

The Urgent Need for Reporting Accurate and Fair Student Comparisons in the Medical Student Performance Evaluation

Scott A. Cottrell, EdD

Jason S. Hedrick, PhD

Anna Lama, EdD

Sarah Sofka, MD

Norman D. Ferrari III, MD

A common criticism of the Medical Student Performance Evaluation (MSPE) is that it provides insufficient objective data for residency program directors to use to distinguish applicants.^{1,2} The Association of American Medical Colleges (AAMC) has encouraged schools to include summative assessments into the MSPE for relative medical student comparisons and to differentiate levels of performance.³ Groupings and rankings, however, vary widely across medical schools.⁴ A minority of schools are reporting a quartile, from first (top) to fourth quartile.⁵ Some schools provide ill-defined descriptors to characterize the categories, and the number of students falling into each category is unequal.⁶ An example includes categories ranging from “outstanding, excellent-outstanding, excellent, very good-excellent, very good, and good.” Additionally, these categories are often disproportionately applied across classes with little explanation.⁷ Ambiguous and potentially misleading categories could trigger issues with the National Resident Matching Program. If a residency program matches a student and learns that the MSPE is not “accurate” or incomplete (eg, failing to report professionalism probation), the program may have grounds for a match waiver and potentially a release from a match commitment.⁸ The conclusions drawn from these analyses and supported by the literature are that program directors find it difficult to compare applicants within and across medical schools and may therefore minimize the value of the MSPE.⁹ Medical students, too, have voiced concerns about distinguishing their applications, especially when applying to highly competitive specialties.¹⁰ Accordingly, the authors advocate for best practices to calculate and report a summative assessment into the MSPE. Educators have an obligation to include accurate and fair data in the MSPE, allowing program directors to make informed judgements

concerning residency applications that ultimately have implications on the competency and quality of our future physician workforce.

The Value of Reporting Student Comparisons in the MSPE

Program directors across medical specialties were surveyed about residents’ performance during the first postgraduate year.¹¹ One of the survey items asks: “Reflect on the information provided about this individual from the medical school including the MSPE before or during the transition to residency. Was the information provided useful?” Only 7577 (56%) of the 13 530 respondents answered that the MSPE was helpful; 4938 (36.5%) responded that the MSPE was “somewhat” helpful and 1015 (7.5%) responded that the MSPE “was not” helpful. Some program directors qualified why they felt the MSPE was not helpful. For example, a program director commented in our school-specific survey results:

“The MSPE results are generally of no help and at this point have essentially no discriminatory value in assessing applicants. Particularly with schools no longer categorizing students and most everything is about pass or failing.”

The importance of reporting a summative assessment is punctuated by variability in how schools determine grades. Future modifications to the AAMC recommendations for the MSPE should encourage schools to detail how grades are calculated, allowing program directors to distinguish the different types of grading practices and the potential for grade inflation. Some schools, for example, award a majority of students the highest clerkship grade. As a result, distinguishing students’ clinical performance is nearly impossible.¹² Fair and accurate comparisons, such as a summative quartile ranking of students, help program directors compare students within and between institutions.

DOI: <http://dx.doi.org/10.4300/JGME-D-23-00862.1>

BOX Best Practices for Reporting Accurate and Fair Student Comparisons in the Medical Student Performance Evaluation (MSPE)

- Determine a fair and accurate way to calculate a summative assessment for student comparison.
- Utilize the first attempt only for a course/clerkship/rotation for the calculation of quartiles.
- Communicate the process for calculating a summative assessment (eg, group assignment/quartile/quintiles), in the MSPE, with all stakeholders to enable a fair and transparent process.
- Avoid sharing specific class “ranks” with students (eg, 30 out of 100), which may invite disputes over “tipping” or “falling” into a different group or quartile.
- Avoid descriptors that are ambiguous or exaggerate performance.
- Assign the same number of students in each of the categories or quartiles.
- Encourage contribution from all phases of the curriculum to figure into comparisons.
- Ascertain comparisons between cohorts for students who fall within the same graduating class.
- Report only one final comparison between students in the MSPE.

