

# Variation and Imprecision of Clerkship Grading in U.S. Medical Schools

Erik K. Alexander, MD, Nora Y. Osman, MD, Jessica L. Walling, and Vivian G. Mitchell, MD

## Abstract

### Purpose

Despite standardized curricula and mandated accreditation, concern exists regarding the variability and imprecision of medical student evaluation. The authors set out to perform a complete review of clerkship evaluation in U.S. medical schools.

### Method

Clerkship evaluation data were obtained from all Association of American Medical Colleges–affiliated medical schools reporting enrollment during 2009–2010. Deidentified reports were analyzed to define the grading system and the percentage of each class within each grading tier. Inter- and intraschool

grading variation was assessed in part by comparing the proportion of students receiving the top grade.

### Results

Data were analyzed from 119 of 123 accredited medical schools. Dramatic variation was detected. Specifically, the authors documented eight different grading systems using 27 unique sets of descriptive terminology. Imprecision of grading was apparent. Institutions frequently used the same wording (e.g., “honors”) to imply different meanings. The percentage of students awarded the top grade in any clerkship exhibited extreme variability (range 2%–93%) from school to school, as

well as from clerkship to clerkship within the same school (range 18%–81%). Ninety-seven percent of all U.S. clerkship students were awarded one of the top three grades regardless of the number of grading tiers. Nationally, less than 1% of students failed any required clerkship.

### Conclusions

There exists great heterogeneity of grading systems and imprecision of grade meaning throughout the U.S. medical education system. Systematic changes seeking to increase consistency, transparency, and reliability of grade meaning are needed to improve the student evaluation process at the national level.

*Editor's Note: A commentary by S. Durning and P. Hemmer will appear with this article in the August issue.*

**T**he training and education of a physician is a complex endeavor that focuses on the development of competency within three principle domains: the mastery of medical knowledge, proficiency in the practice of

**Dr. Alexander** is medicine clerkship director and director of medical student education, Brigham and Women's Hospital, Boston, Massachusetts.

**Dr. Osman** is associate clerkship director, Department of Medicine, Brigham and Women's Hospital, Boston, Massachusetts.

**Ms. Walling** is clerkship coordinator, Department of Medicine, Brigham and Women's Hospital, Boston, Massachusetts.

**Dr. Mitchell** is associate clerkship director, Department of Medicine, Brigham and Women's Hospital, Boston, Massachusetts.

Correspondence should be addressed to Dr. Alexander, Brigham and Women's Hospital, 75 Francis St., PBB-B4, Room 415, Boston, MA 02115; telephone: (617) 732-4148; fax: (617) 264-6346; e-mail: ekalexander@partners.org.

Acad Med. 2012;87:XXX-XXX.

First published online

doi: 10.1097/ACM.0b013e31825d0a2a

Supplemental digital content for this article is available at <http://links.lww.com/ACADMED/A93>.

medical skills, and the ability to doctor within a culturally competent and professional environment.<sup>1-4</sup> Though societies have defined healers within their communities for thousands of years, the central tenets of modern-day medical education have their origins in Abraham Flexner's work during the early 20th century. Flexner's 1910 report on medical education in the United States and Canada argued that variation in curriculum standards and training was unacceptable.<sup>5</sup> Shortly after this report was published, nearly all medical school curricula in the two countries were modified to provide a foundation of core basic science material followed by clinical training in the major medical disciplines. Today, medical education in the United States is primarily structured within a four-year developmental curriculum. Student evaluation is based primarily on individual demonstration of competency in the knowledge, skill, and behavioral domains described above.<sup>6</sup> Mastering these competencies is required for promotion toward and attainment of a medical degree.

Although numerous means of student assessment are well described,<sup>7,8</sup> the

process of incorporating such assessment into a final evaluation (or grade) has proven challenging. This is particularly true within the third-year clinical clerkships, when students are taught by numerous faculty and housestaff physicians in an apprenticeship model.<sup>9-12</sup> The Liaison Committee on Medical Education (LCME) mandates that every accredited U.S. medical school have a system in place for the assessment of medical student achievement which employs a variety of measures to assess students' knowledge, skills, and behaviors.<sup>13</sup> Furthermore, the LCME directs each school to ensure that their faculty understand the uses and limitations of various test formats, the benefits of criterion-referenced versus norm-referenced grading, and the reliability and validity of each modality. Indeed, obtaining LCME accreditation implies that both internal and external reviews of clinical curricula meet these standards.

Despite the progress of the last century, however, concern remains that medical school clerkship evaluations are imprecise, highly variable, and difficult to interpret outside of the

microenvironment of a clerkship itself. Recognizing these issues, the Association of American Medical Colleges (AAMC) mandated modifications to the medical student performance evaluation (MSPE), formerly referred to as the “dean’s letter.”<sup>14</sup> Following an extensive effort from 2000 to 2002, the AAMC defined a novel and standard set of measurable professional attributes expected of all medical students. Furthermore, the AAMC requested that all medical schools provide corresponding policies and consistent procedures for a systematic, performance-based assessment of all students across their core clinical clerkships. This has led to increased transparency of medical schools’ clerkship evaluations.

The transparency of the evaluation process is widely considered a positive step forward. However, the inter- and intraschool variability of clerkship grading<sup>15,16</sup> raises questions about the ultimate meaning of each grade. This is of great importance because an MD degree granted in the United States, as viewed by patients and colleagues alike, implies completion of a medical program that met standardized, consistent, and reliable standards. Furthermore, all LCME-accredited schools should seek to assess similar developmental competencies from their students. To date, no complete review of clerkship grading within U.S. medical schools has been performed. Therefore, with this background, we sought to broadly and systematically review clerkship grading across all LCME-accredited AAMC member schools in the United States. Our goal was to catalogue the types and distributions of grading systems and investigate how these systems affect grade distribution. We believe these data will serve as a launching point for an important national discussion on this topic and for movement toward more reliable and competency-based student assessment.

## Method

We sought clerkship grading data for all LCME-accredited, AAMC-member U.S. medical schools that reported enrollment of students in their required clinical clerkships during academic year 2009–2010. For each medical school (with one exception), we obtained all clerkship grading information by review of data

self-reported by each medical school as part of the MSPE. Data pertaining to Harvard Medical School were obtained via internal communication because they are not provided in the MSPE. See Supplemental Digital Appendix 1 (<http://links.lww.com/ACADMED/A93>) for a list of schools included in the analyses. Three study authors (E.A., V.M., N.O.) were part of the Department of Medicine Internship Selection Committee at the Brigham and Women’s Hospital (BWH) and, therefore, had permission and authority to review MSPE data from the deans’ letters of applicants seeking positions in the BWH medical residency program. At the completion of the residency selection process, all identifying information pertaining to any individual applicant was destroyed. Remaining MSPE data pertaining to each school’s self-reported description of clerkship evaluation were then deidentified (removing all individual and medical school identifiers) and collated for this project. All data were therefore anonymous, and no attempt was made to correlate specific grading systems with specific individuals or schools. Because grouped data were not linked to any identifiable individuals, no human subjects were enrolled, and no intervention occurred, this report was deemed exempt from oversight by the BWH investigational review board, and approval was not required.

For each school, we identified the type of grading system used for clerkships, specifically seeking to determine the grading options within an institution. We then divided medical schools into tiered groups. For example, honors–pass–fail was deemed a three-tier grading system, whereas A–B–C–D–F was deemed a five-tier grading system. Thereafter, we determined the percentage of each medical school’s class placed into each grading tier as part of a full year’s academic enrollment. Many schools provided exact numerical percentage data. Other schools provided data in graphic form. When this occurred, two of the authors independently deciphered specific numerical percentage data from the graph and reached consensus. For almost all schools, we were able to obtain accurate data for every grading tier within each required clerkship. A small minority of schools provided data only pertaining to their top grading tier (e.g.,

the percentage receiving honors). In these cases, we included these data only within that category and left remaining grading categories as “unknown.”

Nearly all medical schools provided required clerkship evaluation data for their internal medicine, surgery, family practice (primary care), pediatrics, obstetrics–gynecology, and psychiatry clerkships. A minority of schools also provided data for neurology and anesthesia clerkships. All data provided by the schools were for clerkships required as part of the schools’ MD degree-granting requirements and usually occurred in the third year of medical school. We collected no data for elective clerkship experiences.

Because grade inflation and grade interpretation of the top grading tier has garnered particular attention nationwide,<sup>17–19</sup> we assessed intra- and interschool grading variation in part by comparing percentages of students placed into this top category. We assessed intraschool variation by comparing the proportion of students receiving the top evaluation category within different clerkships from the same medical school, whereas we examined interschool variation by comparing the same clerkships among different medical schools.

Categorical variables were statistically compared using the chi-square test, whereas linear variables were compared with the Student’s *t* test. A *P* value <.05 was considered significant.

## Results

We accurately obtained and analyzed clerkship grading data from 119 of the 123 (97%) AAMC-member, LCME-accredited U.S. medical schools that reported enrollment of students in their required clinical clerkships during the study period (see Supplemental Digital Appendix 1, <http://links.lww.com/ACADMED/A93>). Eight new medical schools had yet to enroll medical students in their third-year clinical rotations and were, thus, excluded from this analysis. Of the 119 schools (46 private institutions, 73 public institutions) for which data were available, all provided information clarifying their type of grading system and the terminology they used. One

hundred ten of the 119 (92%) schools also provided additional data defining the proportion of students receiving each specific grade within their grading system. Three of the remaining schools listed only the proportion of students receiving the top possible grade (such as honors), whereas six others used pass–fail grading systems, though none of the six pass–fail schools revealed the percentage of failing students.\*

The clerkship grading systems used among U.S. medical schools varied greatly with respect to number of grading tiers. We identified eight different grading systems: Six schools used a 2-tier grading system (defined as pass–fail), 16 used a 3-tier system, 63 used a 4-tier system, 27 used a 5-tier system, 4 used a 6-tier system, and 1 school each employed a 7-tier, 9-tier, and 11-tier system (Table 1).

We also observed significant variability with respect to grading terminology. Ninety-seven of the 119 schools (82%) used descriptive terminology to define their grading categories. This often included labels such as “honors,” “satisfactory,” “marginal,” or “unsatisfactory.” There was, however, no clear semantic consistency among schools; the same word had a different meaning at different schools. For example, “honors” was used to define the top grade at some schools, but the same term indicated the second-best grade at other schools. Similarly, “satisfactory” was often used to denote a relatively high grade in a 3-tier or 4-tier system, but it described a relatively low grade in some systems with more grading tiers.

In contrast, 22 of 119 schools (17%) used letter grades between A and F to define their grading categories. Among these, variation between schools was once again profound. Some schools used all letter options A through F, whereas others added “plus” and “minus” to further delineate options. Several others employed the “plus” terminology, though not the “minus.” Finally, one school created combination letter grades, such as AB and BC. A summary of all grading systems and grading terminology currently used in U.S. medical schools

\* Because student performance data were reported only in percentages, we cannot provide raw numerical data for our results related to student performance.

Table 1

**Grading Systems and Grading Terminology at 119 Liaison Committee on Medical Education–Accredited, Association of American Medical Colleges–Affiliated U.S. Medical Schools Reporting Data for Required Clinical Clerkships, 2009–2010**

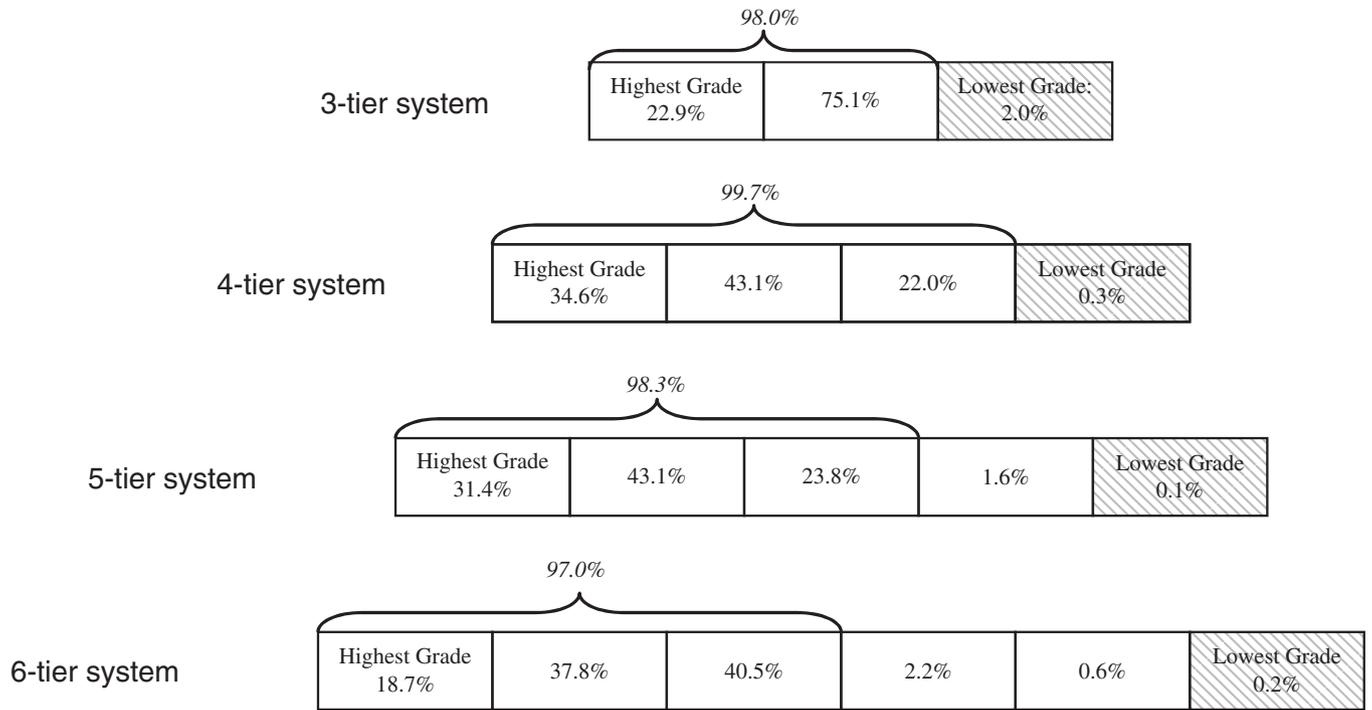
Grading system	No. of schools	Terminology
2-tier	6	<ul style="list-style-type: none"> <li>• Pass/Fail</li> </ul>
3-tier	15	<ul style="list-style-type: none"> <li>• Honors/Pass/Fail</li> <li>• Honors/Satisfactory/Remediation</li> </ul>
4-tier	63	<ul style="list-style-type: none"> <li>• A/B/C/F</li> <li>• Honors/High Pass/Pass/Fail</li> <li>• High Honors/Honors/Pass/Fail</li> <li>• Honors/Near Honors/Pass/Fail</li> <li>• High Honors/Excellent/Good/Fail</li> <li>• Honors/Pass/Conditional Pass/Fail</li> <li>• Honors/Commendable/Satisfactory/Fail</li> <li>• Honors/Satisfactory/Low Satisfactory/Fail</li> <li>• Outstanding/Advanced/Proficient/Unsatisfactory</li> <li>• Honors/Satisfactory Plus/Satisfactory/Unsatisfactory</li> <li>• Honors/Letter of Commendation/Satisfactory/Unsatisfactory</li> </ul>
5-tier	27	<ul style="list-style-type: none"> <li>• A/B/C/D/F</li> <li>• A/B+/B/C/F</li> <li>• Honors/High Pass/Pass/Marginal/Fail</li> <li>• Honors/High Pass/Satisfactory Pass/Marginal Pass/Fail</li> <li>• Outstanding/Above Expected/Expected/Below Expected/Fail</li> <li>• Honors/High Satisfactory/Satisfactory/Low Satisfactory/Unsatisfactory</li> </ul>
6-tier	5	<ul style="list-style-type: none"> <li>• A/B+/C/C+/C/F</li> <li>• A/AB/B/BC/C/F</li> <li>• Excellent With Honors/Excellent/Good/Marginal/Pass/Fail</li> </ul>
7-tier	1	<ul style="list-style-type: none"> <li>• A/B+/B/C+/C/D/F</li> </ul>
9-tier	1	<ul style="list-style-type: none"> <li>• Honors/A/A–/B+/B/B–/C+/C/C–</li> </ul>
11-tier	1	<ul style="list-style-type: none"> <li>• A/A–/B+/B/B–/C+/C/C–/D+/D/D–/F</li> </ul>

is depicted in Table 1. There was no statistical difference in the types of grading systems used in the 46 private versus the 73 publicly funded medical schools.

On average, less than 1% of U.S. medical students failed a required clinical clerkship in internal medicine, surgery, pediatrics, obstetrics–gynecology, psychiatry, or family practice during the reported academic year. The percentage of required clerkships reporting zero failures during the academic year increased with the number of grading tiers. Specifically, only 29 of 83 (35%) required clerkships using a three-tier system reported zero failures, whereas 251 of 375 (67%) using a four-tier system and 138 of 160 (86%) using a five-tier system reported zero failures ( $P < .01$ ). Ninety-seven percent of all U.S. clerkship students were awarded one of the top three grades, regardless of the number of grading tiers (Figure 1).

We next analyzed the proportion of medical students who had received the top grade for their performance in required clerkships. Grading systems with four or more grading tiers were associated with a higher proportion of students receiving the top clerkship grade as compared with systems with three grading options (mean: 33% received top grade in four-tier systems versus 23% received top grade in three-tier systems;  $P < .01$ ). Grading systems with four or more tiers also demonstrated larger school-to-school variation in the proportion of students receiving the top grade (range: 2%–93%) as compared with systems with three grading options (range: 5%–51%). Detailed grading analysis is shown in Table 2 and Figure 2.

