The American Journal of Surgery 217 (2019) 329–334

Contents lists available at ScienceDirect

The American Journal of Surgery

journal homepage: www.americanjournalofsurgery.com

Variation of surgery clerkship grades in US medical schools

Michelle R. Brownstein [a, *], Mary R. Shen [b], Paula D. Strassle [a], Michael O. Meyers [a]

[a] Department of Surgery, University of North Carolina at Chapel Hill, 4008 Burnett-Womack Building, CB #7228, Chapel Hill, NC, 27599-7050, USA
[b] University of North Carolina School of Medicine, 321 S Columbia St, Chapel Hill, NC, 27516, USA

ARTICLE INFO

Article history:
Received 15 May 2018
Received in revised form
6 September 2018
Accepted 17 September 2018

Keywords:
Medical student performance evaluation
Surgery clerkship
Student evaluation
Grade variability
Medical student
Surgical education

ABSTRACT

Background: Despite efforts at standardization, evaluation and reporting of clerkships remains highly variable. This study reviews the current spectrum of surgical clerkship grading.
Methods: Data were reviewed for every medical school from which an application was received to a single surgery residency program in 2017 and were evaluated for core surgical clerkship grading systems, distributions, and components. Fischer's exact tests and Wilcoxon-Mann-Whitney tests were used for analysis.
Results: 133 (49 private) schools were evaluated. Geographic distribution:34 Northeast, 50 South, 31 Midwest and 18 West. 120 reported grading tiers, with public schools (95%) more likely than private (80%) to report this ($p = 0.02$). The number of grading categories ranged from 2 to 11; 90% with 3–5. Over 25% of the schools gave ≥40% of students the highest grade; median of 30% in the highest tier.
Conclusions: Significant variation exists in core surgery clerkship grading between schools. Similarly, a sizeable difference exists in how grades are calculated and the reporting systems used. Standardizing grading schemes across medical schools would be beneficial.

Introduction

Variability of clerkship grades between medical schools has long been recognized. Not only does this make it difficult to compare progress and achievement of students between schools, but this creates challenges for training programs when reviewing applications for residency. Issues that have been raised include grade inflation, the heterogeneity of the grading system used and the components that go into those grades.[1–3] Clerkship grades currently include measures of the student's competencies across such subjective domains as professionalism, interpersonal interactions, work ethic and critical thinking that are not assessed in the quantitative components such as the National Board of Medical Examiners (NBME) subject tests (a.k.a. "shelf" exams). These nonobjective components have contributed to grade inflation that has persisted.[3]

The Association for American Medical Colleges (AAMC) has been actively involved in developing and updating guidelines for the summative evaluations that medical schools send out for applicants since 1989. In 2002, the Medical Student Performance Evaluation (MSPE) guidelines were updated to improve transparency between all AAMC member medical schools such that the individual school's comparative data on clerkship grading schema and distributions were to be included. While this was a step forward, the variability in clerkship grades was now available to be compared across medical schools. Using applications from the 2009–2010 academic cycle, Alexander et al. evaluated MSPEs from 119 schools and reported dramatic variation in grading systems with a non-normal grade distribution skewed high and a sizeable portion of students receiving the highest grades across all core clerkships (Alexander, 2014).[2] These findings were also found in the internal medicine clerkship grades across the country when Hom et al. assessed 117 schools and focused on the IM MSPE reported grades and grading schema. They also found approximately half of the schools were not following the then MSPE guidelines for providing information about the grading schema or distribution of clerkship grades.[4] The additional impact of the different components of the grades and the impact the various weights of each component have on final grades have also been identified as contributing to grade inflation.[1]

In September 2016, the AAMC MSPE Task Force released a document titled Recommendations for Revising the MSPE. There were several principals the task force identified to guide what the

* Corresponding author. Department of Surgery, Division of Acute Care Surgery, Brody School of Medicine, 600 Moye Boulevard, East Carolina University, Greenville, NC, 27834-4354, USA.
E-mail addresses: brownsteinm18@ecu.edu (M.R. Brownstein), pstrass@live.unc.edu (P.D. Strassle), mmeyers@med.unc.edu (M.O. Meyers).

https://doi.org/10.1016/j.amjsurg.2018.09.024
0002-9610/

MSPE would provide. These include supplementing (not reiterating) information already provided in the application, providing a level of transparency and standardization to facilitate residency selection, providing comparative applicant information and applicants' standing on competencies required as a successful resident, increasing the opportunity for program directors to examine applicants holistically and provide qualitative and quantitative assessment of applicants that are easy to read. Among other specific recommendations, one was to include information on how final grades and comparative data are derived. For complete transparency, the components of each clerkship grade and the relative weights were to be specified.[5] There has since been some discussion as to the different points of emphasis in these guidelines, highlighting defining the components will be important to assess concordance between grades and the narrative provided.[6]

We sought to review the spectrum of core surgical clerkship grades reported for students in the 2016-17 academic year. The aim of this study was to evaluate the distribution of core surgery clerkship grades and examine Medical Student Performance Evaluation (MSPE) compliance with reporting grade components with respect to the Association of American Medical Colleges (AAMC) guidelines.

