Key Word Use in Letters of Recommendation for Ophthalmology Residency Applicants According to Race, Gender, and Achievements

Sahil Aggarwal, BS1 Seanna Grob, MD, MAS2 Dhruva Banerjee, BA3 Preston J. Putzel, BS4 Jeremiah Tao, MD, FACS2

1 School of Medicine, University of California, Irvine, California
2 Division of Oculofacial Plastic and Orbital Surgery, Department of Ophthalmology, Gavin Herbert Eye Institute, University of California, Irvine, California
3 Department of Neurobiology and Behavior, University of California, Irvine, California
4 Donald Bren School of Information and Computer Sciences, University of California, Irvine, California

Address for correspondence Jeremiah Tao, MD, FACS, Division of Oculofacial Plastic and Orbital Surgery, Department of Ophthalmology, Gavin Herbert Eye Institute, Irvine School of Medicine, University of California, 850 Health Sciences Road, Irvine, CA 92697 (e-mail: jeremiat@uci.edu).

Keywords
- letters of recommendation
- ophthalmology
- bias
- gender
- underrepresented minority
- Alpha Omega Alpha

Abstract

Objectives To identify differences in letters of recommendation (LORs) of applicants to a single ophthalmology residency program by gender, race, academic performance, and match outcome.

Design This was a retrospective analysis of LORS for 2,523 applicants (7,569 letters) to the University of California, Irvine ophthalmology residency program from 2011 to 2018.

Methods Programming scripts were employed to determine the number of times 22 key words from four thematic categories (standout words, ability, grindstone, and compassion) appeared in LORs for each applicant. A chi-square test was performed to assess for possible differences in the presence of each key word by the following characteristics: gender, underrepresented minority (URM) status, Alpha Omega Alpha (AOA) membership, the United States Medical Licensing Exam (USMLE) Step 1 score, and match outcome. Linear regressions were created to determine the frequency at which words in each thematic category appeared according to the same baseline characteristics.

Results In the LORs, females were more likely to be described as “empathetic” ($p = 0.002$), URMs were more likely to be described as “caring” ($p = 0.002$), high Step 1 scorers ($\geq 240$) were more likely to be described as “outstanding” ($p = 0.002$), and matched students were more likely to be described as “exceptional” ($p = 0.001$), “outstanding” ($p < 0.001$), and “superb” ($p = 0.001$). Standout words appeared more often in the LORs of AOA members, matched candidates, and high Step 1 scorers ($p < 0.001$ for all comparisons). “Competent” appeared more commonly in LORs for low Step 1 scorers ($p < 0.001$) and unmatched applicants ($p = 0.001$).

Conclusion This study identifies differences in LORs by gender, URM status, and achievement including successful ophthalmology residency match. Females and URMs were more likely to be described as “empathetic” and “caring,” respectively; otherwise,
we detected no gender or racial disparities in key word use in LORs. Candidates with high USMLE Step 1 scores or AOA membership had a higher frequency of standout words in their LORs. Whether they were truly more qualified in various dimensions or if they benefited from a halo effect bias warrants further investigation. There was a significant difference in the number of standout words in LORs between matched and unmatched applicants, suggesting that key word frequency may be a relevant metric for LOR appraisal.

Securing an ophthalmology residency position in the United States is a highly selective process.\textsuperscript{1,2} Factors commonly cited as contributors to matching include the United States Medical Licensing Exam (USMLE) scores, grades, research experiences, academic achievements, Medical Student Performance Evaluations (MSPEs, also known as the dean’s letter), and letters of recommendation (LORs).\textsuperscript{3,4}

LORs provide important supplementary information about applicants, such as personality traits, professionalism, and interpersonal skills. However, the usefulness of LORs in the resident selection process has been questioned due to a lack of standardization of content. As part of a 3-year pilot program introduced by the Association of University Professors of Ophthalmology (AUPO), letter writers in the 2018 application cycle had the option to complete a standardized LOR form for candidates as an alternative to a formal LOR.\textsuperscript{5}

This LOR requires writers to rank an applicant in comparison to peers for several competencies.\textsuperscript{6}

