Measuring and Predicting Faculty Consensus Rankings of Standardized Letters of Evaluation

Morgan Sehdev, MD
Benjamin Schnapp, MD, MEd
Nicole M. Dubosh, MD
Al'ai Alvarez, MD
Alexis Pelletier-Bui, MD
Sharon Bord, MD
Caitlin Schrepel, MD
Yoon Soo Park, PhD
Eric Shappell, MD, MHPE

ABSTRACT

Background Standardized letters of evaluation (SLOE) are becoming more widely incorporated into the residency application process to make the letter of recommendation, an already critical component in a residency application packet, more objective. However, it is not currently known if the reviewers of these letters share consensus regarding the strength of an applicant determined by their SLOE.

Objective We measured the level of faculty agreement regarding applicant competitiveness as determined by SLOEs and the ability of 2 algorithms to predict faculty consensus rankings.

Methods Using data from the 2021-2022 Match cycle from the Council of Residency Directors in Emergency Medicine SLOE Database as a blueprint, authors created 50 fictional SLOEs representative of the national data. Seven faculty then rated these SLOEs in order of applicant competitiveness, defined as suggested rank position. Consensus was evaluated using cutoffs established a priori, and 2 prediction models, a point-based system and a linear regression model, were tested to determine their ability to predict consensus rankings.

Results There was strong faculty consensus regarding the interpretation of SLOEs. Within narrow windows of agreement, faculty demonstrated similar ranking patterns with 83% and 93% agreement for "close" and "loose" agreement, respectively. Predictive models yielded a strong correlation with the consensus ranking (point-based system r=0.97, linear regression r=0.97).

Conclusions Faculty displayed strong consensus regarding the competitiveness of applicants via SLOEs, adding further support to the use of SLOEs for selection and advising. Two models predicted consensus competitiveness rankings with a high degree of accuracy.

Introduction

Recently, 2 truths have become glaringly clear in the residency application process: individual programs in most specialties are sorting through more applications annually than ever before, and reviewers are tasked with reviewing more and possibly challenging to interpret evaluative metrics.[1-3] Over the past decade, the number of applications submitted to residency programs by each applicant roughly doubled.[1] The advent of virtual interviews since the 2020 Match cycle was associated with perpetuation of this pattern, as students submitted, on average, nearly 10 more applications than years prior.[4] While the use of various filters could allow residency program leadership to cull the number of applications their committees will review, this approach runs counter to the best practice of holistic applicant review and would still result in a significant amount of data application readers must sift through to make the decisions of whom to interview and rank.[5] As programs aim to approach application review holistically in the face of growing numbers of applicants, it is essential to assess the means by which applicants are efficiently and effectively evaluated through the data provided in each application.

Letters of recommendation are one such piece of data that have been identified as critical in a residency program's decision to interview and rank applicants during the Match process.[6] However, in addition to a boom in applications submitted annually, the past decade has brought the recognition that narrative letters of recommendation are of limited utility, prone to bias, and challenging to interpret.[7,8] These truths coincide in such a way that in 2021 the Coalition for Physician Accountability called for the implementation of standardized evaluations across specialties, and this sentiment is echoed in additional calls to actions such as that written by Tavarez et al in the *Journal of Graduate Medical Education* in

DOI: http://dx.doi.org/10.4300/JGME-D-22-00901.1

Editor's Note: The online supplementary data contains the Emergency Medicine Standardized Letter of Evaluation (2021-2022), a sample study standard letter of evaluation based with blueprinted ratings, faculty-developed narrative, and coded applicant and author identifiers, and further data from the study.

2023.[9,10] Before this year, several specialties, including emergency medicine, obstetrics and gynecology, orthopedic surgery, plastic surgery, otolaryngology, and dermatology have already adopted the use of a standardized letter of evaluation (SLOE) to improve the assessment of applicants' clinical performance.[11,12] The Emergency Medicine (EM) SLOE was the first to be implemented and has been cited as one of the most valuable application components for determining interview offers and location on the program's rank list.[13] Since the SLOE's inception, evaluators have sought to understand its utility as a tool. Primarily, research on the SLOE focuses on interrater reliability (ie, do writers evaluate individual students similarly and do SLOE rankings match other data points about an applicant?) and writer behaviors (ie, how do SLOE writers approach the SLOE and what bias might they bring, if any?).[11,12,14,15] The design, prior validity evidence, and weight of the SLOE in the application process make it a powerful tool to help programs determine applicants' likelihood to match, particularly as other ordinal performance measures (eg, United States Medical Licensing Examination Step 1, clerkship grades) are replaced with dichotomous pass/fail outcomes and narrative descriptions of performance.[16,17] While we may understand how SLOEs are written and the information they convey, limited data exist on whether those frequently reading and reviewing SLOEs agree on an applicant's competitiveness as determined by the SLOE alone. Schwartz's Theory of Basic Values would predict variability in this assessment, suggesting that different individuals (eg, advisors) and groups (eg, residency program leadership) often place priority on different values and hierarchies, depending on their culturally agreed upon structure, relationships, and attitudes.[18] For example, one program could particularly value high levels of professional behavior, while another prefers students with sophisticated patient assessments. Yet, should consensus exist, program directors and, more generally, entire specialties may feel empowered to turn to standardized evaluations as a valuable tool for efficient and effective stratification of applicant performance to aid in the recruitment process.

This study aims to measure the degree of faculty consensus on the competitiveness of applicants based on SLOEs. We also measure the ability of 2 models to predict faculty consensus rankings to determine if they can be used to accurately assess applicants' strength from the SLOE at scale.

KEY POINTS

What Is Known
More specialties are beginning to use a standardized letter of evaluation (SLOE) for the residency application process, with varying degrees of validity evidence for these tools depending on the specialty.

What Is New
This study of the Emergency Medicine SLOE used mock SLOEs to measure agreement of readers' positioning of the applicants on a fictional rank list as well as success of a predictive model.

Bottom Line
Good agreement was found, suggesting other specialties that might be newer to using standardized letters can use these findings when considering the validity of such an approach.

Methods

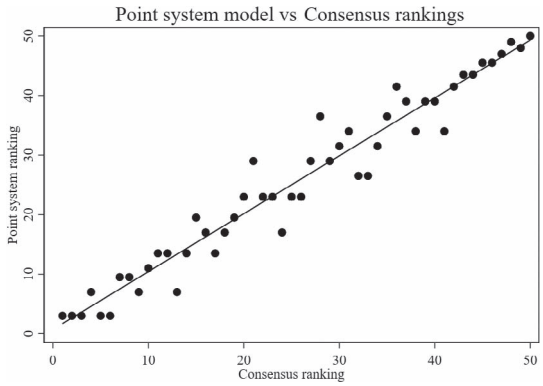
Setting and Faculty Participants

The study was conducted during the 2022 academic year. We recruited a convenience sample of faculty members with experience writing and reading SLOEs from diverse geographic regions across the United States to write representative narratives for our analysis (Association of American Medical Colleges regions represented: 3 Northeast, 2 Western, 1 Central, 1 Southern). At or before the time of the study, 7 of 7 (100%) faculty authors for this study (A.P.B., A.A., B.S., N.D., C.S. S.B., E.S.) regularly read SLOEs (mean ± standard deviation years of experience reading SLOEs=7.8 ± 2.4), and 5 of 7 (71%) regularly wrote SLOEs (mean ± standard deviation years of experience writing SLOEs=5.2 ± 4.1).

Data Acquisition, Blueprinting, Narratives

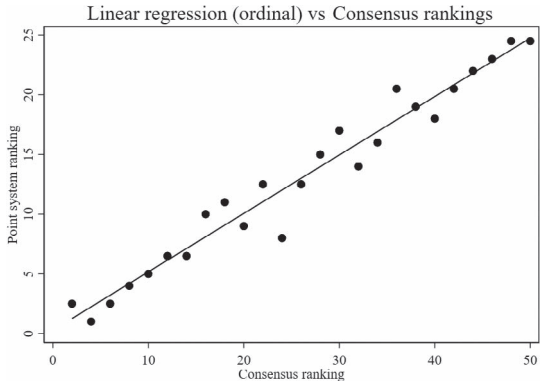
The first part of our intervention included the creation of 50 simulated, mock SLOEs. Data from the "Qualifications for EM" and "Global Assessment" ratings of all applicants in the 2021-2022 Match cycle were obtained from the Council of Residency Directors in Emergency Medicine (CORD) SLOE database. These ratings created a blueprint for 50 fictional SLOEs that would provide a representative national data sample (see online supplementary data FIGURE 1; SLOE field C.2.B.). The distribution of comparative assessment ratings (Qualifications for EM and Global Assessment) was used to populate the assessment ratings for fictional SLOEs (SLOE field C.1.). The final blueprint is detailed in online supplementary data TABLE 1. To avoid "stacking" of lower ratings that would make SLOEs more easily ranked in an ordinal fashion, the scores for B1-7 were assigned randomly for SLOEs otherwise containing the same scores (eg, the top 7 SLOEs in online supplementary data TABLE 1). Block randomization was used to assign letters to each of the 7 faculty authors who would subsequently write the mock SLOEs used in this study. This randomization was done to equalize the rating distributions for authors writing narratives,

A. Point System Model



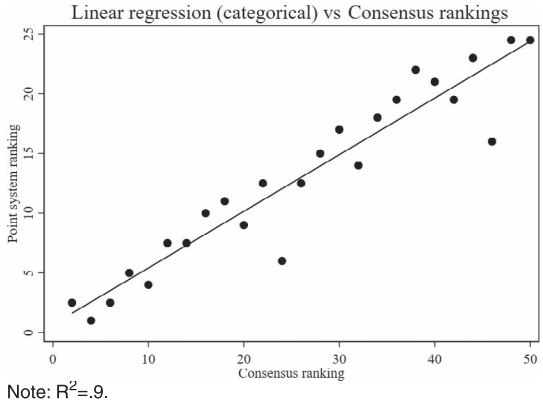
Note: R^2=.95.

B. Linear Regression Model (Ordinal)



Note: R^2=.94.

C. Linear Regression Model (Categorical)



Note: R^2=.9.

FIGURE 1A-C
Predicted vs Consensus Standardized Letter of Evaluation Rankings

similar to the distribution of student rotators throughout the academic year.

Authors then received their randomized set of SLOE blueprinted radio box data and were asked to write matching, genderless narratives using predefined instructions (BOX).

Outcomes Measured—Ranking and Consensus

We compiled narratives and ratings from the blueprint into mock SLOE documents using FileMaker Pro (Claris International Inc). A sample study SLOE is included in online supplementary data FIGURE 2. After the 50 SLOEs were compiled, faculty participants received instruction to rank these 50 SLOEs in order of preference for matching at their program, assuming these documents were the only information available. Consensus rankings were established by ordering SLOEs by average faculty rating. In cases of duplicate average rankings, the group decided a priori that the letter with the lower standard deviation in rank ratings would be assigned the higher consensus rank. Definitions for levels of agreement are detailed in TABLE 1.

Analysis of Outcomes—Predictive Models

We utilize 2 methods to predict consensus rankings: (1) a point system developed by one author (E.S.) before national SLOE data were obtained and (2) a linear regression model. For the point system, we assigned each rating in the Qualifications for EM and Global Assessment sections of the SLOE a point value (online supplementary data TABLE 2) based on experience reading and writing SLOEs. We assigned predicted rankings by point totals and compared them to consensus rankings. For the regression model, half the data set was used to train the model, and the other half was used for validation. SLOEs with odd-numbered consensus rankings were used as a training data set to develop the model. We then applied the model to a validation data set of SLOEs with even consensus rankings. The regression model was run with SLOE rankings both as ordinal and categorical variables. For both prediction algorithms, redundant rankings were managed by averaging the rank for that range (eg, if the top 5 SLOEs had the identical scores, all 5 of those SLOEs would be assigned a ranking of 3). Also, neither model utilized narrative comments to assign scores or subsequent ranking.

This study was deemed exempt by the Institutional Review Board at Mass General Brigham.

Results

The faculty consensus ranking process and the accuracy of predictive models are outlined in TABLE 2. Graphical comparisons of consensus versus predicted rankings are displayed in FIGURE 1A-C.

Faculty Consensus

Consensus among faculty raters was strong (in this case, we adopt "strong" terminology used in the

TABLE 1
Agreement Definitions

Ranking	Consensus: Faculty Ratings	Prediction: Point System	Prediction: Regression
n	350 rankings (7 raters x 50 SLOEs)	50 rankings	25 training rankings/ 25 validation rankings
Exact	Percent of rankings where faculty assign same rank as consensus rank	Percent of rankings with same assigned rank as consensus rank	Percent of rankings in validation set with same assigned rank as consensus rank
Tight	Percent of rankings where faculty rank is within 2 positions (± 4%) of consensus rank	Percent of rankings with assigned rank within 2 positions (± 4%) of consensus rank	Percent of rankings with assigned rank within 1 position (± 4%) of consensus rank
Close	Percent of rankings where faculty rank is within 4 positions (± 8%) of consensus rank	Percent of rankings with assigned rank within 4 positions (± 8%) of consensus rank	Percent of rankings with assigned rank within 2 positions (± 8%) of consensus rank
Loose	Percent of rankings where faculty rank is within 6 positions (± 12%) of consensus rank	Percent of rankings with assigned rank within 6 positions (± 12%) of consensus rank	Percent of rankings with assigned rank within 3 positions (± 12%) of consensus rank

evaluation of correlation coefficient agreement).[19] While there was only 21% "exact" agreement on the ranking of the SLOEs, "tight," "close," and "loose" consensus agreement remained consistently above 66%, with a maximum agreement of 93%.

Predictive Model Performance

Utilizing the same agreement criteria outlined in TABLE 1, faculty consensus was compared to the rankings predicted by the point-based scoring system (online supplementary data TABLE 2). Consensus and point system rankings demonstrated a tight correlation (r=0.97). The predictive accuracy of this system was only marginally lower than faculty consensus agreement (ie, "tight" agreement 62% with point model vs 67% with consensus, "close" agreement 82% with point model vs 83% with consensus).

A linear regression model was also used to predict consensus rankings. The first linear regression model interpreted radio-box entries ordinally. This model closely correlated with the consensus data (r=0.97). Compared to consensus data, this predictive model also had similar "tight" (64%), "close" (92%), and "loose" (96%) ranking agreements. The ordinal model yielded a mean absolute difference of 1 (4%) between the predicted and measured ranks, with a maximum difference of 4 (16%).

The second linear regression model interpreted radio box entries categorically. This second model is

Box Instructions for Letter Writers

Writing

You will receive 7 sets of SLOE rankings for the sections "Qualifications for EM" and "Global Assessment." These rankings have been created from a blueprint of the actual distribution of rankings for SLOEs created between January 1, 2021, and December 21, 2021. For each set of rankings, you are asked to write a SLOE narrative for a fictitious student that is consistent with these rankings. You have creative license to include as much detail or examples that you see fit in these narratives, however:

- Please keep your approach consistent across all letters. Your narratives should read as if the same person authored all of your letters at the same institution. Your author ID code (not your name) will be included in the signature line at the bottom of the letters in the rating process, so raters will be able to see which sets of letters came from the same author.
- Please only include comments about clerkship performance in the narrative (ie, no comments about outside research or other roles in the medical school or community).
- Please do not include language that specifies the fictitious student's gender, race, or sexual orientation.
- To refer to the student by name, use "Student X".
- Use they/them pronouns to refer to the student.

Abbreviations: SLOE, standardized letter of evaluation; EM, emergency medicine.
Note: Before writing their assigned SLOEs, Letter Writers were sent detailed instructions, as outlined here, to make the simulated SLOEs as unbiased as possible while still reading as true to form as possible.

TABLE 2
Ranking Agreement

Ranking	Consensus: Faculty Ratings	Prediction: Point System	Prediction: Regression (Ordinal)	Prediction: Regression (Categorical)
Exact	21%	12%	20%	0%
Tight	67%	62%	64%	52%
Close	83%	82%	92%	88%
Loose	93%	90%	96%	92%
Correlation with consensus ratings	N/A	.97	.97	.98

Note: Using the definitions outlined in TABLE 1, we determined ranking agreement as reported here for faculty rankings, the predictive point system, and predictive linear regressions.

also closely correlated to the consensus data (r=0.98). Compared to consensus data, this predictive model yielded similar yet slightly lower "tight" (55%), "close" (88%), and "loose" (92%) ranking agreements. The categorical model yielded a mean absolute difference of 1.7 (7%) between the predicted and fitted data with a maximum deviation of 7 (28%).

Discussion

We found strong consensus among faculty raters regarding the interpretation of SLOEs. Faculty demonstrated similar ranking patterns within narrow windows of agreement, with 83% and 93% agreement for "close" and "loose" agreement, respectively. These findings support that faculty generally share a unified perspective on applicant competitiveness based on SLOEs.

The geographic and institutional differences among our study's raters strengthen the value of the overall high level of consensus observed as it reflects a degree of universal agreement of "competitiveness" as determined by SLOEs alone. These differences also help explain why the "exact" and "tight" constraints yielded "poor" to "fair" agreement: it is expected that a diverse pool of raters reviewing SLOEs will have slightly different priorities depending on the program they represent.[18,20-22] Regardless, our initial data show that faculty generally possess similar perspectives on the interpretation of SLOEs, which provides further support for using SLOEs, as demonstrated by the iterative and thoughtful design of the SLOE.[23,24] Our fictional SLOEs likely helped minimize evaluative differences, bias, and possible rank inflation known to impact SLOE writing and presumably subsequent ranking.[11,25-27] Characterizing consensus using actual SLOEs will be valuable if studied while considering the different contexts in which they are written (writer/student differences, narrative components, etc).

A limitation of this study includes the relatively small group of reviewers reading a fraction of the SLOEs that a residency program would receive in a typical application cycle.[28] It is unknown if the results of this study would be reproducible with a larger group, including reviewers with less experience or from specialties less familiar with SLOEs. While diverse SLOE writers and readers were used to author the fictional SLOEs used in this study for fidelity, it is possible that actual SLOEs may be written, constructed, or interpreted somewhat differently given different institutional, program, or geographic differences.

Another limitation is the bidirectional nature of competitiveness in the Match. While applicant competitiveness remains pivotal in the Match, residency program competitiveness as perceived by applicants likely creates an interplay between application availability, ranking, and the degree to which an applicant is viewed as "competitive."[29,30] In emergency medicine, there is no universally accepted program competitiveness ranking, as program types, qualities, and missions vary widely.[31-33] Future research should explore whether this poorly understood characteristic of perceived program strength alters the strong consensus we observed among representatives from 7 unique programs.

A final limitation of the study is the deidentified nature of the SLOEs used for the review. Previous studies have demonstrated that SLOEs are preferentially reviewed and ranked depending on various contextual identifiers, including home vs away status, the program from which the SLOE was written, and the letter writer's experience.[34-36] These subjective identifiers may skew the interpretation of any SLOE, making a prediction model based on objective radio box data less accurate. The narrative comments are not considered in the predictive point system or the regression models used in this study. Prior work on this section of the SLOE shows gender and racial bias in SLOEs.[14,37] While the lack of these identifiers in our study limits the contribution to the issue of bias, our finding that faculty have

strong consensus in the competitiveness rankings of deidentified SLOEs lays the groundwork for further research in this domain, serving as a potential comparison group for SLOE rankings that include these identifiers.

In future work addressing SLOE utility and implementation, it will be necessary to understand if these findings can be replicated using increasingly complex and higher-fidelity contextual factors, including real SLOEs and information typically available to reviewers such as medical school, rotation site, gender, and race/ethnicity information. Given the strong correlation between predictions of models without narratives and faculty rankings of SLOEs with narratives, further work is warranted to understand when and how narrative comments influence perceptions of the applicant through a SLOE independent of context established by radio buttons. Data suggests bias in the stories presented in this narrative commentary[2,38]; however, we still need to understand whether biased narratives change the perception of radio box SLOE data. As SLOEs vary across specialties and change over time, future studies should examine how specific features of SLOEs impact the degree of faculty consensus. For instance, it will be necessary to understand if our results are reproducible with new updates to the EM SLOE format[39] and SLOEs from other specialties. It is also possible that specialties with less experience using SLOEs would have lower consensus ratings, given different formats and familiarity with the process; this hypothesis should be tested in future studies. Comparative performance measures based on varied SLOE designs could be used to inform future SLOE revisions to optimize desired performance outputs. Finally, characterizing how "competitiveness" differs between residency programs within and across specialties would be valuable to understand generalizability and the changes based on institutional cultures, values, or preferences.

As faculty rankings yielded good agreement on SLOE strength, this work also created a means of comparison for our "competitiveness" prediction algorithms. The point-based and linear regression models strongly correlated with consensus competitiveness rankings. Intuitively, the ordinal regression model seems most appropriate given that the SLOE radio boxes require a tiered evaluation. This appropriateness was further supported by the more accurate predictions of this model, demonstrating both lower average and lower maximum differences between predicted and measured competitiveness rankings. Therefore, we believe the ordinal regression model is the most appropriate for practical application. Given the predictive similarity between the point-based scoring system and the ordinal regression, either could serve as a viable future prediction model for SLOE strength. Moving forward, it is possible that predictive models could gain additional validity evidence and later be applied to the application review process—either screening or sorting applicants. As we continue to investigate the role of artificial intelligence in the application review process, these and more advanced predictive models may be helpful in the search for a streamlined yet holistic review process for the annually growing influx of submitted residency applications.[39,40]

While it may not be feasible to limit the number of applications received by any given program each application cycle, it is possible to control, scrutinize, and improve the data obtained in each application. This work helps demonstrate that, despite diverse writing styles and approaches, standardized letters of evaluation convey information in such a way that reviewers exhibit strong levels of consensus when ranking applicants using SLOEs alone, regardless of unique programmatic goals.

Conclusions

Faculty displayed strong consensus regarding the competitiveness of applicants based on SLOEs, adding validity evidence to using SLOEs for residency selection. Two models predicted consensus competitiveness rankings with a high degree of accuracy.

References

1. Carmody JB, Rosman IS, Carlson JC. Application fever: reviewing the causes, costs, and cures for residency application inflation. *Cureus*. 2021;13(3):e13804. doi:10.7759/cureus.13804
2. Ginsburg S, Regehr G, Lingard L, Eva KW. Reading between the lines: faculty interpretations of narrative evaluation comments. *Med Educ*. 2015;49(3):296-306. doi:10.1111/medu.12637
3. Association of American Medical Colleges. ERAS statistics. Accessed October 19, 2023. https://www.aamc.org/data-reports/data/eras-statistics-data
4. Meyer AM, Hart AA, Keith JN. COVID-19 increased residency applications and how virtual interviews impacted applicants. *Cureus*. 14(6):e26096. doi:10.7759/cureus.26096
5. Garber AM, Kwan B, Williams CM, et al. Use of filters for residency application review: results from the Internal Medicine In-Training Examination program director survey. *J Grad Med Educ*. 2019;11(6):704-707. doi:10.4300/JGME-D-19-00345.1
6. Saudek K, Saudek D, Treat R, Bartz P, Weigert R, Weisgerber M. Dear program director: deciphering letters of recommendation. *J Grad Med Educ*. 2018;10(3):261-266. doi:10.4300/JGME-D-17-00712.1

7. Saudek K, Treat R, Goldblatt M, Saudek D, Toth H, Weisgerber M. Pediatric, surgery, and internal medicine program director interpretations of letters of recommendation. *Acad Med*. 2019;94(suppl 11 Association of American Medical Colleges Learn Serve Lead: Proceedings of the 58th Annual Research in Medical Education Sessions):64-68. doi:10.1097/ACM.0000000000002919
8. Zhang N, Blissett S, Anderson D, O'Sullivan P, Qasim A. Race and gender bias in internal medicine program director letters of recommendation. *J Grad Med Educ*. 2021;13(3):335-344. doi:10.4300/JGME-D-20-00929.1
9. Coalition for Physician Accountability. The Coalition for Physician Accountability's Undergraduate Medical Education-Graduate Medical Education Review Committee (UGRC): Recommendations for Comprehensive Improvement of the UME-GME Transition. Published 2021. Accessed December 6, 2023. https://physicianaccountability.org/wp-content/uploads/2021/08/UGRC-Coalition-Report-FINAL.pdf
10. Tavarez MM, Baghdassarian A, Bailey J, et al. A call to action for standardizing letters of recommendation. *J Grad Med Educ*. 2022;14(6):642-646. doi:10.4300/JGME-D-22-00131.1
11. Kukulski P, Ahn J. Validity evidence for the Emergency Medicine Standardized Letter of Evaluation. *J Grad Med Educ*. 2021;13(4):490-499. doi:10.4300/JGME-D-20-01110.1
12. Jackson JS, Bond M, Love JN, Hegarty C. Emergency Medicine Standardized Letter of Evaluation (SLOE): findings from the new electronic SLOE format. *J Grad Med Educ*. 2019;11(2):182-186. doi:10.4300/JGME-D-18-00344.1
13. The National Resident Matching Program. Results from the 2020 NRMP Program Director Survey. Updated November 28, 2022. Accessed on June 21, 2023. https://public.tableau.com/app/profile/national.resident.matching.program/viz/PDSurvey2020-Final/Desktoptablet
14. Alvarez A, Mannix A, Davenport D, et al. Racial bias in medical student standardized letters of evaluation (SLOE). *West J Emerg Med*. 2022;23(suppl 4):18.
15. Love JN, Doty CI, Smith JL, et al. The Emergency Medicine Group standardized letter of evaluation as a workplace-based assessment: the validity is in the detail. *West J Emerg Med*. 2020;21(3):600-609. doi:10.5811/westjem.2020.3.45077
16. McDonald JA, Lai CJ, Lin MYC, O'Sullivan PS, Hauer KE. "There is a lot of change afoot": a qualitative study of faculty adaptation to elimination of tiered grades with increased emphasis on feedback in core clerkships. *Acad Med*. 2021;96(2):263-270. doi:10.1097/ACM.0000000000003730
17. Humphrey HJ, Woodruff JN. The pass/fail decision for USMLE Step 1—next steps. *JAMA*. 2020;323(20):2022-2023. doi:10.1001/jama.2020.3938
18. Schwartz SH. An overview of the Schwartz Theory of Basic Values. *Online Read Psychol Cult*. 2012;2(1). doi:10.9707/2307-0919.1116
19. McHugh ML. Interrater reliability: the kappa statistic. *Biochem Med (Zagreb)*. 2012;22(3):276-282.
20. Hartman ND, Lefebvre CW, Manthey DE. A narrative review of the evidence supporting factors used by residency program directors to select applicants for interviews. *J Grad Med Educ*. 2019;11(3):268-273. doi:10.4300/JGME-D-18-00979.3
21. Katzung KG, Ankel F, Clark M, et al. What do program directors look for in an applicant? *J Emerg Med*. 2019;56(5):e95-e101. doi:10.1016/j.jemermed.2019.01.010
22. Bhat R, Takenaka K, Levine B, et al. Predictors of a top performer during emergency medicine residency. *J Emerg Med*. 2015;49(4):505-512. doi:10.1016/j.jemermed.2015.05.035
23. Love JN, Smith J, Weizberg M, et al. Council of Emergency Medicine Residency Directors' standardized letter of recommendation: the program director's perspective. *Acad Emerg Med*. 2014;21(6):680-687. doi:10.1111/acem.12384
24. Hegarty CB, Lane DR, Love JN, et al. Council of Emergency Medicine Residency Directors standardized letter of recommendation writers' questionnaire. *J Grad Med Educ*. 2014;6(2):301-306. doi:10.4300/JGME-D-13-00299
25. Hansroth JA, Davis KH, Quedado KD, et al. Lower-third SLOE rankings impede, but do not prevent, a match in emergency medicine residency training. *J Med Educ Curric Dev*. 2020;7:2382120520980487. doi:10.1177/2382120520980487
26. Kang HP, Robertson DM, Levine WN, Lieberman JR. Evaluating the standardized letter of recommendation form in applicants to orthopaedic surgery residency. *J Am Acad Orthop Surg*. 2020;28(19):814-822. doi:10.5435/JAAOS-D-19-00423
27. Oyama LC, Kwon M, Fernandez JA, 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. doi:10.1111/j.1553-2712.2010.00882.x
28. Weissbart SJ, Kim SJ, Feinn RS, Stock JA. relationship between the number of residency applications and the yearly Match rate: time to start thinking about an application limit? *J Grad Med Educ*. 2015;7(1):81-85. doi:10.4300/JGME-D-14-00270.1
29. Zigrossi D, Ralls G, Martel M, et al. Ranking programs: medical student strategies. *J Emerg Med*. 2019;57(4):e141-e145. doi:10.1016/j.jemermed.2019.04.027
30. Love JN, Howell JM, Hegarty CB, et al. Factors that influence medical student selection of an emergency medicine residency program: implications for training programs. *Acad Emerg Med*. 2012;19(4):455-460. doi:10.1111/j.1553-2712.2012.01323.x