Finally, we analyzed intraschool clerkship variation by comparing the proportion of medical students receiving the top grade among required clerkships within the same medical school. Similar to



**Figure 1** Percentage of medical students receiving each grade within various grading systems. The figure reflects data from 110 Liaison Committee on Medical Education–accredited, Association of American Medical Colleges–affiliated U.S. medical schools that reported the proportion of students receiving each specific grade within their grading system, 2009–2010.

above, even within the same medical school, we detected extensive variability in the proportion of students awarded the top grade. After excluding schools with pass–fail grading systems, 30 of 113 (27%) medical schools had two or more required clerkships in which the

percentage of students receiving the top grade in the same academic year differed by  $\geq 30\%$  (Table 3). One comparison is exemplary: At one school, 82% of students were awarded the top grade in the neurology clerkship, whereas only 20% were awarded such in the psychiatry

clerkship. Another school conversely awarded 72% of students the top grade in the psychiatry clerkship, though only 25% received the top grade in the neurology clerkship.

**Discussion**

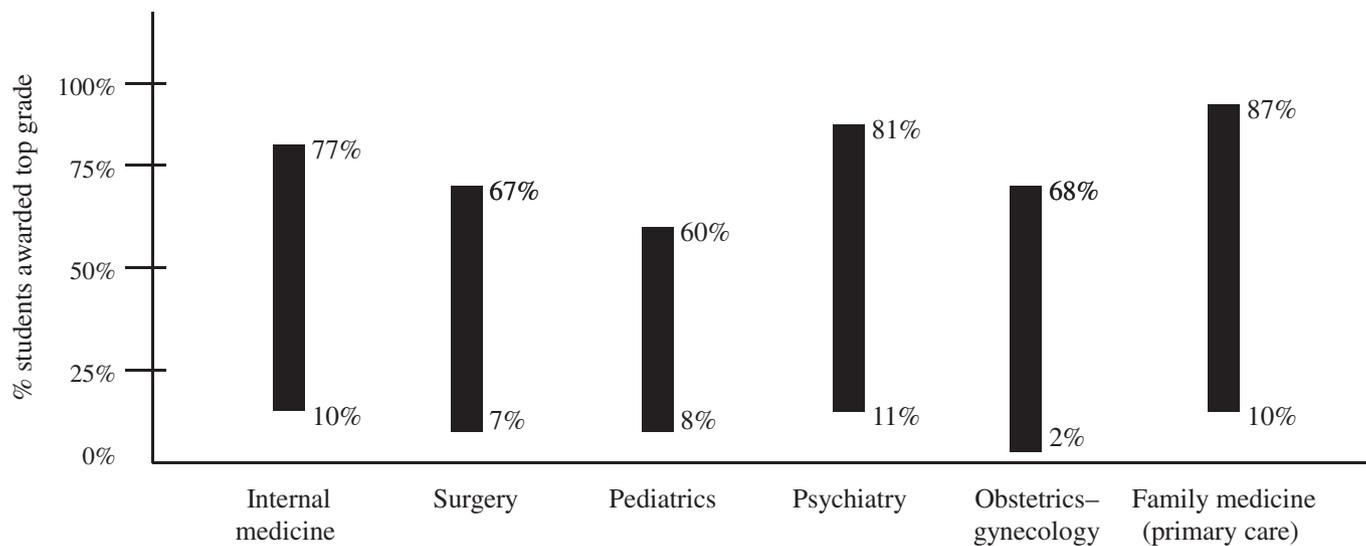
For the first time, these data provide a broad and unbiased assessment of medical student clerkship grading in the United States. Although LCME accreditation requires similarity and standardization of core clerkship curricula, we found that there are no standardized requirements for the evaluation and grading of medical students within these clerkships. Indeed, we found that both the *language* and the *meaning* of evaluation and grading vary dramatically among U.S. medical schools. We identified eight different grading systems with 27 unique sets of descriptive terminology. Furthermore, institutions frequently use the same terminology (such as “honors” or “satisfactory”) to imply different meaning. The percentage of medical students awarded the top grade in any required clerkship similarly exhibits great variation from school to school and is even highly variable when comparing clerkships within a single institution. Together, these data suggest

**Table 2**  
**Distribution and Variation in the Percentage of Students Awarded the Top Clerkship Grade at 105 Liaison Committee on Medical Education–Accredited, Association of American Medical Colleges–Affiliated U.S. Medical Schools Using 3-Tier, 4-Tier, and 5-Tier Grading Systems, 2009–2010\***

Clerkship	% of class awarded top grade		
	3-tier grading system	4-tier grading system	5-tier grading system
<b>Internal medicine</b>	21.5	30.5	27.9
<b>Surgery</b>	21.1	31.4	33.4
<b>Pediatrics</b>	24.0	31.6	31.0
<b>Obstetrics–gynecology</b>	26.8	32.8	31.9
<b>Psychiatry</b>	25.2	39.3	36.4
<b>Family medicine</b>	20.6	39.1	34.5
<b>All clerkships</b>			
Mean	23.2 <sup>†</sup>	33.9 <sup>†</sup>	32.5 <sup>†</sup>
Median	22.0	31.0	31.0
Range	46 (5–51)	82 (2–84)	86 (7–93)

\*Two-tier schools were excluded because no failures were reported at these schools. Data from four 6-tier schools, and one each of 7-tier, 9-tier, and 11-tier schools, are not shown because of exceptionally small numbers which prevent effective comparison. Raw numerical data were not reported for student performance; therefore, we can report only percentages.