Best Practices for Reporting Accurate and Fair Student Comparisons in the MSPE

There are a number of ways to determine and report comparisons between students. The authors argue that strategies should include sharing best practices. See the BOX for a summary of best practices, which include:

Determine a fair and accurate way to calculate a summative assessment for student comparison. For example, some schools assign “points” to course grades (eg, fail=0, pass=2, high pass=4, and honors=6). These points are aggregated, and the number of points students acquire relative to their peers are used to ascertain a group assignment. Another common way to ascertain students’ ranking across courses is to utilize T-scores, which have a mean of 50 with a standard deviation of 10. The process of converting students’ final course numerical scores to T-scores allows for comparisons between students across diverse courses. Each course or clerkship has an overall final numerical score between 0 and 100. The “weight,” or credit hours, of each course and clerkship is also considered in the calculation. The calculation to convert a student’s final numerical score for each course is: $T\text{-score} = 50 + [10 \times ((\text{final numerical score} - \text{class mean})/\text{SD})]$.

An example of how to calculate a student’s overall T-score and weighting across courses and academic years is provided in the TABLE.

TABLE

Example of How to Calculate a Student’s Ranking Utilizing T-Scores

Course/ Clerkship	T-Score	Credit Hours	T-Score × Credit Hours
Course A	52	3	156
Course B	61	5	305
Clerkship A	50	8	400
Clerkship B	46	6	276

The total number of credit hours=22 (ie, 3 + 5 + 8 + 6). The total number of T-scores multiplied by credit hours is 1137 (ie, 156 + 305 + 400 + 276). The student in this example earned an overall average T-score of 51.6 (ie, 1137/22=51.6). In order to determine the quartile ranges for all students, arrange all overall average T-scores from least to greatest. The median (50th percentile) of the overall averages distinguishes the class into 2 halves. The median is also determined for each of the 2 halves to ascertain 25th and 75th percentiles, which results in 4 quartiles. Each student’s overall average T-score will fall into 1 of the 4 quartiles, which is designated in the MSPE.

Utilize the first attempt only for a course/clerkship/rotation for the calculation of quartiles. Students who remediate and earn a higher score on the second attempt contribute to an unfair comparison between students. Students who pass the course on the first attempt are not given an opportunity to score higher by repeating the course. The first attempt score for all students, including students who had an incomplete grade or take a leave of absence and then return to the curriculum, are ascertained only when the entire course is completed in full and final grades are submitted.

Communicate the process for calculating a summative assessment (eg, group assignment/quartile/quintiles), in the MSPE, with all stakeholders to enable a fair and transparent process. Our school, for example, includes an “MSPE Policy” in both the student and faculty handbooks, which details how quartiles are calculated and reported in the MSPE.

Avoid sharing specific class “ranks” with students (eg, 30 out of 100), which may invite disputes over “tipping” or “falling” into a different group or quartile. Our school, for example, shares quartile data with students at the end of each academic year, allowing students to monitor their progress throughout the curriculum.

Avoid descriptors that are ambiguous or that exaggerate performance. The distribution between assessment categories such as outstanding, exemplar, and stupendous

are not accurate. Disingenuous attempts to rank students into categories with varnished descriptors only denude the potential of the MSPE to be a valued component of the holistic application review process by program directors. A continuum of quartiles/quintiles is recommended. Schools are encouraged to explicitly label the “top” category, such as first (top), second, third, and fourth quartiles.

Assign the same number of students in each of the categories or quartiles. Some classes will have an odd number and some overall average scores will be identical, making it impossible to have exactly equal groups. The intention of this practice is that students should have an approximately equal chance to be assigned to each category.

Encourage contribution from all phases of the curriculum to figure into comparisons. Students should strive to do their best during all phases of the curriculum, including both pre-clerkship and clerkship courses.

Ascertain comparisons between students who fall within the same graduating class. For example, students who take a leave of absence should be removed from their original cohort until their graduating class is determined.

Report only one final comparison between students in the MSPE. The final comparison can include quartiles, evaluation descriptors, and core competency indicators. A clear and succinct statement can provide a helpful comparison of applicants. For example, the language in the MSPE may include: “The student is ranked in the first (top) quartile academically, which is calculated using first attempt in each course of the first 3 years of the curriculum weighted by the number of credit hours in each course.”

Medical School Concerns

There may be different reasons why schools are reluctant to report summative assessments such as quartiles. Some educators, for example, believe that medical students are an outstanding group of learners and that placing them into ranked groups is not fair.⁴ Educators may also suggest that comparisons between students may foment competition between medical students. However, the benefits of reporting accurate and fair comparisons between students outweigh potential costs. Embedding summative assessments into the MSPE addresses program directors’ calls for more transparency in the MSPE, including summative assessments across medical schools.¹³ Some educators argue that summative assessments are prone

to bias.¹⁴ Indeed, schools should examine patterns of grade distribution across race and other demographics to investigate potential sources of systematic grading biases.¹⁵ Removing objective data from the MSPE, however, may result in unintended disparities for students attending non-marque schools or schools that have limited research opportunities. Holistic review does not mean that objective data should be purged. Rather, data like common quartile rankings complement other information shared in the MSPE, such as “noteworthy characteristics.”

Conclusions

Adopting best practices for accurate and fair comparisons of students is important for our overall health care system by meeting our educational obligation to society. These efforts will advance the reputation and value of the MSPE, which will benefit both educators and students in the residency application process.

References

1. 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
2. Bird JB, Friedman KA, Arayssi T, Olvet DM, Conigliaro RL, Brenner JM. 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). Accessed September 20, 2023. <https://www.aamc.org/media/23311/download?attachment>
4. 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. doi:10.1097/ACM.0000000000001180
5. Brenner JM, Bird JB, Brenner J, Orner D, Friedman K. Current state of the medical student performance evaluation: a tool for reflection for residency programs. *J Grad Med Educ*. 2021;13(4):576-580. doi:10.4300/JGME-D-20-01373.1
6. Kiefer CS, Colletti JE, Bellolio MF, et al. The “good” dean’s letter. *Acad Med*. 2010;85(11):1705-1708. doi:10.1097/ACM.0b013e3181f55a10
7. 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
8. National Resident Matching Program. Match participation agreement for medical schools: 2024 Main Residency Match and Supplemental Offer and Acceptance Program (SOAP). Accessed September 21, 2023. <https://www.nrmp.org/wp-content/uploads/2023/09/2024-MPA-Main-Match-MedicalSchool-2.pdf>
 9. Green M, Jones P, Thomas JX Jr. Selection criteria for residency: results of a national program directors survey. *Acad Med.* 2009;84(3):362-367. doi:10.1097/ACM.0b013e3181970c6b
 10. Girard AO, Qiu C, Lake IV, Chen J, Lopez CD, Yang R. US medical student perspectives on the impact of a pass/fail USMLE Step 1. *J Surg Educ.* 2022;79(2): 397-408. doi:10.1016/j.jsurg.2021.09.010
 11. Association of American Medical Colleges. Resident Readiness Survey Summary Report. Published April 2023. Accessed September 21, 2023. <https://www.aamc.org/data-reports/students-residents/report/rrs-project>
 12. 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
 13. Dambro AB, Newberry Z, Parascando J, Anderson A. Primary care residency perspectives on medical student performance evaluations. *PRiMER.* 2023;7:12. doi:10.22454/PRiMER.2023.763651
 14. Lucey CR, Hauer KE, Boatright D, Fernandez A. Medical education’s wicked problem: achieving equity in assessment for medical learners. *Acad Med.* 2020;95(suppl 12):98-108. doi:10.1097/ACM.0000000000003717
 15. The Coalition for Physician Accountability’s Undergraduate Medical Education-Graduate Medical Education Review Committee (UGRC). Recommendations for Comprehensive Improvement of the UME-GME Transition. Accessed September 20, 2023. <https://physicianaccountability.org/wp-content/uploads/2021/08/UGRC-Coalition-Report-FINAL.pdf>



Scott A. Cottrell, EdD, is Professor, Department of Medical Education, West Virginia University School of Medicine, Morgantown, West Virginia, USA; **Jason S. Hedrick, PhD**, is Assistant Professor, Department of Medical Education, West Virginia University School of Medicine, Morgantown, West Virginia, USA; **Anna Lama, EdD**, is Assistant Professor, Department of Medical Education, West Virginia University School of Medicine, Morgantown, West Virginia, USA; **Sarah Sofka, MD**, is Professor and Vice Chair of Education, Department of Internal Medicine, West Virginia University School of Medicine, Morgantown, West Virginia, USA; and **Norman D. Ferrari III, MD**, is Professor and Chair, Department of Medical Education, West Virginia University School of Medicine, Morgantown, West Virginia, USA.

Corresponding author: Scott A. Cottrell, EdD, West Virginia University School of Medicine, Morgantown, West Virginia, USA, scottrell@hsc.wvu.edu