In addition to lack of standardization, LORs may suffer from bias that may disadvantage some groups. Studies conducted in different academic settings identified significant differences in LOR content by gender.\textsuperscript{7–10} In a large retrospective study of 6,000 applicants to 16 Yale residency programs, Ross et al identified key differences in words such as “exceptional,” “best,” and “outstanding” in MSPEs by race and gender.\textsuperscript{11} Isaac et al identified similar biases in MSPEs of diagnostic radiology applicants.\textsuperscript{12}

Besides gender and race, other biases in residency applications may create an unlevel playing field in the match process. USMLE Step 1 performance and Alpha Omega Alpha (AOA) academic honor society membership have been shown to carry disproportionate weight in selection.\textsuperscript{4} The extent to which these academic factors influence LOR content has not been studied.

The halo effect is a cognitive bias in which an assessment of a person, place, or thing assumes ambiguous information from concrete information.\textsuperscript{13,14} High academic achievers may also have strong interpersonal and professionalism qualities; however, the senior author (J. T.), who has read LORs for over a decade, hypothesizes that through the halo effect, candidates with high board scores receive positive descriptive words for competencies beyond what written examinations assess. AOA membership may similarly confer additional praise through this heuristic that may be especially common for medical students who often have very limited time and context with their letter writers.

In this study, we evaluate the ophthalmology residency LORs of candidates to a single institution over eight application cycles to explore differences in key word appearance and frequency by gender, underrepresented minority (URM) status, academic achievements, and match outcome.

Materials and Methods
This retrospective cohort study was granted exempt status by the University of California, Irvine (UCI) Institutional Review Board; therefore, no informed consent was required. From 2011 to 2018, a total of 2,524 allopathic (MD) and osteopathic (DO) candidates applied to the UCI ophthalmology residency program and were reviewed for this study. Applicants who did not provide USMLE Step 1 scores were removed from analyses. Applications of those who were reapplying after a prior year(s) unsuccessful match were included in analyses as independent applications due to presumed differences in LOR content by year of submission.

A schematic of our methods is provided in Fig. 1. Each application was exported from the SF Match as a Portable Document Format (PDF) file. Python (Python Software Foundation) was used to deidentify all applications (including identifiers for both the applicants and letter writers) and to write scripts for automated extraction of baseline characteristics (USMLE Step 1 and Step 2 scores, URM status, and AOA status). The 2018 match was the first time applicants reported URM status, and all statistical analyses associated with URM were confined to this group only.

At the time of application submission, AOA status was not reported for 1,092 applicants, meaning that they either did not have an AOA chapter at their school or their AOA status was not yet determined. To ensure an accurate depiction of the role of AOA status on letter content, only those applications with definitive reporting of yes AOA status (n = 1,431) at the time of application were included in the analyses. Adobe Acrobat Professional’s Optical Character Recognition software (Adobe Systems Incorporated) was used to convert all scanned pages of each PDF (including LORs) into searchable text. For processing of the searchable LORs, several Python programming scripts were written to extract key words. Using methodology as previously described, 22 key words from four thematic categories were investigated (Table 1).\textsuperscript{11,15} These scripts reviewed each LOR line by line for the key words, and the output data were the number of times each word appeared in all three LORs for each
applicant. The scripts also accounted for different forms of each key word, including “compassionate” ("compassion), "diligent" ("diligence"), "empathy" ("empathetic"), and "talented" ("talent"). Because gender is not explicitly reported in each application, gender of each applicant was determined based on the frequency of the personal pronouns “he” and “she” in the LORs.

The final database was deidentified in Python and imported into the statistics package R (The R Foundation for Statistical Computing, Vienna, Australia) for analysis. Two-sample t-tests were completed to compare USMLE Step 1 and Step 2 scores by gender, URM status, AOA status, and match outcome. A Pearson’s chi-square test was performed to compare AOA status by gender, URM status, and match status. An α level of 0.05 was used to assess for significance in these analyses.

