Current State of the Medical Student Performance Evaluation: A Tool for Reflection for Residency Programs

Judith M. Brenner, MD
Jeffrey B. Bird, MA
Jason Brenner, BS
David Orner, MPH
Karen Friedman, MS, MD

ABSTRACT

Background The Medical Student Performance Evaluation (MSPE) provides important information to residency programs. Despite recent recommendations for standardization, it is not clear how much variation exists in MSPE content among schools.

Objectives We describe the current section content of the MSPE in US allopathic medical schools, with a particular focus on variations in the presentation of student performance.

Methods A representative MSPE was obtained from 95.3% (143 of 150) of allopathic US medical schools through residency applications to the Zucker School of Medicine at Hofstra/Northwell in select programs for the 2019–2020 academic year. A manual data abstraction tool was piloted in 2018–2019. After training, it was used to code all portions of the MSPE in this study. The results were analyzed, and descriptive statistics were reported.

Results In preclinical years, 30.8% of MSPEs reported data regarding performance of students beyond achieving "passes" in a pass/fail curriculum. Only half referenced performance in the fourth year including electives, acting internships, or both. About two-thirds of schools included an overall descriptor of comparative performance in the final paragraph. Among these schools, a majority provided adjectives such as "outstanding/excellent/very good/good," while one-quarter reported numerical data categories. Regarding clerkship grades, there were numerous nomenclature systems used.

Conclusions This analysis demonstrates the existence of extreme variability in the content of MSPEs submitted by US allopathic medical schools in the 2019–2020 cycle, including the components and nomenclature of grades and descriptors of comparative performance, display of data, and inclusion of data across all years of the medical education program.

Introduction

The Medical Student Performance Evaluation (MSPE), formerly known as the "Dean's letter," is a key application component for residency program directors. In the 2018 NRMP Program Director Survey,[1] the MSPE ranked third in terms of importance in the residency application process, with 81% of respondents citing it as a factor behind only USMLE Step 1 and letters of recommendation in the specialty. With the change in USMLE Step 1 to pass/fail reporting in the near future, the MSPE may become an even more important component of the residency application.[2] An essential element of undergraduate medical education (UME) to graduate medical education (GME) communication, the MSPE is a comprehensive summary of a medical student's performance across 3 plus years of medical school. The MSPE is intended to provide an honest and objective summary of a student's personal attributes, experiences, and academic accomplishments.[3] Historically, program directors and other end users have found the MSPE difficult to decipher, widely variable among schools, and lacking in transparency.[2,4–11]

In an effort to improve the usefulness of the MSPE, the Association of American Medical Colleges (AAMC) convened an MSPE Task Force, which made recommendations in 2016.[3] These recommendations, which included page length and uniformity of presentation, were intended to make the letter easier to read and interpret as well as attain "a level of standardization and transparency that facilitates the residency selection process."[3]

To date, most US allopathic medical schools have adopted the 2016 MSPE Task Force recommendations in terms of overall structure, guided by MSPE Task Force templates.[12,13] The next step is to determine whether there is variability in the data presented in each MSPE section. Variability can present itself in several ways, including content chosen to represent the category (eg, components of clerkship grade), visual display of the content (eg,

DOI: http://dx.doi.org/10.4300/JGME-D-20-01373.1

type of graphic display), and nomenclature (eg, different adjectives).

The aim of this study is to review the content and presentation variations in student performance data in US MSPEs.

Methods

A single MSPE was obtained from 95.3% (143 of 150) of allopathic US medical schools with graduating seniors through applications to the Zucker School of Medicine at Hofstra/Northwell residency programs in internal medicine, dermatology, orthopedic surgery, and urology for the 2019–2020 academic year.

A manual data abstraction tool was created (J.B.B., D.O.) to align with the 6 major sections in the MSPE Task Force recommendations: identifying information, noteworthy characteristics, academic history, academic progress, summary, and medical school information. Additional data abstracted focused on content of the sections: (1) performance in the preclinical years, (2) performance in the fourth-year experiences, (3) grades in the clinical year, (4) summary adjective, (5) summary narrative, and (6) authorship of the MSPE. The manual abstraction tool is available from the authors upon request.