Methods

Data were reviewed for every medical school from which an application was received to a single surgery residency program in 2017. The primary source of data was the MSPE, where information on each school's core surgical clerkship grading criteria and composition were extracted. This included the number of grading tiers, components used to calculate grades, and the percentage of students assigned to each grade tier when available in addition to demographic information about each school. We did not use individual student grades and no student-specific information was collected. Only LCME-accredited, AAMC-member U.S. medical schools, excluding Puerto Rico, were included. Individual schools were categorized by United States Census region, and private versus public status using freely available data provided by the AAMC including schools using the Texas Medical and Dental School Application.[7,8]

Number of grading tiers was recorded as the number of possible grading options available for a student at each school; for example, honors/high pass/pass/fail would be considered four tiers. Additionally, the terminology used at different schools within each grading system was recorded (e.g., for a 3-tier grading system, honors/pass/fail and A/B/C) to evaluate for any consistency in nomenclature between schools. Grading schema was generally listed in the MSPE or on the medical school website. School-specific reported comparative student performance within the MSPE was used to calculate the percentage of students in each grading tier.

Univariate analyses were used to describe the prevalence of grading tiers and distribution of grading component weights. Fisher's exact tests were used to assess whether reporting varied across region or private/public school status. Wilcoxon-Mann-Whitney tests were used to assess differences in grade distributions. Schools with 2-tier systems were excluded from these analyses.

Results

Core surgical clerkship grading data from 133 of 147 (90%) AAMC-member, LCME-accredited U.S. medical schools were obtained. This included 34/34 (100%) of medical schools in the Northeast, 50/53 (94%) of schools in the South, 31/35 (89%) of schools in the Midwest, and 18/24 (75%) of schools in the West. Of the schools included, 49 (37%) were classified as private medical schools and 84 (63%) were public.

Grading tiers

The vast majority of medical schools (120/133, 90%) reported their number of grading tiers. While no regional variation in reporting was seen ($p = 0.72$), public schools were significantly more likely to report their grading tiers compared to private schools (95% vs. 82%, $p = 0.02$).

The number of tiers used to report grades ranged from two to eleven tiers, with the majority of institutions reporting 3-, 4-, or 5-tiers (32%, 33%, and 25% of reporting schools, respectively) (Table 1). Additionally, even within systems with the same number of tiers, substantial variation in grading terminology exists. And while almost half of schools ($n = 61$, 48%) included a failing grade in their terminology, a few schools opted for euphemisms like “Unsatisfactory” ($n = 9$), “Noncredit” ($n = 1$), or “No Pass” ($n = 3$). 52 schools (41%) did not include any type of failing grade in their terminology. No regional ($p = 0.57$) or public/private ($p = 0.35$) differences were seen between schools who did not include a failing grade.

Grade distribution

Among schools with ≥3 tiers, the median number of students falling into the highest grade was 30% (interquartile range [IQR] 21–40). Over 25% of schools gave ≥40% of their students the highest grade (maximum 73%). Roughly the same proportion of students was given the highest grade among 3-, 4-, and 5- tiered systems (35%, 31%, and 31%, respectively) (Fig. 1). And despite having more tiers, schools with 4-, 5-, and 6- tiers had <5% of students beyond the third lowest grade. Among schools that included a failing grade ($n = 67$), 51 schools (76%) failed 0% of students, 9 schools (13%) failed 1% of students, 3 schools (4%) failed 2% of students, and 4 schools (6%) failed 3% of students. No significant differences existed in the median proportion of students given the top grade between private and public schools (27% vs. 33%, $p = 0.17$) or across regions (Northeast 34%, Midwest 28%, South 29%, and West 25%, $p = 0.19$) (Fig. 2).

Clerkship grade components

Only 53% of schools (71/133) reported the grade components that contributed to a student's overall surgery clerkship grade. There were no regional differences in reporting ($p = 0.55$) or between private/public schools ($p = 0.99$).

The two most common grade components were clinical evaluation (median 50%, IQR 60%–75%) and the NBME shelf exam (median 30%, IQR 25%–33%), Fig. 3. Most of the schools that reported grade components did not use the objective structured clinical exam (OSCE) to determine surgery clerkship grades (median 0%, IQR 0%–10%). Other assignments generally included attendance, professionalism, written assignments, intra-departmental exams, or presentations. This category represented anywhere from 0% to 65% of the final grade (median 15%, IQR 5%–25%) (Fig. 3).

Discussion

Surgery clerkship grades have long been considered one of the more important aspects of medical student evaluation. Traditionally they have been relied upon heavily for evaluating applicants to training programs and some have reported clerkship grades as an important determinant in which students choose to matriculate into surgery residency programs.[9,10] The MSPE is intended to give

Table 1
Clerkship grading systems among accredited medical schools in the United States (n = 130).

Number of Grading Tiers	N (%)	Terminology[a]
2	5 (4)	Honors, Pass (2)
		Honors, Satisfactory (2)
		Pass, Fail
3	42 (32)	A, B, C
		High Honors, Excellent, Good
		Honors, Credit, Noncredit
		Honors, High Pass, Fail
		Honors, High Pass, Pass (15)
		Honors, High Satisfactory, Satisfactory (2)
		Honors, Letter of Commendation, Satisfactory
		Honors, Near Honors, Pass
		Honors, Pass, