Read all three of our prestigious publications, each offering high-quality content to keep you informed with the latest developments in the field.

**Laryngoscope**

**Editor-in-Chief:** Michael G. Stewart, MD, MPH

The leading source for information in head and neck disorders.

[Laryngoscope.com](http://www.Laryngoscope.com)

---

**Laryngoscope Investigative Otolaryngology**

**Editor-in-Chief:** D. Bradley Welling, MD, PhD, FACS

Rapid dissemination of the science and practice of otolaryngology-head and neck surgery.

[InvestigativeOto.com](http://www.InvestigativeOto.com)

---

**ENTtoday**

**Editor-in-Chief:** Alexander Chiu, MD

Must-have timely information that Otolaryngologist-head and neck surgeons can use in daily practice.

[Enttoday.org](http://www.Enttoday.org)
Using Word Clouds to Re-envision Letters of Recommendation for Residency Applicants

Sinehan B. Bayrak, MD; Jennifer A. Villwock, MD; Mark R. Villwock, MS; Alexander G. Chiu, MD; Kevin J. Sykes, PhD, MPH

**Objectives/Hypothesis:** To develop a "word cloud"–based visual letter of recommendation (VLOR) and to evaluate its efficiency in discerning applicant quality compared to narrative letters of recommendation (NLORs).

**Study Design:** Cross-sectional cohort study.

**Methods:** NLORs for 48 otolaryngology residency applicants interviewed from the 2016 application cycle were identified and mined for descriptive terms to generate a word cloud, referred to as a VLOR. Eight individuals reviewed and rated a total of 187 blinded NLORs and 48 VLORs on a four-point scale (negative to exceptional). Median VLOR and NLOR scores and the time to review for each candidate were compared using the Wilcoxon signed rank test.

**Results:** It took significantly more time to review the NLORs in comparison to the VLORs (67 seconds, interquartile range [IQR]: 41–98 seconds vs. 17 seconds, IQR: 11–26 seconds, P < .001). There was no significant difference between median scores for VLORs and NLORs (P = .136). Review time and score correlated positively for VLORs and was statistically significant (ρ = 0.459, P = .001), indicating that more time spent reviewing equates to higher scores. The same relationship appeared with NLORs, but was not statistically significant (ρ = 0.276, P = .058).

**Conclusions:** VLORs are a novel and efficient additive tool for screening candidates for otolaryngology residency interview slots. Their scores do not significantly vary from NLOR scores and are significantly faster to evaluate.

**Key Words:** Letter of recommendation, residency, otolaryngology, application.

**Level of Evidence:** 2b

---

**INTRODUCTION**

Letters of recommendation (LORs) are an important element of the residency application process. On average, each applicant submits three to four letters from an assortment of faculty members with whom they have interacted during their medical school training. These LORs are traditionally narrative in nature. In general, they typically describe 1) the relationship between the author and the applicant, 2) the applicant’s academic accomplishments, and 3) the applicant’s personality. Previous publications have identified specific disadvantages of these narrative letters of recommendation (NLORs), including inability to predict future clinical performance, suboptimal interrater reliability, and intrinsic gender bias.

To address these shortcomings, a standardized letter of recommendation (SLOR) was developed in 1995 by the Council of Emergency Medicine Residency Directors and described in the emergency medicine literature in 1999. Advantages of this SLOR include less time required to evaluate applicants and better interrater reliability compared to the traditional NLOR. The SLOR was first piloted in a cohort of pediatric otolaryngology fellowship applicants in 2012, and then replicated in a cohort of otolaryngology residency program applicants that same year. Unfortunately, disadvantages of the SLOR became apparent shortly thereafter. These include a lack of variation in ratings, preventing differentiation amongst applicants, and the questionable utility of individual domains within the SLOR. Although the SLOR was arguably developed and implemented to replace the NLOR, the majority of institutions continue to submit both. The objective of this study was to evaluate the efficiency of our novel “word cloud”–based visual letter of recommendation (VLOR) that extracts key words and phrases from the NLOR and distills them into a word cloud. Also referred to as “tag clouds,” word clouds are a relatively novel technique to visualize textual data. The first known word cloud was created in 1976 by Stanley Milgram, a social psychologist, to illustrate mental attitudes toward the city of Paris. Decades later, a German edition of Mille Plateaux by the French philosopher Gilles Deleuze was published, with a word cloud summarizing the book’s content on the cover. The current popularity of word clouds is most attributable to their use by Flickr, a popular photo sharing platform, starting with its launch in 2004. Despite their extensive use in the fields of business, education, and computer science, word clouds have never been used to summarize letters of recommendation. We hypothesized that this method will preserve the descriptive value of the NLOR without sacrificing the efficiency of the SLOR.
MATERIALS AND METHODS

Study Design

Approval for this project was obtained from the University of Kansas Medical Center Institutional Review Board. Forty-eight otolaryngology residency applicants were interviewed at the study site for the 2016 application cycle, and all were included in this study. Otolaryngology residency applicants who did not interview at the study site for the 2016 application cycle were not included. All NLORs for all otolaryngology residency applicants interviewed from the 2016 application cycle were reviewed. NLORs for each applicant were mined for descriptors (e.g., adjectives, adverbs) by the first author. Using QSR Nvivo 11 (QSR International [Americas] Inc., Burlington, MA) qualitative data analysis software, all descriptors from all NLORs for each applicant were narrowed by synonym and queried into one word cloud-based VLOR for each applicant. Additional VLOR examples, along with excerpts from one of the multiple NLORs used to create them, can be seen in Figures 1 and 2.

NLORs were blinded of all identifying information to remove all references to names that would identify the applicant. After explicit verbal and written instructions were given by the primary author, eight individuals (two attending otolaryngologists, one rhinology fellow, one otolaryngology resident, two research personnel, and two administrative staff) employed by the Department of Otolaryngology–Head and Neck Surgery at the University of Kansas Medical Center reviewed and rated blinded narrative letters of recommendation. Each individual reviewed and rated all NLORs for six random applicants to provide an equal division of labor. Letters were ranked on a Likert-type scale as “negative,” “average,” “above average,” or “exceptional.” The length of time (measured in minutes and seconds) required to review each NLOR was also recorded via an imbedded REDCap function in which participants clicked to start and stop an internal timer. Once each participant had completed reviewing and rating their allocated NLORs, they then similarly reviewed and rated VLORs on the same scale. Forty-seven VLORs were reviewed by eight different reviewers, and one VLOR received seven reviews. Once both NLORs and VLORs were reviewed and rated, participants completed a validated instrument to measure usefulness and ease of use.14 All data were collected via REDCap.15

Results

Forty-eight otolaryngology residency applicants were included (47.9% male, 52.1% female). A total of 187 blinded NLORs were reviewed and rated. Afterward, a total of 48 corresponding VLORs were reviewed and rated in an identical fashion.

There was no significant difference in the distribution of median scores for VLORs and NLORs (P = .136) (Fig. 3). It took significantly more time to review the NLORs in comparison to the VLORs (67 seconds, interquartile range [IQR]: 41–98 seconds vs. 17 seconds, IQR: 11–26 seconds; P < .001). Review time and score correlated positively for VLORs and was statistically significant (ρ = 0.459, P = .001), indicating that more time spent evaluating equates to higher scores. This same positive correlation appeared with NLORs, but was not statistically significant (ρ = 0.276, P = .058).

Moreover, among males, VLOR and NLOR scores were statistically equivalent (P = .079). The same was also true for females (P = .747). There were also no significant differences in the median male or female VLOR scores (P = .124) or NLOR scores (P = .827).

After reviewing and rating NLORs and corresponding VLORs, seven participants completed a validated instrument to measure usefulness and ease of use. One participant did not complete the instrument, but the

Statistical Analysis

Data were analyzed using SPSS version 24 (IBM Corp., Armonk, NY). The median VLOR and NLOR scores and time of review for each candidate were compared using Wilcoxon signed rank tests. Subanalyses were performed within each gender to examine any differences between VLOR and NLOR also using the Wilcoxon signed rank test. The gender effect was further investigated using Mann-Whitney U tests to compare the median VLOR and NLOR score between men and women. Correlation between median review time and median score was evaluated using Spearman’s ρ. Significance was set at α = .05.
seven participants who did indicated that they believed the VLORs were moderately or extremely likely to enable completion of the review of letters more quickly and found the VLORs to be easy to use. Six out of seven (86%) reported that they found VLORs were moderately (three) or extremely (three) useful.

DISCUSSION

The ideal LOR should convey unique aspects of the individual applicant and provide reviewers with the ability to distinguish between the many high-performing applicants to identify those that best match the characteristics they seek in future residents. The objective of this study was to propose and evaluate a new additive tool for LORs utilized in applications for residency programs. To accomplish this objective, we developed a word cloud–based VLOR that preserves the descriptive context of the NLOR without sacrificing the efficiency of the SLOR. We evaluated the VLOR’s efficiency in discerning applicant quality and compared it to the standard NLORs. We found that NLOR and VLOR scores are not...
significantly different from one another, and that VLORs are significantly faster to evaluate.

The utility of LORs has been questioned in many fields including otolaryngology, especially as it may introduce gender bias and generally presents an overly positive view of all applicants. Prior studies have revealed that the vast majority of responses for the 10 domains of the SLOR were in the upper two deciles with a small standard deviation, indicative of a tightly grouped distribution. This lack of variability in ratings ultimately prevents differentiation amongst applicants. Similarly, Kimple et al. demonstrated that scores for professionalism, procedural skills, research, commitment to otolaryngology, and commitment to academic medicine were not correlated with successful matching. This further questions the utility of the individual domains within the SLOR.

Previous research has shown that the current standard of practice regarding letters of recommendation for otolaryngology residency applicants fails to fairly convey distinct characteristics (i.e., the SLOR) and is time consuming (i.e., the NLOR). Although existing literature has highlighted the shortcomings of SLORs, it has not yet proposed a solution to the problem. The majority of LOR authors continue to submit both NLORs and the SLORs, presumably in an effort to address the disadvantages of each; however, what was initially meant to shorten the applicant review process has now, ironically, lengthened it. This redundancy ultimately reflects on the perceived ineffectiveness of both.

The primary benefit of SLORs is efficiency, whereas the primary benefit of NLORs is contextual information. A study by Young showed that 61% of interviewers continue to prefer NLORs because they provide more information. As proven by our results, our VLORs preserve the qualitative content of NLORs without sacrificing the efficiency of SLORs by constructing a visual representation of narrative information. We also found that increased evaluative times positively correlated with score for both VLORs and NLORs, which may reflect more diverse, interesting, and engaging feedback within the VLOR or NLOR. Additionally, the majority of reviewers found VLORs to be both useful and easy to use.

Reading a NLOR that regurgitates the statistics of an applicant that are readily available in the curriculum vitae provides no value added and arguably distracts from the purpose. As an alternative, by illuminating key terms or characteristics directly from faculty and residents working with medical students on research projects and in clinical settings, letter authors can prospectively create a VLOR. This may provide programs reviewing applicants with a higher quality understanding of each individual applicant. Our study provides evidence that the VLOR is rated comparably to the NLOR and is significantly faster to evaluate than reading a standard NLOR.

Our findings are limited by the single-center nature of this study, the mining of NLORs by a single individual, and the evaluation of retrospectively created VLORs. This limited sample may explain the lack of a statistically significant difference between VLOR and NLOR scores across our total population; however, this was a pilot study, and without previous research to predict the variability in scores, we were not able to produce power calculations. Further research with a larger cohort of applicants and reviewers is needed to better address these limitations. Although it is unlikely that letter writers will significantly alter their style of writing, efforts to try to game the system by increasing the use of superlatives in their letters will be mitigated by the ability of the Nvivo algorithm to group synonyms together.

Although the applicant review processes may vary substantially across residency programs, we consider LORs to be distinguishing instruments for applications that contribute only partially to our assessment of an application. The VLOR or NLOR should help sort applicants who are otherwise comparable by all other measures considered valuable by program review committees. These letters only represent part of the picture. Currently, our department has implemented VLORs into our medical student evaluations by sending out an open-ended survey to all attendings and residents at the end of their rotation. For our fourth-year medical students interested in applying to otolaryngology, we have used these VLORs in lieu of a narrative letter from our chairman. In instances where specific anecdotes about the applicant are warranted (e.g., exceptional situations), small supporting paragraphs were included as well. Because VLORs only took an average of 17 seconds to evaluate, we do not think that the inclusion of these anecdotes significantly increases the reviewers’ burden. In the future, we plan to implement a similar system to evaluate resident performance.

CONCLUSION
LORs are an integral component of residency applications and often serve the purpose of describing the applicant’s academic accomplishments and the applicant’s personality. Although the former can be gleaned from objective data within an application, LORs are often the only glimpse residency programs have into the latter. Developing both an effective and efficient method of discerning applicants from one another may help to reduce the valuable time and effort expended by program directors and other applicant screeners.

In this study, we have created a novel LOR that combines the best elements of both SLORs (i.e., efficiency) and NLORs (i.e., meaningful content). We found that VLOR ratings did not significantly vary from NLOR ratings. Additionally, VLORs had the added benefit of significantly decreasing evaluative time. Future research on this topic will focus on whether or not prospective VLOR data generation on medical students and residents can create more effective snapshots of their achievements and strengths.

ACKNOWLEDGMENTS
The authors thank Vidur Bhalla, MD, Samantha Hall, BA, Andrew Holcomb, MD, Bryan Humphrey, AAS, RRT-NPS, Shannon Kraft, MD, and Dana Reinert, BA, MA, for reviewing and rating the visual and narrative
letters of recommendation, and Kavindu K. Ndeti, BA for assisting with data collection.

BIBLIOGRAPHY