The data abstraction tool was piloted using MSPEs from the 2018–2019 academic year. Pilot testing was conducted by 2 authors (J.M.B., J.B.) who individually reviewed 13 MSPEs and adjudicated discrepancies by discussion and constituted training. Following training, data from MSPEs from the 2019–2020 application season were abstracted by a single author (J.B.) with ongoing discussion with an additional author (J.M.B.). All data were collected in Microsoft Excel. Data analysis was performed to determine descriptive statistics.

There was no funding for the project, and the Northwell Institutional Review Board deemed this educational project exempt from review.

Results

A total of 143 MSPEs were reviewed from schools across the United States (Northeast 24.5%, Midwest 24.5%, South 38.5%, and West 12.5%; TABLE). About one-third (30.8%, 44 of 143) of MSPEs included a narrative of student performance in the preclinical curriculum, while the remaining 69.2% (99 of 143) either omitted data or reported passing all coursework. About half (51.7%, 74 of 143) referenced performance in the fourth year, including electives, acting internships, or both. Almost all schools (98.6%, 141 of 143) included clerkship grades in the academic progress section, 81.6% (115 of 141) included what determines clerkship grades, and 88.7% (125 of 141) included a grade distribution. The visual display of this data varied greatly among schools with no clear standard identified. In addition, there were numerous variations in grade type (eg, honors type vs numerical). About two-thirds (69.2%, 99 of 143) included an overall descriptor of comparative performance (ie, adjective) in the final paragraph, and 92.9% (92 of 99) provided detail on what components contributed to the comparative performance. Representative examples are provided in the BOX. Beyond the descriptor term, 60.8% (87 of 143) provided a narrative describing the student in the summary paragraph. Lastly, more than half of MSPEs (58.0%, 83 of 143) were signed by a dean of student affairs, leaving 42.0% (60 of 143) signed by other faculty. Among the "others" were the dean, career advisors, academic advisors, and select faculty.

BOX Examples of Descriptors of Comparative Performance

- Exemplary, superior, very strong, strong, satisfactory
- Superlative, exceptional, excellent, very good, good
- Highest, most highly, highly, recommend, satisfactory
- 90+, 75–89, 50–74, 25–49, 0–24
- No. of letters of distinction
- Lower, middle, highest

Discussion

In moving past mere compliance with the 2016 MSPE Task Force guidelines, this analysis demonstrates the existence of extreme variability in the components and nomenclature of grades and descriptors of comparative performance (ie, adjectives), the display of data, and the inclusion of all years of the curriculum in the MSPE across US allopathic medical schools. The only consistent data included were the clerkship grades. Furthermore, the letters were signed by individuals with a variety of roles.

The clerkship grades, while consistently reported by most schools, is an area ripe for misinterpretation and creates barriers in making comparisons across schools due to the variations in content. While 81.6% of schools report on the assessments that contribute to clerkship grades, the actual components vary from school to school. The vast majority (88.7%) of schools include the grade distribution; however, there are numerous variations in grade type (eg, honors type vs numerical). This is compounded by variations in visual display, making readability challenging. Turning to clinical literature, consistency in reporting is considered necessary for interpretation.[14] Similarly, business literature has suggested that readability

TABLE
Medical Student Performance Evaluation (MSPE) Content Analysis

Content	n (%)
MSPE details[a]	
Preclinical years: personal narrative	44 (30.8)
Clerkship grades	141 (98.6)
Fourth-year performance (acting internships, electives)	74 (51.7)
Summary adjective	99 (69.2)
Components contributing to adjective[b]	92 (92.9)
Narrative in summary	87 (60.8)
Dean of student affairs authorship	83 (58.0)
Clerkship grade types[c]	
Variant of honors type	86 (61.0)
Variant of letter grade type	22 (15.6)
Variant of numerical type	6 (4.3)
Variant of pass/fail	4 (2.8)
Other	23 (16.3)
Summary adjective types[b]	
Variant of outstanding type	69 (69.7)
Variant of numerical type	25 (25.3)
Other	5 (5.0)
Components contributing to adjective types[d]	
Whole curriculum	46 (50.0)
Clerkship performance only	26 (28.3)
Holistic review	9 (9.8)
Inclusion of USMLE Step 1 score	9 (9.8)

