The "Good" Dean's Letter

Christopher S. Kiefer, MD, James E. Colletti, MD, M. Fernanda Bellolio, MD, MS,
Erik P. Hess, MD, MSc, Dale P. Woolridge, MD, PhD, Kristen B. Thomas, MD,
and Annie T. Sadosty, MD

Abstract

Purpose
To determine whether a correlation exists between the term "good" on the summative, comparative assessment of a student's Medical Student Performance Evaluation (MSPE) and his or her actual performance in medical school.

Method
The authors reviewed the MSPEs submitted to three residency programs to determine the presence of the term "good" in either the summary paragraph or the appendices. Next, they noted, for institutions using "good," the percentile rankings of those students who received "good" as a descriptor. To examine the consistency among institutions regarding the percentile ranking denoted by "good," they dichotomized the data into students below and above the bottom 25th percentile. They analyzed the data using a nonparametric test because of their nonnormal distribution.

Results
The authors collected MSPEs from 122 of the 125 Liaison Committee on Medical Education–accredited medical schools that were graduating students in 2008. Of these 122 institutions, 34 (28%) used the term "good." All 34 institutions used the term to characterize students in the bottom 50% of the graduating class. The authors found a significant difference in the percentile ranking of students described as "good" between institutions using it to describe the bottom 25% and institutions using the term to describe those in the 25th to 50th percentiles (median ranking of 12.5% versus 30%, $P < .0001$).

Conclusions
Overall, the term "good" in the MSPE describes students in the bottom 50% of the class; therefore, the term "good," as used to describe performance in medical school, consistently indicates below-average performance.

Dr. Kiefer is chief resident, Department of Emergency Medicine, Mayo Clinic, Rochester, Minnesota.

Dr. Colletti is assistant professor and residency program director, Department of Emergency Medicine, Mayo Clinic, Rochester, Minnesota.

Dr. Bellolio is assistant professor, Department of Emergency Medicine, Mayo Clinic, Rochester, Minnesota.

Dr. Hess is assistant professor, Department of Emergency Medicine, Mayo Clinic, Rochester, Minnesota.

Dr. Woolridge is assistant clinical professor of emergency medicine and pediatrics and assistant residency director of the combined emergency medicine/pediatrics residency, University of Arizona, Tucson, Arizona.

Dr. Thomas is assistant professor of radiology and residency program director, Department of Radiology, Mayo Clinic, Rochester, Minnesota.

Dr. Sadosty is assistant professor of emergency medicine and former residency program director, Mayo Clinic, Rochester, Minnesota.

Correspondence should be addressed to Dr. Colletti, 200 First Street SW, Rochester, MN 55905; telephone: (507) 255-5388; fax: (507) 255-6592; e-mail: Colletti.james@mayo.edu.

Each year, residency program directors throughout the United States seek to fill their open positions with the most highly qualified applicants. Program directors have both subjective and objective data to guide them in their decision-making process. The data currently available to compare and contrast potential residents include clinical grades, United States Medical Licensing Examination scores, and the dean's letter or Medical Student Performance Evaluation (MSPE).

The MSPE is designed to present an objective report of the medical student's performance during both the preclinical and clinical years. Historically, the content of these evaluations has been highly variable and selectively laudatory.[1–3] In 1989, the Association of American Medical Colleges (AAMC) proposed guidelines for standardizing the content of these letters, including the insertion of "comparative performance data" to allow program directors to better weigh differences among potential candidates.[1] In addition, the AAMC states explicitly that the MSPE is to be a letter of evaluation—not a letter of recommendation. In 2002, the AAMC updated the guidelines by recommending that the MSPE include six distinct sections: Identifying Information, Unique Characteristics, Academic History, Academic Progress, Summary Statement, and Appendices.[2] The summary statement should contain "a summative assessment of the student's comparative performance during medical school relative to his or her peers." In this section, medical schools often use descriptive adjectives, such as "excellent," "outstanding," or "good," to summarize a student's performance.