<sup>†</sup>  $P < .01$  for difference between groups.



**Figure 2** National variation in the percentage of medical students awarded the top grade in each required clerkship (range: high to low). The figure reflects data from 110 Liaison Committee on Medical Education–accredited, Association of American Medical Colleges–affiliated U.S. medical schools that reported the proportion of students receiving each specific grade within their grading system, 2009–2010.

the need to rethink and reform the medical student evaluation process across U.S. medical schools. They also suggest that a standardized approach to grading that employs a common evaluative structure may hold the greatest promise for optimally achieving this end point.

The LCME has addressed the importance of clear and consistent educational objectives as well as fair and transparent evaluation.<sup>13</sup> The LCME outlines 47 educational directives, which clerkship directors are expected to understand and apply to their courses. Educational Directive 1 (ED1-A)<sup>13</sup> addresses the need for core content and consistent course objectives defined in part by public expectations of a physician's competency. Educational Directive 30 (ED 30)<sup>13</sup> states that clerkship directors must also design and implement a system of fair and timely evaluation for each course. What is not addressed, however, is the expectation of consistency in evaluation and grading terminology. Even among U.S. medical schools using the same grading systems, we document impressive differences in the proportion of medical students awarded the top grade. Among LCME-accredited U.S. medical schools, we found that this percentage ranges from 2% to 92%.

Despite all of these inconsistencies, less than 1% of all U.S. medical students fail a required clerkship, regardless of the grading system used. Our data suggest that increasing the available grading options within a school does not lead to

a more rational distribution of grades. Rather, it contributes to grade inflation, with most students grouped in the highest grading groups, and vanishingly small numbers of students awarded the lowest grades (Figure 1). Ninety-seven percent of all U.S. medical students were awarded one of the top three grades regardless of the grading system used.

Concern regarding medical student clerkship evaluation has been raised previously. In a pilot study of 15 Canadian medical schools, McLeod and colleagues<sup>15</sup> also witnessed inconsistent grading terminology and grade distribution. Similarly, Magarian and Mazur<sup>20</sup> reported parallel findings in U.S. medical schools, though in-depth and comparative analysis was limited. More recently, Hemmer and colleagues<sup>21</sup> performed a nonvalidated survey of internal medicine clerkships seeking to define evaluation methods. They noted increased use of descriptive evaluation among clerkship directors, with many schools using the RIME (Reporter, Interpreter, Manager, Educator)<sup>22</sup> framework. Such evaluation metrics have indeed proven increasingly useful and popular among clerkship directors, though unfortunately our data suggest that such tools have failed to reduce grading variation and imprecision. Regardless, it is clear that the desire for a universal, standardized, valid, and reliable clerkship grading system has been present

for decades despite limited meaningful progress toward that end point.

The downstream influence of variable and imprecise grading is notable. In 2000, an advisory committee was formed by the AAMC to address the decline in the quality of the deans' letters, largely attributable to the imprecision and variation in student evaluation.<sup>14</sup> In 2002, the advisory committee's recommendations were adopted by the AAMC Executive Committee and implemented. Although this process significantly improved the consistency of content within the newly termed MSPE, it fell short of creating consistency in the language and meaning of grades.

This lingering inconsistency deserves emphasis. The MSPE often contains descriptive comments which allow for a more nuanced understanding of the student's performance than might otherwise be garnered from the grade found on the transcript. These paragraphs, however, are often drafted to highlight the student's strengths and force the reader to infer meaning from select code words or pertinent negatives. Furthermore, such narrative comments are unavailable to the general public or licensure boards, who must rely on transcript grades to convey assessment of competency. The transcript, with its transparent list of courses and grades, is therefore assumed to be a clear and reliable description of the student's

Table 3

**Detailed Analysis of 30 Liaison Committee on Medical Education–Accredited, Association of American Medical Colleges–Affiliated U.S. Medical Schools Exhibiting  $\geq 30\%$  Variation in the Proportion of Medical Students Receiving the Top Grade in Two Required Clerkships, 2009–2010**

