male students. Perhaps female medical students actually were less available than male students[41]; alternatively, the assumptions about their gendered roles may have led evaluators to interpret their behaviors as such.[42]

Male and female authors interpret male and female student behavior differently. The interpretation of similar behaviors may differ according to student gender (e.g., she "showed enthusiasm" versus he "took initiative"). The use by female authors of fewer "positive emotion" words in describing male students, more "motion" words for female students, and more "space" words for male students may reflect this. Research by Biernat and Eidelman[43] suggests that the use of

different terms to describe similar behaviors for male and female applicants may be beneficial. They found that when equivalent language was used in a letter of recommendation to describe male and female students in a traditionally masculine domain, raters translated those letters into less favorable judgments of qualifications for female than for male applicants.[43]

Gender-stereotyped assumptions could account for the factor structure involving family medicine. Family medicine may be viewed as more appropriate for women physicians: It is a relatively nurturing specialty, of comparatively low status, with a high proportion of female residents.[44] Whether these facts conspire to eliminate family medicine from the factor structure of the male-authored letters for male students or account for the positive correlation between family medicine and "standout" words when female authors write female student MSPEs is speculative but consistent. When strings of positive words appeared, male authors did not associate these with family medicine for female students, and female authors did not associate these with family medicine for male students, consistent with expectations of gender and status.[34] These findings could belie subtle but clearly different social messages given to male and female medical students throughout their training. Such socialization may contribute to the overrepresentation of women in low-status, primary care specialties relative to higher-status surgical and medical subspecialties.[44]

Henderson et al[9] and Eger[10] have described an "advocacy factor" in which authors wrote better letters of recommendation for people like themselves in personality, ideology, and gender. We found mixed support for an advocacy factor in our study. In support, female authors used more "positive emotion" words for female students than for male students. As evidence against an advocacy factor, men wrote the shortest letters for male students, and only in the cross-gender pairings of author and student do we see "certain" words that add emphasis to positive words.

Different linguistic patterns for male and female authors. The clearest difference we found between male and female authors was length of MSPEs;

male authors wrote shorter letters. Consistent with our findings, Watson[11] has found that the shortest letters of recommendation for graduate school applicants were by male authors for male applicants, and Henderson et al[9] have found that female authors wrote longer letters of recommendation than male authors. Trix and Psenka[8] have found that male authors wrote shorter letters of recommendation for female applicants for faculty positions, whereas Schmader et al[7] found no difference. The importance of letter length to the success of an applicant remains unclear.[19,45]

Other studies of letters of recommendation

Residency program directors value letters of recommendation in ranking students for admission to residency programs.[12–14,46,47] Because the MSPE is an attempt to standardize the dean's letter, our study is of particular interest. Our findings suggest that gender can override attempts at standardization of medical student performance and assessment. Trix and Psenka[8] found more adjectives about hard work ("grindstone" words) in letters written for female than male faculty members. They interpreted this as suggesting that if a woman is performing at a high level in a traditionally male field, she must be working very hard. Differences in descriptions of hard work between the male student "eager beavers" and the female student "worker bees" in our study might also warrant this interpretation.

Limitations

The gender of the authors of the text imported from rotation evaluations in the MSPE is unknown to us, leaving unanswered how much of the student gender difference is due to what supervising physicians in clerkships wrote. However, the significant difference in word count and the group difference in the summary statement suggest that the gender of the senior administrator compiling the MSPE did influence the final product. Of the 297 MSPEs we examined, 35 were for NRMP-ranked students, but we do not know the rank order or the actual academic achievements of the students. Finally, this study is limited to one radiology program.

Conclusions

Our linguistic analysis of MSPEs found differences attributable to both student and author gender. These did not seem to be correlated with whether students were ranked. An implicit socialization of female students toward family medicine merits further study. Our research underscores the complex effects of gender in the performance and evaluation of medical students, and more attention should be paid to the possible impact of gender in preparing evaluations and letters of recommendation for medical students.

Dr. Isaac is a research associate, University of Wisconsin–Madison Center for Women's Health Research, Madison, Wisconsin.

Dr. Chertoff is vice chair, Department of Diagnostic Radiology, program director, Diagnostic Radiology Residency, director, Section of Gastrointestinal Radiology, and assistant medical director, Medical Staff Affairs, Dartmouth–Hitchcock Medical Center; assistant dean for clinical affairs, Dartmouth Medical School; and associate professor, Radiology and Obstetrics and Gynecology, Dartmouth–Hitchcock Medical Center, Lebanon, New Hampshire.

