

The CORD Standardized Letter of Evaluation: Have We Achieved Perfection or Just a Better Understanding of Our Limitations?

DANIEL R. MARTIN, MD
ROBERT McNAMARA, MD

In the early 1990s, an emergency medicine (EM) program director remediated a resident for over a year, to no avail. The resident's contract was not renewed, and a recommendation made that the resident consider another specialty. When this decision was discussed with the department chair and clerkship director, who had written a very positive and "flowery" narrative letter of recommendation (NLOR), the chair said, "We knew this resident would struggle." The "fluffed up" letter was a disservice to colleagues and to the resident, who spent a difficult year in over her head. At the time, the general discussion among EM program directors indicated that accurate information transfer was a common limitation of the NLOR. Often, NLORs included no objective data (not even the EM clerkship grade) and provided no global comparison to other students. It was perceived that, often, one could not get to a "bottom line" view of the candidate despite a lengthy letter that was time consuming to prepare.

In an attempt to address the problems with NLORs, a council of EM residency directors (CORD) subcommittee developed a standardized letter of recommendation (SLOR) in 1995, which was initiated in 1997.¹ The SLOR offered more objective data than the NLOR, including an evaluation of EM clerkship performance and a prediction by the writer of how their program might rank the student. The original SLOR included the following 4 sections: (A) background information (clerkship performance); (B) qualifications for EM (personal characteristics relating to the choice of EM); (C) global assessment (comparisons to other students); and (D) written comments.

As discussed in a paper by Girzadas et al,² it quickly became clear that the SLOR was easier to prepare and read with the new format. Whether it made information transfer more reliable was a separate question, and that led to research of potential SLOR limitations. That research,

similar to what Girzadas et al² had found, suggested some problems.^{3,4} Potential biases resulting in grade inflation of the SLOR were uncovered. Areas of concern included gender bias, inexperience by letter writers, and the duration of time the letter writer knew the applicant.^{3,4} Another paper by a CORD task force demonstrated evidence of SLOR "grade inflation," as 40% of reviewed SLORs rated their applicants in the "top 10%," and over 95% of these SLORs rated applicants in the top third.⁵ Finally, when rank lists were compared with the global assessment question regarding estimated rank list position, overestimation on the SLOR occurred 66% of the time.⁶

The SLOR is central to 2 papers in this issue of the *Journal of Graduate Medical Education*. The study by Diab et al⁷ shows that the SLOR, with its measurable categories, allows research into the application process. Diab et al⁷ demonstrated a significant increase in the global assessment ranking of "outstanding" in letters where applicants did not waive their Family Educational Rights and Privacy Act (FERPA) rights, suggesting that if a faculty member is aware that an applicant may read their SLOR, the grade may be inflated.⁷ Thankfully, 93% of applicants waived their FERPA rights. The study is limited in that we do not know whether applicants who did not waive their rights were representative of the whole population of applicants, but it does suggest one should consider the possibility of bias if no waiver is present.

The paper by Hegarty et al⁸ describes the work of a CORD SLOR task force that was convened in 2011 to review the SLOR and determine whether improvements could be recommended. Although only 37% of the group surveyed had read the CORD guidelines in the previous year, these guidelines were very general and did not include specific recommendations for each question or even for each of the 4 sections. The consequence was great variability in how the question regarding "One Key Comment from ED Faculty Evaluations" was addressed in section A. The manner in which answers were interpreted was similarly variable in section B.

Perhaps the most interesting finding from the 2011 CORD SLOR task force work was the way question 12, regarding score inflation, was answered by those surveyed.

Daniel R. Martin, MD, is Professor and Vice Chair of Education, Department of Emergency Medicine, The Ohio State University; and **Robert McNamara, MD**, is Professor and Chair, Department of Emergency Medicine, Temple University.

Corresponding author: Daniel R. Martin, MD, Department of Emergency Medicine, The Ohio State University, 750 Prior Hall, 376 West 10th Avenue, Columbus, OH 43210, Daniel.Martin@osumc.edu

DOI: <http://dx.doi.org/10.4300/JGME-D-14-00172.1>

The results indicated that 36.2% reported they “rarely” inflated, 21.4% reported that they “sometimes” inflated, and 2.6% frequently inflated. My math ($36.2 + 21.4 + 2.6 = 60.2$) tells me that 60.2% of those surveyed admitted to inflating scores. This clearly agrees with earlier studies that showed SLOR inflation with a number of potential contributors.

It seems clear that the SLOR is neither completely objective nor highly accurate in terms of applicant ranking. Readers of the SLOR ideally would need to determine where the writer falls on the inflation scale, but that is not easily done. We are left to wonder what factors, such as FERPA or experience or time knowing the applicant or never really learning one’s fractions, are at play when the SLOR is written. The main question underlying the inflation issue is how best to honestly advocate for our students without doing them or one’s colleagues a disservice.

The recommendations of the CORD task force resulted in a new 2013–2014 edition of the SLOR with a mindful name change: the standardized letter of evaluation (SLOE). The SLOE applied task force recommendations in an attempt to standardize the writer response process. In section B, the “Qualifications” questions compare students to “peers” rather than use a number of adjectives that were hard to define. Other changes included eliminating confusing questions or parts of questions, such as the “Key Comments” requested in section A and the multiplication numbers in section C that described rank (2×, 4×, and 6×). However, the most significant change may be the theoretical shift that the new name represents. This is not a form to be used to blindly and subjectively recommend our students but rather to objectively evaluate their performance and attributes and accurately identify how they compare to a competitive pool of applicants.

Despite this aim, there are persistent limitations to the SLOE. For example, all program directors want to meet the students who score highest in every SLOE category, and we avoid those with the lowest SLOE scores. However, most applicants do not fit this profile. Thus, we must seek to understand the applicant’s true ranking from how much positive or negative concordance exists among SLOE sections, from the ever-shrinking narratives provided or by

prioritizing the items or sections that we feel are the best predictors of performance in our specific programs.

The greatest positive result from the CORD SLOR task force study is the effort by CORD to continue to evaluate and refine this important tool. Future efforts are needed to improve the accuracy and reliability of the SLOE, especially with a consistently increasing applicant pool. Further studies to better elucidate causes of rank inflation could include the use of more detailed instructions that are section-based or question-based. Additional outcome data for the applicants who are scored in the middle or lower thirds may help program directors have more confidence in considering these applicants.

We all recognize that despite the limitations discussed, the SLOE is not only the best tool for EM educators, it may be the best tool among any specialty. At the end of the day, even though we know the SLOE is not perfect, EM faculty spend less time creating and interpreting these imperfect recommendations than we did when the NLOR was the only option.

References

- 1 Keim SM, Rein JA, Chisholm C, Dyne PL, Hendey GW, Jouriles NJ, et al. A standardized letter of recommendation for residency application. *Acad Emerg Med*. 1999;6(11):1141–1146.
- 2 Girzadas DV Jr, Harwood RC, Dearie J, Garrett S. A comparison of standardized and narrative letters of recommendation. *Acad Emerg Med*. 1998;5(11):1101–1104.
- 3 Beskind DL, Hiller KM, Stolz U, Bradshaw H, Berkman M, Stoneking LR, 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(4):544–550. doi: 10.1016/j.jemermed.2013.08.025. Epub 2013 Oct 7.
- 4 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(9):988–991.
- 5 Love JN, Deiorio NM, Ronan-Bentle S, Howell JM, Doty CI, Lane DR, et al. Characterization of the Council of Emergency Medicine Residency Directors’ standardized letter of recommendation in 2011–2012. *Acad Emerg Med*. 2013;20(9):926–932.
- 6 Oyama LC, Kwon M, Fernandez JA, Fernández-Frackelton M, Campagne DD, Castillo EM, et al. Inaccuracy of the global assessment score in the emergency medicine standard letter of recommendation. *Acad Emerg Med*. 2010;17(suppl 2):38–41.
- 7 Diab J, Overton D, Riley S, Downes A, Gaeta T, Hern HG, et al. A multicenter study of the Family Education Rights and Privacy Act and the Standardized Letter of Recommendation: impact on emergency medicine residency applicant and faculty behaviors. *J Grad Med Educ*. 2014;6(2):292–295.
- 8 Hegarty CB, Lane DR, Love JN, Doty CI, Delorio NM, Ronan-Bentle S, et al. Council of Emergency Medicine Residency Directors standardized letter of recommendation writers’ questionnaire. *J Grad Med Educ*. 2014;6(2):301–306.