Pearson’s chi-square test was performed to assess possible differences in the presence of each key word by gender, URM, AOA status, USMLE Step 1 score, and match status. For this analysis, the median Step 1 score (240) was used to separate applicants into two groups: "low" scores and "high" scores. For this analysis, a Bonferroni correction was applied to account for multiple comparisons (corrected α = 0.05/22, or 0.002). We also performed linear regressions to predict the number of times each thematic category appeared in LORs from the applicant’s gender, URM status, AOA status, USMLE Step 1 score, and match status, again applying a Bonferroni correction (corrected α = 0.05/20, or 0.0025).

### Results

Of 2,524 applicants, 1 did not report a USMLE Step 1 score and was removed from the study. A total of 2,523 applicants were included in the final analyses. Females comprised 42.2% (n = 1,064) of all applicants. The mean USMLE Step 1 score was 238.3 ± 15.8 (median: 240), and the mean USMLE Step 2 score was 243.6 ± 14.4 (median: 246). Among the sample, 479 (19%) were elected into AOA and 1,092 (43.3%) reported either no AOA chapter at their school or unknown AOA status at the time of submission. Of the 387 applicants in the 2018 match cycle who could provide their URM status, 40 (10.3%) self-classified themselves as URM

<table>
<thead>
<tr>
<th>Thematic category</th>
<th>Words examined</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standout words</td>
<td>Exceptional, best, outstanding, superb, stellar, excellent, phenomenal</td>
</tr>
<tr>
<td>Ability</td>
<td>Intelligent, bright, talent, brilliant, competent, smart, gifted</td>
</tr>
<tr>
<td>Grindstone</td>
<td>Organized, hardworking, conscientious, diligent</td>
</tr>
<tr>
<td>Compassion</td>
<td>Caring, kind, empathy, compassionate</td>
</tr>
</tbody>
</table>
members (p < 0.001 for Steps 1 and 2), and unmatched applicants scored significantly lower than matched applicants (p < 0.001 for Steps 1 and 2) ({\textit{Fig. 2}}).

The distribution of each of the 22 key words and 4 thematic categories of words are shown in {\textit{Fig. 3}}; all distributions were heavily positively skewed. Chi-square testing showed differences in the presence of key words in LORs by gender, URM status, USMLE Step 1 score, and match outcome. Females were more likely to be described as "empathetic" (p = 0.002), URMs were more likely to be described as "caring" (p = 0.002), high Step 1 scorers (\(\geq 240\)) were more likely to be described as "outstanding" (p = 0.002), and matched students were more likely to be described as "exceptional" (p = 0.001), "outstanding" (p < 0.001), and "superb" (p = 0.001). "Competent" appeared more commonly in low Step 1 scorers (p < 0.001) and unmatched applicants (p = 0.001) ({\textit{Table 3}}).

The results of the linear regression revealed only the applicant's AOA status, Step 1 score, and match status were predictive of the number of times key words appeared in LORs ({\textit{Table 4}}). Specifically, standout words appeared significantly more often in the LORs of AOA members than nonmembers, matched than unmatched applicants, and "high" versus "low" Step 1 scorers (p < 0.001 for all comparisons). There was a trend toward female applicants being described with terms from the compassionate category (p = 0.009), but this effect failed to cross significance according to our preset criteria corrected for multiple comparisons (p < 0.0025). URM status did not predict the frequency at which each category of key words appeared in LORs.

**Discussion**

Ophthalmology is among the most competitive specialties in the United States, with most applicants achieving high board scores, several research publications, and excellent clinical grades.\(^1\) Comparing applicants is a difficult task due to few objective metrics beyond USMLE board scores. Additionally, several qualifications in the applications are linked, such as AOA status and high USMLE score. This study of a large set of residency applications suggests that there may exist other interdependencies between application components. In particular, halo effects and other biases may extend into LORs. Moreover, race and gender may affect LOR content.