Despite the development of these guidelines, previous investigators have noted wide variability in the content and structure of the MSPE as well as a continuing tendency to present students in only a positive light. Shea and colleagues[2] noted that in 2005, three years after the most current AAMC recommendations, only 75% of MSPEs were, considering the guidelines, "adequate." Furthermore, wide variation exists in the students described by the same descriptor at different institutions; for example, researchers recently found that 75% of the schools in their sample used the word "excellent" in MSPEs and that the variance in performance of students whose MSPEs included the word "excellent" ranged from the 3rd to the 97th percentile of the graduating class.[3] This variability can present a challenge for residency directors who desire to use MSPEs to assess the caliber and potential of their applicants.

We undertook this study to determine whether a correlation exists between the

term "good" on the summative assessment of a student's MSPE and his or her actual performance in medical school. We hypothesized that on MSPEs, the term "good" would indicate students with below-average academic performance. To test our hypothesis, we documented the frequency of the term "good" in MSPEs and the percentile ranking of students whose MSPEs included the term.

Method

The Mayo Clinic institutional review board deemed this study exempt.

Study design

The MSPEs submitted by Liaison Committee on Medical Education (LCME)-accredited medical schools through the Electronic Residency Application Service (ERAS) during the application year 2008–2009 were reviewed by two abstractors (a faculty member [E.P.H.] and the chief resident [C.S.K.]) for the presence of the term "good" in the summary statement. If "good" was not present in the summary statement, the abstractors examined the accompanying appendices to see whether a category entitled "good" was present in the explanation of student rankings (the appendices, one of the six sections of the MSPE, contain a graphic representation of a student's performance relative to his or her classmates). Abstractors received one hour of training in the navigation of ERAS and MSPEs. The training emphasized locating the key word "good" in either the summary statement or the appendices of the MSPE.

We randomly selected 26% (32 of 122) of the reviewed MSPEs to assess the interrater reliability of the data abstractors. In these cases, the two trained abstractors extracted data independently and in duplicate, and they resolved disagreements by consensus. In addition, the emergency medicine program director at the Mayo Clinic Rochester site (J.E.C.) monitored abstractor performance by randomly reviewing a representative sample of 30 MSPEs.

To blind the abstractors to institution, we assigned each LCME institution an identifier code. We tracked, using these codes, the institutions that used the word "good" in a Microsoft Excel spreadsheet.

Study setting and population

We undertook this study through the emergency medicine residency at the Mayo Clinic Rochester, the combined emergency medicine/pediatrics residency at the University of Arizona, and the radiology residency at the Mayo Clinic Rochester using the 2008 ERAS application cycle. The study comprised all MSPEs submitted to the above programs. Our criterion for *in*cluding MSPEs in the study was submission from a graduate of an LCME-accredited medical school. Our criteria for *ex*cluding MSPEs were (1) a draft date before the year 2000 and (2) indeterminate or unclear use of the term "good" (i.e., "good" appeared in the letter but was not defined in either the letter or the appendices).

Study protocol

Two abstractors reviewed each MSPE's summary statement for the presence or absence of the term "good," the category "good," or a graphic with an equivalent descriptive category.

Measurements

After the abstractors reviewed MSPEs and entered data, noting the use of the term "good," we calculated the percentage of institutions using the term "good" as well as the percentage of institutions not using "good." In addition, we tabulated the number of institutions using the term to describe students in different quartiles or other rankings.

Data analysis

The data we obtained in the study are categorical and are presented as percent frequency of occurrence. We calculated the Cohen unweighted kappa to assess the interrater reliability of data abstraction. Because the data were not normally distributed, we used the Wilcoxon/Kruskal–Wallis test to test our hypothesis that institutions use the term "good" to indicate significantly different percentile rankings. We used JMP software Version 7.0 (SAS Institute, Chicago, Illinois) for our data analysis.

Results

We reviewed MSPEs from 122 of the 125 LCME-accredited schools graduating students in 2008. Independent data abstractors agreed on 30 of 32 cases (94%). The interobserver agreement for

data abstraction was near-perfect (kappa = 0.85, 95% confidence interval [CI] 0.66–1.00). Overall, 34 institutions (28%) used the term "good" to classify students. Two additional institutions (2%) used the word "good" to describe a group of students, but did not clearly define the term.