Dr. Lee is a consultant in statistics and evaluation from Tarpon Springs, Florida, retired research associate, University of South Florida, Louis de la Parte Florida Mental Health Institute, Tampa, Florida, and school psychologist, Hillsborough County Public Schools, Tampa, Florida.

Dr. Carnes is director, Center for Women's Health Research, University of Wisconsin–Madison and Meriter Hospital; professor, Department of Medicine and Psychiatry, School of Medicine and Public Health; professor, Department of Industrial and Systems Engineering, College of Engineering, University of Wisconsin–Madison, Madison, Wisconsin; and director, Women Veterans Health Program, William S. Middleton Memorial Veterans Hospital, Madison, Wisconsin.

References

1 Magrane D, Jolly P. The changing representation of men and women in academic medicine. AAMC Analysis in Brief. July 2005;5.

2 Wright AL, Schwindt LA, Bassford TL, et al. Gender differences in academic advancement: Patterns, causes, and potential solutions in one U.S. college of medicine. Acad Med. 2003;78:500–508. http://journals.lww.com/academicmedicine/Fulltext/2003/05000/Gender_Differences_in_Academic_Advancement_.15.aspx. Accessed September 30, 2010.

3 Carnes M, Morrissey C, Geller S. Women's health and women in academic medicine: Hitting the same glass ceiling? J Womens Health (Larchmt). 2008;17:1453–1462.

4 Ash AS, Carr PL, Goldstein R, Friedman RH. Compensation and advancement of women in academic medicine: Is there equity? Ann Intern Med. 2004;141:205–212.

5 National Academy of Sciences; National Academy of Engineering; Institute of Medicine of the National Academies. Beyond Biases and Barriers: Fulfilling the Potential of Women in Academic Science and Engineering. Washington, DC: National Academies Press; 2006.

6 Isaac C, Lee B, Carnes M. Interventions that affect gender bias in hiring: A systematic review. Acad Med. 2009;84:1440–1446. http://journals.lww.com/academicmedicine/Fulltext/2009/10000/Interventions_That_Affect_Gender_Bias_in_Hiring__A.36.aspx. Accessed September 30, 2010.

7 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;57:509–514.

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

9 Henderson J, Briere J, Hartsough R. Sexism and sex roles in letters of recommendation to graduate training in psychology. Can Psychol. 1980;21:75–80.

10 Eger H. Writer Perception, Writer Projection: The Influence of Personality, Ideology, and Gender on Letters of Recommendation [dissertation]. Chicago, Ill: University of Illinois; 1991.

11 Watson C. Sex-linked differences in letters of recommendation. Women Lang. 1987; 10:26–28.

12 Wagoner NE, Suriano JR. A new approach to standardizing the dean's letter. Acad Med. 1989;64:688–689. http://journals.lww.com/academicmedicine/Abstract/1989/11000/A_new_approach_to_standardizing_the_dean_s_letter.19.aspx. Accessed September 30, 2010.

13 Wagoner NE, Suriano JR. Program directors' responses to a survey on variables used to select residents in a time of change. Acad Med. 1999;74:51–58. http://journals.lww.com/academicmedicine/Abstract/1999/01000/Program_Directors__Responses_to_a_Survey_on.23.aspx. Accessed September 30, 2010.

14 Wagoner NE, Suriano JR, Stoner JA. Factors used by program directors to select residents. J Med Educ. 1986;61:10–21.

15 Clemente M, Michener MW. The dean's letter of recommendation and internship performance. J Med Educ. 1976;51:590–592.

16 Delisa JA, Jain SS, Campagnolo DI. Factors used by physical medicine and rehabilitation residency training directors to select their residents. Am J Phys Med Rehabil. 1994;73:152–156.

17 Johnson M, Elam C, Edwards J, et al. Medical school admission committee members' evaluations of and impressions from recommendation letters. Acad Med. 1998;73(10 suppl):S41–S43. http://journals.lww.com/academicmedicine/Fulltext/2001/06000/Letters_of_Recommendation__Comparing_Perceptions.20.aspx. Accessed September 30, 2010.

18 DeZee KJ, Thomas MR, Mintz M, Durning SJ. Letters of recommendation: Rating, writing, and reading by clerkship directors of internal medicine. Teach Learn Med. 2009;21:153–158.

19 Greenburg AG, Doyle J, McClure DK. Letters of recommendation for surgical residencies: What they say and what they mean. J Surg Res. 1994;56:192–198.