Our analysis of the baseline characteristics of this sample found significant differences in USMLE Step 1 scores by URM status, gender, AOA status, and match outcome. A high USMLE score is frequently a prerequisite for induction into AOA at most medical schools\(^16\); therefore, the finding that members scored significantly higher than nonmembers is not surprising. Significant difference in USMLE Step 1 scores by match outcome was evident in this cohort, as previously described.\(^4\) Though a small difference, women scored significantly lower on USMLE Step 1 than men, a finding that has been previously documented.\(^17\) This difference in scores may be attributed to differences in educational backgrounds. More men pursue undergraduate degrees in basic sciences than women, and the USMLE is geared toward the basic sciences.\(^17,18\) Interestingly, despite this difference in mean USMLE Step 1 scores, there were significantly more women inducted into AOA than men in our sample. URMs scored significantly lower than non-URMs on the USMLE, a finding consistent with previous studies.\(^19,20\) Disparities in financial and other resources likely play a major contributory role; some studies suggest baseline cognitive differences, but the reasons remain not completely clear.\(^21\)

We identified differences in key words used to describe URMs and females. URMs were more commonly described as "caring," and females were more commonly described as "empathetic." However, there was no difference in the frequency at which the categories of key words (standout words, ability, grindstone, or compassion) appeared by URM status or gender. Overall, these data corroborate previous studies documenting differences in LOR and MSPE content by race and gender.\(^10–12\) The relevance may be negligible, as "empathetic" and "caring" were not more or less likely to appear in the LORs of matched applicants.

An interesting finding was that the key word "competent" was seen more commonly in LORs of unmatched applicants.
and in those with low Step 1 scores, suggesting that it may not be a good indicator of the strength of an applicant. Ross et al report that Black applicants were more likely to be described as “competent.” Our findings may corroborate their suspicions about racial discrimination by evaluators.

Conversely, LORs of high Step 1 scorers were more likely to contain the word “outstanding,” and matched students were more likely to be described as “exceptional,” “outstanding,” and “superb.” Standout words also appeared at a higher frequency in the LORs of matched versus unmatched applicants and AOA members versus nonmembers. Standout word frequency also correlated directly with USMLE Step 1 score. We deduce that there may be a halo effect in which a high Step 1 score or AOA induction may earn a candidate positive key words beyond what these achievements define. We also acknowledge the possibility of the alternative theory that those applicants who are top academic performers are also strong in other dimensions. Further research is warranted.

LORs offer information regarding an applicant that examination scores and clinical grades cannot reveal, including personality traits, work ethic, and interactions with patients. Despite their subjective content, strong LORs have been shown to correlate well with both faculty evaluations and examination scores during residency. At face value, nearly all LORs to residency programs are positive in nature, owing to applicant selection of individuals who they believe will write strong LORs. Furthermore, there is no standardization of LOR length or content, creating difficulties in differentiating the strength of LORs, which is the likely basis for the AUPO’s pilot standardized LOR initiative. This study showed that standout words appeared significantly more often in matched applicants compared with unmatched applicants. Key word frequency may offer some objectivity for appraising LORs. It may also offer the ability to stratify LORs from the same letter writer.

This study has several limitations. While the self-reported application data are meant to be truthful, we cannot confirm

---

Fig. 2 (A) Mean United States Medical Licensing Exam (USMLE) Step 1 score by baseline characteristics. (B) Proportion of applicants who are Alpha Omega Alpha (AOA) by baseline characteristic. p-Values with asterisks represent statistically significant differences (p < 0.05). URM, underrepresented minority.
the accuracy of the information (such as URM status) each applicant provided. Almost half (1,092) of all applicants in our study did not report their AOA status in their application, meaning that they either did not have an AOA chapter at their school or their AOA status was undetermined at the time of submission. Our study was unable to distinguish which of the two reasons AOA status was not reported for each applicant. However, 96% of all U.S. medical schools have an AOA chapter; therefore, it is likely that most of these candidates simply did not know their AOA status when they submitted their applications. Our study revealed that AOA status may influence descriptive words found in LORs; hence, late induction may disadvantage some members. We acknowledge that many of the 1,092 candidates were able to update their AOA status before selecting their LORs and submitting their applications.
Table 3  Chi-square results of the presence of descriptive words by baseline characteristic