School	Highest % of students receiving top grade		Lowest % of students receiving top grade		% Intraschool variation (range)
	Clerkship	% Students	Clerkship	% Students	
School 1	Family medicine	81	Surgery	18	63
School 2	Neurology	82	Psychiatry	20	62
School 3	Family medicine	78	Obstetrics–gynecology	18	60
School 4	Family medicine	84	Pediatrics	25	59
School 5	Psychiatry	72	Neurology	25	47
School 6	Family medicine	84	Surgery	38	46
School 7	Neurology	67	Family medicine	22	45
School 8	Neurology	65	Family medicine	20	45
School 9	Neurology	93	Pediatrics	49	44
School 10	Medicine	72	Pediatrics	28	44
School 11	Family medicine	81	Psychiatry	41	40
School 12	Family medicine	57	Obstetrics–gynecology	17	40
School 13	Psychiatry	65	Medicine	26	39
School 14	Psychiatry	67	Pediatrics	28	39
School 15	Obstetrics–gynecology	51	Surgery	12	39
School 16	Psychiatry	56	Neurology	18	38
School 17	Surgery	70	Medicine	32	38
School 18	Obstetrics–gynecology	49	Surgery	12	37
School 19	Obstetrics–gynecology	48	Family medicine	12	36
School 20	Neurology	58	Obstetrics–gynecology	22	36
School 21	Neurology	56	Family medicine	22	34
School 22	Surgery	45	Medicine	12	33
School 23	Psychiatry	67	Medicine	34	33
School 24	Surgery	57	Pediatrics	25	32
School 25	Obstetrics–gynecology	52	Medicine	21	31
School 26	Obstetrics–gynecology	48	Psychiatry	17	31
School 27	Psychiatry	47	Family medicine	16	31
School 28	Psychiatry	49	Surgery	19	30
School 29	Psychiatry	47	Family medicine	17	30
School 30	Medicine	52	Surgery	22	30

clerkship performance. Our analysis suggests that this assumption is not true. There exists little consistency among medical schools: Grades such as “honors” or “satisfactory” at one school convey meaning which is very different from that same grade at another school. It should be stressed that transparency regarding the level of students’ achieved competency carries even greater importance within a system where less than 1% of medical students fail medical school after acceptance.

Beyond raising awareness of this problem, our data provide important findings which may assist in formulating

a solution. Most important, consistent and transparent grading terminology should be established on a national level. LCME accreditation standards should recommend use of a single and consistent grading system throughout all U.S. medical schools, much as content exposure to internal medicine, surgery, pediatrics, obstetrics–gynecology, and psychiatry are mandated. Those tasked with creating and implementing any standardized grading system should reflect that the above analysis suggests a detriment in creating excessive grading categories. Our analysis suggests that grading systems with fewer available grade options demonstrate less grade

inflation while simultaneously reducing interschool variation.

Although a true pass–fail grading system may best mirror a true competency-based curriculum, we also acknowledge that a single pass–fail evaluation provided for an entire clerkship limits comparative analysis of individual students. However, we note that several medical schools define their grading systems as pass–fail but also use unique attributes or secondary evaluative scales to further describe student performance. One school evaluates competency as pass or fail, though it also lists overall course performance based on a numerical score

of 0 to 100. Separate schools provide “letters of distinction” or “special acknowledgement” in lieu of raw scores. In doing so, these systems effectively provide a three-tiered evaluation system regardless of their external label.

Though several reasonable solutions exist to the problem described above, we believe that all clerkship evaluations should be competency based and should convey a student’s “good and adequate” performance when such competencies are achieved. Outstanding student performance above this level would qualify for an honors evaluation. Similar to other LCME educational directives, a reasonable standard could also be set clarifying the expected proportion of students who should receive the honors grade annually. Finally, these proposals by no means exclude the importance of descriptive evaluation in conjunction with any summative grade.

We acknowledge limitations to our study. First, intraschool grade variation may simply reflect a superiority to the teaching, curriculum, and/or learning environment in one clerkship over another while, nonetheless, accurately depicting a student’s mastery (or lack) of competency. Second, these data are self-reported, and thus the accuracy of the data must be assumed. Furthermore, they represent data from only a single academic year. However, given the accreditation oversight of these schools, it seems highly likely that the findings are accurate and depict a reasonable overview of the current U.S. medical school system.

In summary, our data confirm the extreme heterogeneity of evaluative systems and imprecision of grade meaning which presently exist throughout the U.S. medical education system. Systematic changes which seek to increase consistency of terminology, transparency of grade distribution, and reliability of grade meaning are needed to

improve the student evaluation process on a national scale. This, in turn, will lead to better assessment of individual student performance which ultimately will benefit all involved. Arguably, this is of most importance for the students themselves.

*Funding/Support:* None.

