

The AAMC Medical Student Performance Evaluation Task Force Recommendations: Do They Go Far Enough?

Jillian S. Catalanotti, MD, MPH
Kathryn M. Andolsek, MD, MPH
Jeffrey S. Berger, MD, MBA

Autumn: the season that turns foliage red and program directors' hair gray. In 2016, residency programs received an average of 912 applications, with some receiving more than 5 times this amount.^{1,2} To make well-informed interview and rank decisions, program directors require transparent information about applicants that is efficiently and consistently presented. As the Association of American Medical Colleges (AAMC) Medical Student Performance Evaluation (MSPE) Task Force has released new recommendations for the MSPE, we are commenting on the new recommendations from the perspective of residency program leaders.

The MSPE remains the third most common factor cited by residency program directors in selecting applicants to interview.¹ Despite the AAMC's prior efforts to encourage medical school deans to include comparative information and standardize MSPEs,³ compliance with both the letter and spirit of these recommendations has been limited.⁴ MSPEs that provide a standardized, transparent, and comparative assessment and a holistic review of students would not only improve program director efficiency in reviewing applications, but would also likely result in matching students to programs that best suit their strengths.

As part of its "Optimizing Graduate Medical Education" initiative, the AAMC recently released "Recommendations for Revising the Medical Student Performance Evaluation (MSPE)"⁵; the document emphasizes that the MSPE is a "letter of evaluation, not a letter of recommendation." Guiding principles for the MSPE include that it should be standardized and transparent, allow for more holistic review of applicants, provide supplemental value to the Electronic Residency Application Service (ERAS) application (including comparative information in both qualitative and quantitative assessments), and evaluate applicants on the competencies relevant to success

in residency. While we strongly agree with these guiding principles, and feel that they are important next steps, we believe the AAMC's recommendations could go even further.

Comparative Assessment

As applications to residency programs balloon, program directors must strategically choose whom to offer an interview slot. Comparing medical students is understandably worrisome to medical school deans, who may fear reputational damage from failure to match all students. However, transparent assessment of students in MSPEs would facilitate the honest conversations needed when advising students about both specialty and program choices. For example, deans frequently lament the student who fails to heed clear advice, but students, inexperienced in the nuances of MSPE language, see a glowing letter and may be emboldened to aim high. Program directors distrust the linguistic gamesmanship inherent to many MSPEs. They understand that medical students largely represent the most accomplished individuals in society. At the same time, program directors recognize that half of all students must be assigned to the bottom 50% of the class. Most graduates become wonderful physicians, given the proper environment, and programs are highly motivated to fill all available slots.

By design, the National Resident Matching Program requires residency programs to rank order applicants, making comparative information on overall performance crucial.⁶ Ideally, comparative data would be standardized not only within, but also between medical schools to allow program directors to make meaningful comparisons among applicants. Within individual schools, summative comparative assessments must create subgroups that are small enough to be meaningful, and must provide the distribution of students within each group. Although the AAMC has recommended comparative data with a key for interpretation in Appendix D for more than

a decade, only 61% of schools' MSPEs provide complete comparative keyword data with a full list of terms and distribution.⁴ In addition, keywords lack standardization and group sizes vary; program directors reading hundreds of applications encounter "excellent" students ranging from the 1st through 95th percentiles compared with their classmates.⁷ In a recent report, 10% of schools listed the size distribution of only their top group, an act that obfuscates rather than informs the reader.⁴ Separating students into a "top 10%," followed by a group containing the next 70%, fails to provide useful comparative information for the majority of students.

Thus, we echo the call by Boysen Osborn et al⁷ for consistent names and sizes of comparative groups across schools, including the existence of a bottom group that may be variable in size and contains only those students with significant performance issues, such as course failures or professionalism concerns. We further recommend that course failures and professionalism be separated, as the ability to effectively remediate may vary between them.

Separating overall comparative assessments into 3 axes (such as academic performance, service/leadership, and scholarly activity) may ease the fears of medical school deans about the condemning effects of 1 overall assessment, and may help achieve "best fit" matches both for applicants and programs. If schools created standardized benchmarks for each of these domains, students could be accurately compared across schools. Residency programs with robust research opportunities may prioritize students with demonstrated excellence in scholarly activity, while those that provide patient care to underserved communities may prioritize students with demonstrated excellence in service.

Concordance Between Grades and Narrative

We agree with the AAMC's recommendation that in addition to overall clerkship grades, MSPEs should state the defined components and weights that are used to calculate them. It is essential to include a student's score for each component, especially when the relative weights are similar. For example, because many program directors prioritize applicants' clinical performance level, reporting an overall clerkship grade and noting that there is a 33% contribution from clinical score, shelf examination score, and a course project may not be as helpful as the grade for each component.

The validity and utility of the MSPE rests on the degree of concordance between grades and narrative comments. Often, MSPEs for students in the bottom

BOX Summary of Recommendations

- Report summative comparative assessments with small subgroups and transparent distributions that are consistent across schools; include a bottom group for students with course failures. Professionalism issues should be separated out.
- Develop overall comparative assessments with standardized benchmarks in 3 domains: academic performance, service/leadership, and scholarly activity.
- State the defined components and weights that are used to calculate overall clerkship grades.
- Use standardized anchoring questions to improve concordance between rotation grades and narrative comments.
- When describing an unusual hardship, allow the flexibility to expand beyond 2 sentences of description.
- Align the Medical Student Performance Evaluation (MSPE) release date to coincide with Electronic Residency Application Service application submission.
- After a suitable comment and response period, piloting, and modifications, mandate universal adoption of MSPE recommendations and establish rules for oversight and enforcement.

quartile appear to have been scrubbed clean and lack critical appraisal. Ironically, program directors often find "negative" comments only in MSPEs of those students near the top of their class. The recommendations for a new MSPE should repair this disconnect. One way to improve concordance between grades and narratives is to use anchor questions for each grade level on a given rotation. For example, if a faculty member selects a below average grade for a rotation (often "pass"), his or her narrative comments would be guided by specific questions, such as "What aspects of this student's performance were below the level of his or her peers?" Narrative responses, unedited for content, would be transferred to the MSPE to provide context for applicants' grades.