<table>
<thead>
<tr>
<th>Word categories</th>
<th>Gender</th>
<th>URM</th>
<th>USME Step 1</th>
<th>AOA member</th>
<th>Match outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Male (%)</td>
<td>Female (%)</td>
<td>p-Value</td>
<td>Yes (%)</td>
<td>No (%)</td>
</tr>
<tr>
<td>Standout words</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Exceptional</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Best</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Outstanding</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Superb</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Excellent</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phenomenal</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ability</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intelligent</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bright</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Talented</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Brilliant</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Competent</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Smart</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gifted</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grindstone</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Organized</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hardworking</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Conscientious</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diligent</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Compassion</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Abbreviations: AOA, Alpha Omega Alpha; URM, underrepresented minority; USMLE, United States Medical Licensing Exam.
Note: proportions are rounded to the nearest percent. Bold-faced p-values represent statistically significant results (corrected α level of significance is 0.002).
their AOA status after submitting their applications as they were inducted. However, owing to inconsistencies in timing and method by which AOA was reported, 1,092 candidates were excluded from analyses of AOA.

Our study also reflects only the applicants to the ophthalmology program at the UCI Gavin Herbert Eye Institute and not the national cohort of annual applicants, making it difficult to extrapolate our results to the entire applicant pool. Also, our data abstraction methods using programming script may miss other positive descriptive words and do not capture subtle intonations and other semantics that a reader may glean from reading an LOR in its entirety. Finally, URM status was only reported in the 2018 application cycle, and may glean from reading an LOR in its entirety. Finally, URM status was only reported in the 2018 application cycle, and as such, our LOR analysis was therefore limited to 387 applicants. As such, our LOR analysis was therefore limited to 387 applicants. As future applicants provide their URM status, revisiting these data from a larger sample would be valuable.

In summary, LORs remain one of the most difficult aspects of the application to assess in an objective, meaningful fashion. Biases and halo effects may further confound LORs. An algorithmic approach looking for the presence and frequency of key words may offer insights into the relevance of LORs as well as tendencies of letter writers.

Table 4 Linear regression results

<table>
<thead>
<tr>
<th>Independent variable</th>
<th>Standout Regression coefficient</th>
<th>p-Value</th>
<th>Ability Regression coefficient</th>
<th>p-Value</th>
<th>Grindstone Regression coefficient</th>
<th>p-Value</th>
<th>Compassion Regression coefficient</th>
<th>p-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Step 1 score</td>
<td>0.02</td>
<td>&lt;0.001</td>
<td>0.001</td>
<td>0.256</td>
<td>&lt;0.001</td>
<td>0.852</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gender</td>
<td>-0.30</td>
<td>0.012</td>
<td>-0.11</td>
<td>0.012</td>
<td>0.01</td>
<td>0.851</td>
<td>-0.11</td>
<td>0.009</td>
</tr>
<tr>
<td>URM status</td>
<td>0.07</td>
<td>0.909</td>
<td>-0.17</td>
<td>0.391</td>
<td>-0.08</td>
<td>0.584</td>
<td>0.05</td>
<td>0.805</td>
</tr>
<tr>
<td>AOA</td>
<td>0.59</td>
<td>&lt;0.001</td>
<td>-0.02</td>
<td>0.670</td>
<td>0.05</td>
<td>0.217</td>
<td>0.02</td>
<td>0.648</td>
</tr>
<tr>
<td>Match status</td>
<td>0.95</td>
<td>&lt;0.001</td>
<td>0.03</td>
<td>0.519</td>
<td>0.01</td>
<td>0.728</td>
<td>0.02</td>
<td>0.624</td>
</tr>
</tbody>
</table>

Abbreviations: AOA, Alpha Omega Alpha; URM, underrepresented minority.
Note: Bold-faced p-values represent statistically significant results (corrected α level of significance is 0.0025).

References
11 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(08):e0181659


Edmond MB, Deschenes JL, Eckler M, Wenzel RP. Racial bias in using USMLE step 1 scores to grant internal medicine residency interviews. Acad Med 2001;76(12):1253–1256

Tekian A, Han Y, Hruska L, Kainik AJ. Do underrepresented minority medical students differ from non-minority students in problem-solving ability? Teach Learn Med 2001;13(02):86–91