*Other disclosures:* None.

*Ethical approval:* Because grouped data were not linked to any identifiable individuals, no human subjects were enrolled, and no intervention occurred, this report was deemed exempt from oversight by the Brigham and Women’s Hospital investigational review board, and approval was not required.

*Previous presentations:* These data were presented as part of the 2012 Harvard Medical School Education Day.

## References

- Albanese MA, Mejicano G, Mullan P, Kokotailo P, Gruppen L. Defining characteristics of educational competencies. *Acad Med.* 2008;42:248–255.
- Miller GE. The assessment of clinical skills/competence/performance. *Acad Med.* 1990;65(9 suppl):S63–S67.
- Duffy FD, Gordon GH, Whelan G, et al. Participants in the American Academy on Physician and Patient’s Conference on Education and Evaluation of Competence in Communication and Interpersonal Skills. Assessing competence in communication and interpersonal skills: The Kalamazoo II report. *Acad Med.* 2004;79:495–507.
- Goldstein EA, Maclaren CF, Smith S, et al. Promoting fundamental clinical skills: A competency-based college approach at the University of Washington. *Acad Med.* 2005;80:423–433.
- Flexner A. Medical Education in the United States and Canada: A Report to the Carnegie Foundation for the Advancement of Teaching, Bulletin No. 4. Boston, Mass: Updike; 1910.
- Pangaro L. A primer of evaluation: Definition and important distinctions in evaluation. In: Fincher RM, ed. *Guidebook for Clerkship Directors*. 3rd ed. Omaha, Ne: Alliance for Clinical Education; 2005:142–150.
- Pangaro L. A new vocabulary and other innovations for improving descriptive in-training evaluations. *Acad Med.* 1999;74:1203–1207.
- Kogan JR, Hauer KE. Brief report: Use of the mini-clinical evaluation exercise in internal medicine core clerkships. *J Gen Intern Med.* 2006;21:501–502.
- Zahn CM, Nalesnik SW, Armstrong AY, Satin AJ, Haffner WH. Variation in medical student grading criteria: A survey of clerkships in obstetrics and gynecology. *Am J Obstet Gynecol.* 2004;190:1388–1393.
- Kumar A, Gera R, Shah G, Godambe S, Kallen DJ. Student evaluation practices in pediatric clerkships: A survey of the medical schools in the United States and Canada. *Clin Pediatr (Phila).* 2004;43:729–735.
- Ravelli C, Wolfson P. What is the “ideal” grading system for the junior surgery clerkship? *Am J Surg.* 1999;177:140–144.
- Lurie SJ, Mooney CJ. Relationship between clinical assessment and examination scores in determining clerkship grade. *Med Educ.* 2010;44:177–183.
- Liaison Committee on Medical Education. *Functions and Structure of a Medical School: Standards for Accreditation of Medical Education Programs Leading to the M.D. Degree*. Washington, DC: Liaison Committee on Medical Education; 2010. <http://www.lcme.org/functions2010jun.pdf>. Accessed April 12, 2012.
- Tally RC, Nolan RJ, Brett-Major D, et al. *A Guide to the Preparation of the Medical Student Performance Evaluation*. Washington, DC: Association of American Medical Colleges; 2002. <https://www.aamc.org/download/139542/data/mspe.pdf>. Accessed April 12, 2012.
- McLeod PJ. So few medical schools, so many clerk rating systems! *CMAJ.* 1992;146:2161–2164.
- Plymale MA, French J, Donnelly MB, Iocono J, Pulito AR. Variation in faculty evaluations of clerkship students attributable to surgical service. *J Surg Educ.* 2010;67:179–183.
- Speer AJ, Solomon DJ, Fincher RM. Grade inflation in internal medicine clerkships: Results of a national survey. *Teach Learn Med.* 2000;12:112–116.
- Roman BJ, Trevino J. An approach to address grade inflation in a psychiatry clerkship. *Acad Psychiatry.* 2006;30:110–115.
- Cacamese SM, Elnicki M, Speer AJ. Grade inflation and the internal medicine subinternship: A national survey of clerkship directors. *Teach Learn Med.* 2007;19:343–346.
- Magarian GJ, Mazur DJ. A national survey of grading systems used in medicine clerkships. *Acad Med.* 1990;65:636–639.
- Hemmer PA, Papp KK, Mechaber AJ, Durning SJ. Evaluation, grading, and use of the RIME vocabulary on internal medicine clerkships: Results of a national survey and comparison to other clinical clerkships. *Teach Learn Med.* 2008;20:118–126.
- Pangaro LN. Investing in descriptive evaluation: A vision for the future of assessment. *Med Teach.* 2000;22:478–481.