Holistic Review

The MSPE should provide information needed for holistic review of applicants, especially those who come from backgrounds historically underrepresented in medicine and those who have overcome personal or academic hardship. The "noteworthy characteristics" section is best used in this way, and should provide verifiable context for academic hardships during medical school. This information may not be available elsewhere in the ERAS application, and is of key importance to accurately assess students' medical school performance and the possible need for additional support during residency. Students with academic records at or above average likely would not require more than the AAMC-recommended 3

brief bullet points.⁵ However, we fear that this limitation may disadvantage those students who truly have overcome hardship before and/or during medical school, and have complex stories that cannot be “described in 2 sentences or less.”⁵ Inclusion of this information in the MSPE may free these students to use their personal statements to focus on other areas of their experience or career plans. It would also better allow program directors to tailor training to incoming interns’ needs, including early initiation of study plans, assigning mentors, or adjusting schedules to ensure a successful transition.

Timing

Currently, the MSPE becomes available in ERAS on October 1, which is 2 weeks after applications become available to program directors. The Task Force does not address the nonsynchronous release of the MSPE in their recommendations.⁵ Anecdotally, many programs offer interviews prior to the release of the MSPE, despite the fact that this compromises holistic review. The degree to which this represents competitive market forces versus a disregard for the value for the MSPE is uncertain. Aligning the MSPE release to coincide with application submissions would enhance the holistic review of applicants. This change would level the playing field for program directors who prioritize holistic review and value the MSPE, but currently feel compelled to offer interviews before October 1 rather than risk applicants filling their schedules with interviews from other programs before MSPEs are released.

Enforcement

After a suitable comment and response period, piloting, and modifications, the AAMC recommendations should be adopted by all schools, without exception. The Liaison Committee on Medical Education could add a standard for oversight of institutional compliance with the AAMC’s MSPE recommendations as part of the self-study process. Alternatively, the AAMC could enforce recommendations through their management of ERAS, prohibiting noncompliant MSPEs from being uploaded.

Future Directions

We welcome the new AAMC recommendations for the MSPE but believe they could go further in some areas. MSPEs need more specific language to provide transparent comparative assessment, concordance between grades and narrative comments, a timely release, and enhanced holistic narrative—particularly for students who have overcome personal and/or

academic hardships (BOX). We believe that MSPEs with these characteristics will benefit program directors and students by increasing efficiency, improving appropriateness of fit between matched interns and programs, and enabling early intervention to support the success of lower-performing students.

We look forward to the next phase of the MSPE Task Force’s work, including their recommendations for reporting negative actions after October 1, which we strongly support in the interest of transparency and successful learner handoffs. Finally, we encourage the AAMC to gather comments and responses; to conduct trials of the recommendations; to measure the impact of these new recommendations on all stakeholders, including students and program directors; and to adopt the final recommendations as mandatory for all institutions with robust enforcement.

References

1. National Resident Matching Program. Results of the 2016 NRMP program director survey. <http://www.nrmp.org/wp-content/uploads/2016/09/NRMP-2016-Program-Director-Survey.pdf>. Accessed November 8, 2016.
2. Association of American Medical Colleges. Statistics. <https://www.aamc.org/services/eras/stats/359278/stats.html>. Accessed November 8, 2016.
3. Association of American Medical Colleges. A guide to the preparation of the medical student performance evaluation. 2002. <https://www.lmunet.edu/uploads/dcom/pdfs/aamc-mspe-guidelines.pdf>. Accessed November 8, 2016.
4. Hom J, Richman I, Hall P, et al. The state of medical student performance evaluations: improved transparency or continued obfuscation? *Acad Med*. 2016;91(11):1534–1539.
5. Association of American Medical Colleges. Recommendations for revising the medical student performance evaluation (MSPE). <https://www.aamc.org/download/470400/data/mspe-recommendations.pdf>. Accessed November 8, 2016.
6. Benziger R. From the desk of the program director: show me the rankings! *J Grad Med Educ*. 2014;6(3):413–414.
7. Boysen Osborn M, Mattson J, Yanuck J, et al. Ranking practice variability in the medical student performance evaluation: so bad, it’s “good.” *Acad Med*. 2016;91(11):1540–1545.



Jillian S. Catalanotti, MD, MPH, is Associate Professor of Medicine and of Health Policy Management, and Internal Medicine Residency Program Director, The George Washington University School of Medicine & Health Sciences; **Kathryn M. Andolsek, MD, MPH**, is Associate Editor, *Journal of Graduate*

EDITORIAL

Medical Education (JGME), and Professor of Community and Family Medicine and Assistant Dean for Premedical Education, Duke University School of Medicine; and **Jeffrey S. Berger, MD, MBA**, is Associate Editor, *JGME*, and Associate Professor of Anesthesiology and Critical Care Medicine, Associate Dean for Graduate Medical Education, and Designated Institutional Official, The George Washington University School of Medicine & Health Sciences.

The authors would like to thank Ms Laura Abate for her assistance with format and editing.

Corresponding author: Jillian S. Catalanotti, MD, MPH, Medical Faculty Associates of The George Washington University, Division of General Internal Medicine, 2150 Pennsylvania Avenue NW, Suite 5-416, Washington, DC 20037, 202.741.2621, fax 202.741.2788, jcatalanotti@mfa.gwu.edu