Abbreviation: USMLE, United States Medical Licensing Examination.
[a] Percent of details included in MSPE out of 143 schools.
[b] Percent of MSPEs out of those that provided a summary adjective (n = 99).
[c] Percent of MSPEs out of those that provided clerkship grades (n = 141).
[d] Percent of MSPEs out of those that provided components contributing to adjective (n = 92).

variability on annual reports results in obfuscation.[15] Furthermore, extreme variability appears to be at odds with the MSPE Task Force's stated goals to "achieve a level of standardization and transparency that facilitates the residency selection process."[3] Currently, program directors are expected to read and interpret nearly 150 different versions of the MSPE.

The issue of standardization has been tackled by specialty organizations, most notably emergency medicine in the development of the Student Letter of Evaluation (SLOE). The SLOE, first established in 1996, has evolved over time to be a more standardized, concise, and discerning document.[16] However, like the MSPE, it has limitations related to the reliability of data included, consistency across letter writers (individual versus group SLOE), and discrimination of applications due to limited spread of applicants across rating categories.[17–19] The question still remains of what is the "right" amount of standardization in MSPEs. Ideally, the letter should highlight the unique features of an applicant in a document that enables end users to evaluate candidates critically and more easily across multiple different schools.[20] Consider the metaphor of a car. Each has unique features, but all manufacturers report the miles per gallon as a standard metric of comparison. Might the equivalent in UME be a welcome addition?

This study was limited by having a single person extract data from each MSPE. Despite perceived consistency of MSPEs by school based on author's (K.F.) experience as program director, an additional limitation is the selection of a single MSPE per school for analysis. Analysis of a single MSPE may not have captured additional formats if more than one was used by a single school. Lastly, the MSPEs used were sent to a small number of residency programs at a single institution.

Ultimately, it is important that the UME and GME communities serve learners across the continuum. Thus, the community needs to address whether the

current degree of variability in the presentation of data is serving its learners best. This question will only be answered by considering the perspective of the primary MSPE "consumers," most often program directors who must address what data they need in order to make the best selections of learners for their programs. Thus, future areas of study should include investigating more standardized ways to present data as well as exploring ways to facilitate comparisons of applicants across institutions with standardized key metrics while ultimately promoting readability by the end user.

Conclusions

This analysis demonstrates the existence of extreme variability in the content of MSPEs submitted by US allopathic medical schools in the 2019–2020 academic year, including the components and nomenclature of grades and descriptors of comparative performance (ie, adjectives), the display of data, and inclusion of data across all years of the medical education program.

References

1. National Resident Matching Program. Results of the 2018 NRMP Program Director Survey. https://www.nrmp.org/wp-content/uploads/2018/07/NRMP-2018-Program-Director-Survey-for-WWW.pdf. Accessed April 30, 2021.
2. Bird JB, Friedman K, Arayssi T, Olvet D, Conigliaro R, Brenner J. Review of the Medical Student Performance Evaluation: analysis of the end-users' perspective across the specialties. *Med Educ Online*. 2021;26(1):1876315. doi:10.1080/10872981.2021.1876315
3. Association of American Medical Colleges. Recommendations for Revising the Medical Student Performance Evaluation (MSPE). https://www.aamc.org/system/files/c/2/470400-mspe-recommendations.pdf. Accessed April 30, 2021.
4. Ozuah PO. Variability in deans' letters. *JAMA*. 2002;288(9):1061.
5. Westerman ME, Boe C, Bole R, et al. Evaluation of medical school grading variability in the United States: are all honors the same? *Acad Med*. 2019;94(12):1939–1945. doi:10.1097/acm.0000000000002843
6. 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. doi:10.1097/acm.0000000000001034
7. Ward MA, Palazzi DL, Lorin MI, Agrawal A, Frankenthal H, Turner TL. Impact of the final adjective in the Medical Student Performance Evaluation on determination of applicant desirability. *Med Educ Online*. 2018;23(1):1542922. doi:10.1080/10872981.2018.1542922
8. Fagan R, Harkin E, Wu K, Salazar D, Schiff A. The lack of standardization of allopathic and osteopathic medical school grading systems and transcripts. *J Surg Educ*. 2020;77(1):69–73. doi:10.1016/j.jsurg.2019.06.016
9. Shea JA, O'Grady E, Wagner BR, Morris JB, Morrison G. Professionalism in clerkships: an analysis of MSPE commentary. *Acad Med*. 2008;83(suppl 10):1–4. doi:10.1097/ACM.0b013e318183e547
10. Boysen-Osborn M, Yanuck J, Mattson J, et al. Who to interview? Low adherence by U.S. medical schools to Medical Student Performance Evaluation format makes resident selection difficult. *West J Emerg Med*. 2017;18(1):50–55. doi:10.5811/westjem.2016.10.32233
11. Brenner JM, Arayssi T, Conigliaro RL, Friedman K. The revised Medical School Performance Evaluation: does it meet the needs of its readers? *J Grad Med Educ*. 2019;11(4):475–478. doi:10.4300/jgme-d-19-00089.1
12. Giang D. Medical Student Performance Evaluation (MSPE) 2017 task force recommendations as reflected in the format of 2018 MSPE. *J Grad Med Educ*. 2019;11(4):385–388. doi:10.4300/jgme-d-19-00479.1
13. Hook L, Salami AC, Diaz T, Friend KE, Fathalizadeh A, Joshi ART. The revised 2017 MSPE: better, but not "outstanding." *J Surg Educ*. 2018;75(6):e107–e111. doi:10.1016/j.jsurg.2018.06.014
14. Tunis SR, Clarke M, Gorst SL, et al. Improving the relevance and consistency of outcomes in comparative effectiveness research. *J Comp Eff Res*. 2016;5(2):193–205. doi:10.2217/cer-2015-0007
15. Hassan MK, Abu Abbas B, Garas SN. Readability, governance and performance: a test of the obfuscation hypothesis in Qatari listed firms. *Corp Gov Int J Bus Soc*. 2019;19(2):270–298. doi:10.1108/CG-05-2018-0182
16. Council of Residency Directors in Emergency Medicine. The Standardized Letter of Evaluation (SLOE) for Off-Service or Other Rotation. https://www.cordem.org/resources/residency-management/sloe/osloe-instructions/. Accessed April 30, 2021.
17. Negaard M, Assimacopoulos E, Harland K, Van Heukelom J. Emergency medicine residency selection criteria: an update and comparison. *AEM Educ Train*. 2018;2(2):146–153. doi:10.1002/aet2.10089
18. 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

19. 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
20. Hauer KE, Giang D, Kapp ME, Sterling R. Standardization in the MSPE: key tensions for learners, schools, and residency programs. *Acad Med*. 2021;96(1):44–49. doi:10.1097/acm.0000000000003290

Judith M. Brenner, MD, is Associate Dean for Curricular Integration and Assessment, and Associate Professor of Science Education and Medicine, Donald and Barbara Zucker School of Medicine at Hofstra/Northwell; **Jeffrey B. Bird, MA,** is Educational Research & Strategic Assessment Analyst, and Assistant Professor of Science Education, Donald and Barbara Zucker School of Medicine at Hofstra/Northwell; **Jason Brenner, BS,** is a Volunteer Research Assistant, Donald and Barbara Zucker School of Medicine at Hofstra/Northwell, and Student, University of Michigan; **David Orner, MPH,** is a Research Assistant, Donald and Barbara Zucker School of Medicine at Hofstra/Northwell; and **Karen Friedman, MS, MD,** is Vice Chair for Education, Department of Medicine, Northwell Health, and Professor of Medicine, Donald and Barbara Zucker School of Medicine at Hofstra/Northwell.

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

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

Part of this data was previously presented at the AAMC MedBiquitous Community Connection Webinar, September 29, 2020.

The authors would like to thank the following individuals for their contribution to this work: Krista Paxton, Saori Wendy Herman, Doreen Olvet, and Joanne Willey.

Corresponding author: Judith M. Brenner, MD, Donald and Barbara Zucker School of Medicine at Hofstra/Northwell, judith.m.brenner@hofstra.edu

Received November 6, 2020; revisions received February 11, 2021, and April 5, 2021; accepted April 18, 2021.