20 Lee AG, Golnik KC, Oetting TA, et al. Re-engineering the resident applicant selection process in ophthalmology: A literature review and recommendations for improvement. Surv Ophthalmol. 2008;53:164–176.

21 O'Halloran C, Altmaier E, Smith W, Franken E. Evaluation of resident applicants by letters of recommendation: A comparison of traditional and behavior-based formats. Invest Radiol. 1993;28:274–277.

22 Association of American Medical Colleges Dean's Letter Advisory Committee. A Guide to the Preparation of the Medical Student Performance Evaluation, 2002. https://www.aamc.org/linkableblob/64496-6/data/mspeguide-data.pdf. Accessed November 8, 2010.

23 National Resident Matching Program. Results and Data: 2010 Main Residency Match. Washington, DC: Association of American Medical Colleges; 2010.

24 Linguistic Inquiry and Word Count [computer program and manual]. Mahwah, NJ: Lawrence Erlbaum; 2007.

25 Mehl MR, Eid M, Diener E. Quantitative Text Analysis. Handbook of Multimethod Measurement in Psychology. Washington, DC: American Psychological Association; 2006:141–156.

26 Kahn JH, Tobin RM, Massey AE, Anderson JA. Measuring emotional expression with the Linguistic Inquiry and Word Count. Am J Psychol. 2007;120:263–286.

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

28 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;57:509–514.

29 Richards L. Handling Qualitative Data: A Practical Guide. Repr ed: London, UK: Sage Publications; 2006.

30 Stevens J. Applied Multivariate Statistics for the Social Sciences. 4th ed. Mahwah, NJ: Lawrence Erlbaum Associates; 2002.

31 Grimm LG, Yarnold PR, eds. Reading and Understanding More Multivariate Statistics. Washington, DC: American Psychological Association; 2000.

32 Valian V. Why So Slow? The Advancement of Women. Cambridge, Mass: MIT Press; 1998.

33 Ridgeway CL. Gender, status, and leadership. J Soc Issues. 2001;57:637–655.

34 Ridgeway CL, Bourg C. Gender as status: An expectation states theory approach. In: Eagly AH, Beall AE, Sternberg RJ, eds. The Psychology of Gender. 2nd ed. New York, NY: Guilford Press; 2004:217–241.

35 Carli LL. Gender, interpersonal power, and social influence. J Soc Issues. 1999;55:81–99.

36 Heilman M. Description and prescription: How gender stereotypes prevent women's ascent up the organizational ladder. J Soc Issues. 2001;57:657–674.

37 Rudman LA. Self-promotion as a risk factor for women: The costs and benefits of counterstereotypical impression management. J Pers Soc Psychol. 1998;74:629–645.

38 Carli LL, LaFleur SJ, Loeber CC. Nonverbal behavior, gender, and influence. J Pers Soc Psychol. 1995;68:1030–1041.

39 Phelan JE, Moss-Racusin CA, Rudman LA. Competent yet out in the cold: Shifting criteria for hiring reflect backlash toward agentic women. Psychol Women Q. 2008;32:406–413.

40 Carli LL. Gender, language, and influence. J Pers Soc Psychol. 1990;59:941–951.

41 Brown AJ, Swinyard W, Ogle J. Women in academic medicine: A report of focus groups and questionnaires, with conjoint analysis. J Womens Health. 2003;12:999–1008.

42 Foschi M. Double standards for competence: Theory and research. Annu Rev Sociol. 2000;26:21–42.

43 Biernat M, Eidelman S. Translating subjective language in letters of recommendation: The case of the sexist professor. Eur J Soc Psychol. 2007;37:1149–1175.

44 Leadley J. Women in U.S. Academic Medicine Statistics and Benchmarking Report 2008–2009. Washington, DC: Association of American Medical Colleges; 2009.

45 Bosk C. Resident evaluation: A sociological perspective. In: Bartlett RH, ed. Medical Education: A Surgical Perspective. Chelsea, Mich: Lewis; 1986:267.

46 Wagoner NE, Suriano JR. Recommendations for changing the residency selection process based on a survey of program directors. Acad Med. 1992;67:459–465. http://journals.lww.com/academicmedicine/Abstract/1992/07000/Recommendations_for_changing_the_residency.11.aspx. Accessed September 30, 2010.

47 Wagoner NE, Gray GT. Report on a survey of program directors regarding selection factors in graduate medical education. J Med Educ. 1979;54:445–452.