The Feasibility of Blinding Residency Programs to USMLE Step 1 Scores During GME Application, Interview, and Match Processes

Kathy W. Smith, MD
Richard Amini, MD
Madhulika Banerjee, BS
Conrad J. Clemens, MD, MPH

ABSTRACT

Background With the recent announcement that Step 1 score reporting will soon change to pass/fail, residency programs will need to reconsider their recruitment processes.

Objective We (1) evaluated the feasibility of blinding residency programs to applicants' Step 1 scores and their number of attempts throughout the recruitment process; (2) described the selection process that resulted from the blinding; and (3) reviewed if a program's initial rank list, created before scores were known, would be changed before submission for the Match.

Methods During the 2018–2019 and 2019–2020 recruitment seasons, all programs at a single sponsoring institution were invited to develop selection criteria in the absence of Step 1 data, and to remain blinded to this data throughout recruitment. Participating programs were surveyed to determine factors affecting feasibility and metrics used for residency selection. Once unblinded to Step 1 scores, programs had the option to change their initial rank lists.

Results Of 24 residency programs, 4 participated (17%) in the first year: emergency medicine, neurology, pediatrics, and psychiatry. The second year had the same participants, with the addition of family and community medicine and radiation oncology (n = 6, 25%). Each program was able to determine mission-specific qualities in the absence of Step 1 data. In both years, one program made changes to the final rank list.

Conclusions It was feasible for programs to establish metrics for residency recruitment in the absence of Step 1 data, and most programs made no changes to final rank lists after Step 1 scores were known.

Introduction

The United States Medical Licensing Examination Step 1 and the National Board of Osteopathic Medical Examiners COMLEX-USA Level 1 score reporting will soon change to pass/fail.[1,2] The need for residency programs to develop recruitment processes in the absence of Step 1 scores will become a priority, and programs may need to develop alternative selection criteria.

The controversial use of Step 1 scores in resident selection has been well described, and stakeholders have recognized the unintended consequences of its use in resident selection.[3–5] Recently, more emphasis has been placed on improving holistic review in residency recruitment and decreasing the impact of Step 1 scores; however, these attempts have largely involved concealing Step 1 scores from individuals involved in recruitment after applicants have been selected to interview.[6–8] To our knowledge, there have not been previous attempts to assess the feasibility of removing Step 1 data from the entire recruitment process.

Our study objectives were (1) to assess the feasibility of blinding residency programs to an applicant's USMLE Step 1/COMLEX-USA Level 1 examination score and the number of attempts throughout the entire selection process; (2) to determine other metrics to use in resident selection; and (3) to determine if a program's initial rank list, created before Step 1/Level 1 scores are known, would be changed before final submission for participation in the Match.

Methods

This was a single center, prospective, cohort study conducted at an academic medical center over 2 successive recruitment seasons (2018–2020). The project idea was presented to the Graduate Medical Education Committee and in conversations with program directors (PDs) and associate program directors (APDs). Programs who volunteered to participate were asked to blind themselves to Step 1/COMLEX-USA Level 1 scores, and the number of

DOI: http://dx.doi.org/10.4300/JGME-D-20-00653.1

Editor's Note: The online version of this article contains the survey and questionnaires used in the study.

attempts applicants made, during every step of the selection process and when creating an initial rank list. Programs were also asked to describe the selection process and metrics used in the absence of Step 1 data.

Programs were not required to blind themselves to Step 2 data. For this study, "Step 1" includes both Step 1 and COMLEX-USA Level 1.

Data were collected from September to March of each year. Programs tasked coordinators to develop a way to ensure Step 1 data were removed from application materials. PDs were asked to ensure faculty and residents involved in recruitment were also blinded. On the day the Electronic Residency Application Service (ERAS) opened, and weekly thereafter, programs were queried via an online survey instrument created by the study authors (provided as online supplementary data) to determine whether and how Step 1 scores were unblinded during the preceding week and to identify issues that may have arisen. If programs missed a weekly survey, unblinding episodes of Step 1 data were asked to be reported the following week.

Using semistructured questionnaires created by the study authors (provided as online supplementary data), 2 focus groups were held at the midpoint and end of each recruitment season. At the end of recruitment, programs reported the number of applicants that applied and were interviewed and submitted 2 deidentified rank lists. The first was created while still blinded to Step 1 data, and the second was the final rank list created after unblinding occurred and submitted to the National Resident Matching Program.

The study protocol was reviewed and approved by the University of Arizona College of Medicine-Tucson Institutional Review Board, and PDs and APDs consented prior to participation.

Objectives
We evaluated the feasibility of blinding residency programs to an applicant's Step 1 data throughout the entire recruitment process and described the selection process that resulted from the blinding.

Findings
It was feasible for programs to establish metrics for residency recruitment in the absence of Step 1 data.

Limitations
The choice of alternative variables in the absence of Step 1 data were left to the discretion of the programs and may not necessarily be better predictors of resident performance.

Bottom Line
Residency programs will soon need to decide how to select and rank applicants in the absence of Step 1 data, and this study provides foundational information which can be used by programs to meet this challenge.

Results

Of 24 residency programs, emergency medicine, neurology, pediatrics, and psychiatry agreed to participate in the first year (n = 4). In addition to these, family and community medicine and radiation oncology participated in year 2 (n = 6).

Weekly Survey Data

Year 1: A total of 3779 applications were submitted to the 4 programs (TABLE 1). Of these, 199 (5.3%) were unblinded, with the number of unblinded applicants ranging from 1% to 11.4%. Weekly survey completion rates ranged from 67% to 100%.

Year 2: A total of 5149 applications were submitted to the 6 programs. Of these, 159 (3.1%) were unblinded, with the number of unblinded applicants ranging from 0% to 7.4%. Weekly survey completion rates were 100%.

The most common reasons for unblinded scores are listed in TABLE 2.

Focus Groups

Programs Participating for the First Time (n = 6; 4 in year 1, 2 in year 2): When asked about challenges during implementation, 4 programs estimated 5% to 20% more time was required to review applications, and 2 programs reported no additional time was needed. Reasons for extra time included (1) removing Step 1 data from applicant materials such as redacting Step 1 data before printing for distribution to individuals involved in recruitment or copying non-Step 1 data to the "Notes" section in ERAS to avoid opening the transcript section; (2) reviewing more applications; and (3) ensuring a thorough review of each application. Programs reported that Step 1 scores were revealed in many ways throughout the recruitment process, making it challenging to stay blinded to this data (TABLE 2).

Two programs reported faculty had initially expressed concern about participation in the study but were reassured when it was understood that the scores could be revealed at the end of the season. One program noted one faculty was concerned the process would take more time.

One program informed applicants of their participation in the study when they came to interview, while the remaining 5 programs told applicants about the study only if they attempted to divulge their score, or if they specifically inquired. When applicants were informed of the study, they responded favorably overall. One applicant had a negative response to not having their Step 1 score be known.

TABLE 1
Weekly Survey Data: Step 1 Unblinding Episodes per Program

Program	Year 1 (2018–2019)			Year 2 (2019–2020)		
	Total Applicants	No. (%) Unblinded	Weekly Response Rate, No. (%), N = 18	Total Applicants	No (%) Unblinded	Weekly Response Rate, No. (%), N = 18
Emergency medicine	885	9 (1.0)	12 (67)	858	15 (1.7)	18 (100)
Neurology	653	51 (7.8)	12 (67)	678	9 (1.3)	18 (100)
Pediatrics	1082	123 (11.4)	18 (100)	1262	93 (7.4)	18 (100)
Psychiatry	1159	16 (1.4)	13 (72)	1082	14 (1.3)	18 (100)
Family and community medicine	DNP	N/A	N/A	1187	28 (2.4)	18 (100)
Radiation oncology	DNP	N/A	N/A	82	0 (0.0)	18 (100)
All programs	3779	199 (5.3)	. . .	5149	159 (3.1)	. . .

Abbreviations: DNP, did not participate; N/A, not available.

When asked about previous recruitment processes, 5 programs had a preexisting algorithm that utilized Step 1 scores; of these, 3 kept their original algorithm and removed the Step 1 data as variables, and 2 created novel processes. One program with no previous algorithm created a new one. All programs stated their algorithms accurately captured the type of applicant they were seeking, and all planned to continue using their algorithms. The most common variables used in the absence of Step 1 are listed in the BOX.

PDs reported paying closer attention to clerkship comments and grades in pre-clerkship blocks in the Medical Student Performance Evaluations (MSPEs), and one program chose to wait to review MSPEs before sending interview invitations. PDs suggested the release of MSPEs at the same time as the ERAS opening would allow for more applicant information to be available for review.

Programs agreed blinding Step 1 scores encouraged careful consideration of applicant materials for qualities deemed to be important to their mission and did not constitute a "risk" to their programs. Two programs stated they interviewed applicants they would have missed if they had screened by Step 1 data.

Programs Participating for a Second Year (n = 4, year 2): Similar processes were used for recruitment in the second year. The process of blinding scores was still the most time-consuming aspect. All programs estimated it was less stressful and overall quicker than the previous year. One program estimated 5 extra minutes per application was required to blind Step 1 data.

Three programs informed all interviewees about their participation in the study, while one discussed it only if an applicant asked. Programs reported

TABLE 2
Reasons for Step 1 Examination Scores Unblinding in Order of Decreasing Frequency

Source	Year 1 (2018–2019), n (%)	Year 2 (2019–2020), n (%)
Score disclosed in MSPE	126 (63)	82 (52)
USMLE alert indicator in ERAS notified of irregularity with Step data	33 (16)	1 (0.5)
Score disclosed in personal statement	11 (6)	34 (21)
Score disclosed in advising meeting with PD	9 (5)	0 (0)
Score disclosed in curriculum vitae	6 (3)	1 (0.5)
Score disclosed in letter of recommendation	6 (3)	12 (8)
Score disclosed by applicant during interview	4 (2)	5 (3)
Score disclosed in applicant materials–unspecified	2 (1)	18 (11)
Score disclosed by applicant in email	1 (0.5)	6 (4)
Score known to PD because applicant applied previous year	1 (0.5)	0 (0)
Total	199	159

Abbreviations: MSPE, Medical Student Performance Evaluation; USMLE, United States Medical Licensing Examination; ERAS, Electronic Residency Application System; PD, program director.

applicants continued to have an overall positive response to the blinding of Step 1 data.

Rank List Results

Year 1: Three of the 4 programs unblinded to Step 1 scores before creating final rank lists, while one program chose not to look at Step 1 data at all. One program moved one applicant from spot 62 to 100 on their final list.

Year 2: Four of the 6 programs unblinded to Step 1 scores before creating final rank lists, while 2 chose not to look at Step 1 data at all. One program moved 3 applicants from spot 79 to 178, 144 to 175, and 208 to 219 on their final list.

Discussion

To our knowledge, this is the first study to demonstrate feasibility for programs to blind themselves to Step 1 scores and the number of attempts throughout the entire recruitment process. Programs developed their own processes to blind to Step 1 data; the most common reasons for unblinding events are shown in TABLE 2 and highlight areas that require vigilance to remove this data from the recruitment process. The 2 programs with the highest unblinding rates in year 1 had lower rates in year 2, possibly reflecting an improvement in the blinding process for these programs.

The most common factors used by programs in the absence of Step 1 data are listed in the BOX. Though some programs did use metrics they had not utilized before, they frequently considered factors commonly used in holistic review. Programs commented that blinding to Step 1 scores encouraged careful consideration of applicant materials for qualities deemed to be most important to their missions, and 2 commented specifically that if they had screened by Step 1 scores, they would have missed applicants felt to be an excellent fit for their program. All programs planned to continue using other screening variables in the absence of Step 1 scores.

Programs suggested the release of MSPEs at the time of ERAS opening would allow for all applicant information to be available for consideration from the beginning. Though the MSPE has been criticized for not providing enough objective or individualized information to help programs select applicants to interview,[9] PDs found this document useful.

Though all programs did consider Step 2 CK scores as part of their algorithms, none used this variable alone. While Step 2 CK is thought to be a better predictor of resident performance,[10,11] merely using Step 2 CK in substitution for Step 1 will not address the fundamental issue of relying on a single metric when selecting applicants to interview. Because there is no current proposed plan to change USMLE Step 2 CK reporting, findings from this study may be applicable for programs that want to blind to Step 2 CK data.

BOX Factors Used in Absence of Step 1 Data

- AOA/Gold Humanism Honor Society
- Commitment to underserved communities
- Completed audition rotation with program
- Connection to Arizona/Southwest
- Languages spoken
- Leadership experience
- Leave of absence for academic reasons
- Mission fit
- MSPEs
- Research experience
- Schools with history of previous match at UACOM-T
- Specialty specific pre-clerkship performance
- USMLE Step 2 CK/COMLEX Level 2 CE Scores

Abbreviations: AOA, Alpha Omega Alpha; MSPE, Medical Student Performance Evaluation; UACOM-T, University of Arizona College of Medicine-Tucson; USMLE, United States Medical Licensing Examination; CK, Clinical Knowledge; COMLEX, Comprehensive Osteopathic Medical Licensing Examination; CE, Cognitive Examination.

During the first year of each program's participation, 4 of 6 programs estimated an additional 5% to 20% time was required to blind to Step 1 data; 2 programs reported no additional time was needed. Overall, the blinding process was less time-consuming for programs participating for a second year.

This study has limitations. Since participation was voluntary and involved a single institution, our findings may not be generalizable. Second, all unblinding episodes were self-reported. Not only might this have led to reporting bias, but any unreported unblinding episodes may have had greater influence on recruitment decisions than realized. Third, in year 1 of the study, the weekly survey data completion rate was not 100%. Though programs were asked to account for unblinding episodes that were not reported the following week, underreporting of unblinding episodes may have occurred. Fourth, the choice of alternative variables in the absence of Step 1 data was left to the discretion of the programs and may not necessarily be better predictors of resident performance; factors best equated with resident performance still need further investigation. Finally, our survey instruments were novel, and survey question data were not formally developed. While the authors have significant experience in

GME, survey validation would be needed in future studies.

With the approaching change in Step 1 reporting, residency programs in all specialties will need to decide how to select and rank applicants in the absence of this information. This study provides foundational information which can be used by programs that seek to develop new recruitment processes in the absence of Step 1 scores. Future work will be important to determine which factors may be most applicable to other specialties or institutions.

Conclusions

This study demonstrated that it is feasible to blind programs to Step 1 data in residency recruitment, and only one program changed their final rank list both years after Step 1 scores became known.

References

1. United States Medical Licensing Examination. Change to pass/fail score reporting for Step 1. https://www.usmle.org/incus/. Accessed January 14, 2021.
2. National Board of Osteopathic Medical Examiners. COMLEX-USA Level 1 to Eliminate Numeric Scores. https://www.nbome.org/news/comlex-usa-level-1-to-eliminate-numeric-scores/. Accessed January 14, 2021.
3. Jones MD, Yamashita T, Ross RG, Gong J. Positive predictive value of medical student specialty choices. *BMC Med Educ*. 2018;18(33):1–7. doi:10.1186/s12909-018-1138-x.
4. McGaghie WC, Cohen ER, Wayne DB. Are United States Medical Licensing Exam Step 1 and 2 scores valid measures for postgraduate medical residency selection decisions? *Acad Med*. 2011;86(1):48–52. doi:10.1097/ACM.0b013e3181ffacdb.
5. Rubright JD, Jodoin M, Barone MA. Examining demographics, prior academic performance, and United States Medical Licensing Examination scores. *Acad Med*. 2019;94(3):364–370. doi:10.1097/ACM.0000000000002366.
6. Wusu MH, Tepperberg S, Weinberg JM, Saper RB. Matching our mission: a strategic plan to create a diverse family medicine residency. *Fam Med*. 2019;51(1):31–36. doi:10.22454/FamMed.2019.955445.
7. Association of American Medical Colleges. Aarons CB. Shaking up residency program admissions. May 14, 2019. https://www.aamc.org/news-insights/insights/shaking-residency-program-admissions. Accessed January 14, 2021.
8. Brustman LE, Williams FL, Carroll K, Lurie H, Ganz E, Langer O. The effect of blinded versus nonblinded interviews in the resident selection process. *J Grad Med Educ*. 2010;2(3):349–353. doi:10.4300/JGME-D-10-00051.1.
9. Andolsek KM. Improving the medical student performance evaluation to facilitate resident selection. *Acad Med*. 2016;91(11):1475–1479. doi:10.1097/ACM.0000000000001386.
10. Lee M, Vermillion M. Comparative values of medical school assessments in the prediction of internship performance. *Med Teach*. 2018;40(12):1287–1292. doi:10.1080/0142159X.2018.1430353.
11. Sharma A, Schauer DP, Kelleher M, Kinnear B, Sall D, Warm E. USMLE Step 2 CK: best predictor of multimodal performance in an internal medicine residency. *J Grad Med Educ*. 2019;11(4):412–419. doi:10.4300/JGME-D-19-00099.1.

All authors are with the University of Arizona College of Medicine–Tucson. **Kathy W. Smith, MD,** is Associate Professor, Department of Psychiatry, and Assistant Dean of Student Affairs; **Richard Amini, MD,** is Professor, Department of Emergency Medicine, and Assistant Dean of Student Affairs; **Madhulika Banerjee, BS,** is a Fourth-Year Medical Student; and **Conrad J. Clemens, MD, MPH,** is Professor, Department of Pediatrics and Public Health, and Senior Associate Dean of Graduate Medical Education.

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

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

This work was previously presented at the National Residency Matching Program Conference, Chicago, IL, October 3–5, 2019; Association of American Medical Colleges Annual Meeting, November 8–12, Phoenix, AZ; and National Residency Matching Program Conference, October 8, 2020.

Corresponding author: Kathy W. Smith, MD, University of Arizona College of Medicine-Tucson, kwsmith@arizona.edu

Received June 18, 2020; revisions received October 11, 2020, and December 29, 2020; accepted January 5, 2021.