Of the 34 institutions using "good" to describe students, all used the term to describe students in the bottom half of the graduating class. Specifically, 25 of the institutions (74%) used the term to classify students in the bottom quartile (0%–25%). Four institutions (12%) applied the term to students graduating in the bottom third of the class (0%–33%). Another four institutions (12%) used the term for students ranked between the 20th and 40th percentiles, and the remaining institution (3%) applied "good" to students graduating in the bottom 50% (0%–50%).

When comparing those schools (no. = 25) using "good" to describe the bottom 25% of the graduating class versus those applying the term to students ranked between the 25th and 50th percentiles (no. = 9), we observed a significant difference in the usage of the term "good": median percentile ranking was 12.5% rather than 30% ($P < .0001$).

Discussion

According to Merriam-Webster, "good" is an adjective used to describe someone who is "competent and skillful."[4] As an MSPE summary descriptor, however, "good" is code for a below-average medical student. Over 70% of schools do not use the summary term "good." Among those that do, however, the percentile rank associated with the term is variable. Some schools use this term to describe students in the bottom quartile; others use it for the bottom half.

Residency directors have the challenging task of selecting their future residents, and the MSPE should present a means for accomplishing this task. AAMC guidelines state that the MSPE is intended to be a tool of evaluation—not a letter of recommendation.[2] Program directors should be able to use the MSPE to differentiate between candidates; however, given the variability of the MSPE from institution to institution, using it to consistently compare the

performance of candidates from different schools is difficult.

The potential utility of the MSPE has been illustrated in previous studies. Lurie and colleagues[5] surveyed program directors and found that students in the top two MSPE ranking groups (first and second quartiles) at a single undergraduate medical institution had significantly higher performance rankings after one year of residency than their lower-ranked counterparts. The authors concluded that the MSPE is a significant predictor of performance in internship.[5] Other studies, though, have failed to demonstrate a relationship between MSPE ranking and resident performance.[6] Perhaps the variable ways in which schools use MSPE descriptors not only contribute to these conflicting findings but also result in program directors' relative lack of reliance on the MSPE in general—and/or the MSPE summary statement in particular—as compared with other pieces of the residency application.[7,8]

Our observation of variability in performance for students assigned the summary statement "good" is in accord with the findings of a previous investigation by Naidich and colleagues.[3] These investigators noted that the range of students described by the term "excellent" was anywhere between the 3rd and the 97th percentiles of the graduating class. The variability in the generally positive descriptive terms "excellent" and "good" applied to students makes it very difficult—if not impossible—for program directors to differentiate between applicants based on MSPE summary assessments. One way to improve the utility of the MSPE would be to standardize across all LCME-accredited institutions the qualifications of groups of students described by a particular term. Standardization would require cooperation and, in some cases, curricular revision (as some schools clearly state that their curriculum is not designed to allow for peer comparison), but it has the potential to increase the precision and accuracy of the summary assessment within and across medical schools. Standardization would also allow program directors to differentiate and compare students from different medical schools.

The call for improvement in the MSPE is not new.[9–11] Researchers and reformers have previously examined the many factors that influence the content of a student's MSPE. First, individual medical schools are naturally motivated to present their students in a positive light. At times, schools suppress negative information to enhance a student's candidacy.[12] Second, the negative portrayal of a student by his or her institution could potentially have a negative impact on the student's ability to match into a residency program and, in turn, could reflect poorly on the medical school itself. Third, the LCME rules state that a student has the right to review his or her MSPE before it is submitted to ERAS for review by prospective programs, and the knowledge that students will read the MSPE may alter its objectiveness.[13] Fourth, some MSPE authors may even "fear legal retribution" from a student whose performance has been described negatively.[12] MSPEs and their connection to the process of applying to residency are complex, which may in part explain why MSPE reform has lagged.

Limitations

This study was primarily limited by its retrospective design. As with all studies that use data collected for another purpose (e.g., health record reviews, administrative database studies), we were unable either to standardize the data collected or to ensure completeness for research purposes. Incomplete and unstandardized data resulted in the absence of the term "good" in two-thirds of the MSPEs we examined, limiting generalizability of our findings to those institutions that routinely use the summative term "good" to describe medical student performance. We do, nonetheless, think our findings will be useful to residency directors and others who are involved in interviewing applicants for residency training, as they will invariably encounter the term "good" on the MSPEs they read.