31. Emergency Medicine Residents' Association. Joint Letter to Doximity. Accessed October 11, 2022. http://www.emra.org/be-involved/be-an-advocate/working-for-you/doximity-letter-2020/
32. Yarris LM, DeIorio NM, Lowe RA. Factors applicants value when selecting an emergency medicine residency. *West J Emerg Med*. 2009;10(3):159-162.
33. Farcy D, Muelleman R, Rogers J, et al. Joint Statement to Doximity. Published online June 2019. Accessed October 11, 2022. https://www.saem.org/docs/default-source/saem-documents/position-statements/joint-statement-to-doximity-opposing-the-continued-use-of-the-doximity-residency-navigator-platform-for-emergency-medicine.pdf?sfvrsn=84d7296e_3
34. Feldman MJ, Ortiz AV, Roth SG, et al. An examination of standardized letters of recommendation rating scales among neurosurgical residency candidates during the 2020-2021 application cycle. *Neurosurgery*. 2021;89(6):1005-1011. doi:10.1093/neuros/nyab346
35. Beskind DL, Hiller KM, Stolz U, et al. Does the experience of the writer affect the evaluative components on the standardized letter of recommendation in emergency medicine? *J Emerg Med*. 2014;46(4):544-550. doi:10.1016/j.jemermed.2013.08.025
36. Wilson D, Laoteppitaks C, Chandra S. A comparison of standardized letters of evaluation for emergency medicine residency applicants. *West J Emerg Med*. 2021;22(1):20-25. doi:10.5811/westjem.2020.12.49086
37. Mannix A, Monteiro S, Miller D, et al. Gender differences in emergency medicine standardized letters of evaluation. *AEM Educ Train*. 2022;6(2):e10740. doi:10.1002/aet2.10740
38. Powers A, Gerull KM, Rothman R, Klein SA, Wright RW, Dy CJ. Race- and gender-based differences in descriptions of applicants in the letters of recommendation for orthopaedic surgery residency. *JBJS Open Access*. 2020;5(3):e20.00023. doi:10.2106/JBJS.OA.20.00023
39. Burk-Rafel J, Reinstein I, Feng J, et al. Development and validation of a machine learning-based decision support tool for residency applicant screening and review. *Acad Med*. 2021;96(suppl 11):54. doi:10.1097/ACM.0000000000004317
40. Kibble J, Plochocki J. Comparing machine learning models and human raters when ranking medical student performance evaluations. *J Grad Med Educ*. 2023;15(4):488-493. doi:10.4300/JGME-D-22-00678.1

Morgan Sehdev, MD, is a PGY-2 Resident Physician, Harvard-Affiliated Emergency Medicine Residency, Brigham and Women's Hospital, Massachusetts General Hospital, Boston, Massachusetts, USA; **Benjamin Schnapp, MD, MEd,** is Associate Professor (CHS), Department of Emergency Medicine, University of Wisconsin School of Medicine and Public Health, Madison, Wisconsin, USA; **Nicole M. Dubosh, MD,** is Associate Professor, Department of Emergency Medicine, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, Massachusetts, USA; **Al'ai Alvarez, MD,** is Clinical Associate Professor and Director of Well-Being, Department of Emergency Medicine, Stanford University School of Medicine, Stanford, California, USA; **Alexis Pelletier-Bui, MD,** is Assistant Professor, Department of Emergency Medicine, Cooper Medical School of Rowan University, Camden, New Jersey, USA; **Sharon Bord, MD,** is Assistant Professor, Department of Emergency Medicine, The Johns Hopkins University School of Medicine, Baltimore, Maryland, USA; **Caitlin Schrepel, MD,** is Assistant Professor, Department of Emergency Medicine, University of Washington School of Medicine, Seattle, Washington, USA; **Yoon Soo Park, PhD,** is Associate Professor, Department of Emergency Medicine, Massachusetts General Hospital, Harvard Medical School, Boston, Massachusetts, USA; and **Eric Shappell, MD, MHPE,** is Assistant Professor, Department of Emergency Medicine, Massachusetts General Hospital, Harvard Medical School, Boston, Massachusetts, USA.

Funding: The authors report no external funding source for this study.

Conflict of interest: The authors declare they have no competing interests.

The authors would like to thank the Council of Residency Directors in Emergency Medicine for providing data that informed our blueprinting that was pivotal in the design of our fictional SLOEs.

Corresponding author: Morgan Sehdev, MD, Brigham and Women's Hospital, Massachusetts General Hospital, Boston, Massachusetts, USA, msehdev@mgh.harvard.edu, X @msehdev23

Received November 16, 2022; revisions received July 7, 2023, and October 23, 2023; accepted November 8, 2023.