Strengths

Our study had some strengths as well. We employed an established methodology to limit the intrusion of bias into the data abstraction process.[14–17] We provided, a priori, a clear hypothesis and a description of the database from which the MSPEs were obtained. We also developed a standardized data abstraction form, trained abstractors to reliably identify the term "good" in the MSPE, monitored abstractor performance, blinded abstractors to institution, and assessed the interrater reliability of data abstraction. We believe these methodological safeguards increase the reproducibility and generalizability of our findings.

Conclusions

In our study, the term "good" appeared on MSPEs to describe students in the bottom 50% of the class, thereby making a "good" medical student code for a below-average one. Furthermore, across LCME-accredited institutions, the summary term "good" described a spectrum of students ranging between the 1st and 50th percentiles of their graduating classes.

Despite improvements in standardization of the MSPEs over the last several years, there is still room for improvement. Until then, medical schools may continue to do some students a disservice by failing to differentiate them from their lower- (or higher-) performing peers, and program directors may occasionally be surprised by the performance of a "good" applicant.

References

1 Korn SL, Stagnaro-Green A, Rose S. The dean's letter summary statement: Valuing scholarship, leadership, and academic achievement. Acad Med. 2001;76(10 suppl): S30–S32.

2 Shea JA, O'Grady E, Morrison G, Wagner BR, Morris JB. Medical student performance evaluations in 2005: An improvement over the former dean's letter? Acad Med. 2008;83: 284–291.

3 Naidich JB, Lee JY, Hansen EC, Smith LG. The meaning of excellence. Acad Radiol. 2007;14:1121–1126.

4 "Good." In: Merriam-Webster's Collegiate Dictionary. 10th ed. Merriam-Springfield, Mass: Webster Incorporated; 1993:502.

5 Lurie SJ, Lambert DR, Grady-Weliky TA. Relationship between dean's letter rankings and later evaluations by residency program directors. Teach Learn Med. 2007;19:251–256.

6 Clemente M, Michener WM. The dean's letter of recommendation and internship performance. J Med Educ. 1976;51:590–592.

7 Green M, Jones P, Thomas JX. Selection criteria for residency: Results of a national program directors survey. Acad Med. 2009;84:362–367.

8 Swide C, Lasater K, Dillman D. Perceived predictive value of the Medical Student Performance Evaluation (MSPE) in anesthesiology resident selection. J Clin Anesth. 2009;21:38–43.

9 Wagoner NE, Suriano JR. A new approach to standardizing the dean's letter. Acad Med. 1989;64:688–689.

10 Hunt DD, MacLaren CF, Scott CS, Chu J, Leiden LI. Characteristics of dean's letters in 1981 and 1992. Acad Med. 1993;68:905–911.

11 Lee AG, Golnik KC, Oetting TA, et al. Re-engineering the resident application selection process in ophthalmology: A literature review and recommendations for improvement. Surv Ophthalmol. 2008;53:164–176.

12 Edmond M, Roberson M, Hasan N. The dishonest dean's letter: An analysis of 532 dean's letters from 99 U.S. medical schools. Acad Med. 1999;74:1033–1035.

13 Mallot D. Interview, dean's letter, and affective domain issues. Clin Orthop Relat Res. 2006;449:56–61.

14 Badcock D, Kelly AM, Kerr D, Reade T. The quality of medical record review studies in the international emergency medicine literature. Ann Emerg Med. 2005;45:444–447.

15 Worster A, Bledsoe RD, Cleve P, Fernandes CM, Upadhye S, Eva K. Reassessing the methods of medical record review studies in emergency medicine research. Ann Emerg Med. 2005;45:448–451.

16 Gilbert EH, Lowenstein SR, Koziol-McLain J, Barta DC, Steiner J. Chart reviews in emergency medicine research: Where are the methods? Ann Emerg Med. 1996;27:305–308.

17 Worster A, Haines T. Advanced statistics: Understanding medical record review (MRR) studies. Acad Emerg Med. 2004;11:187–192.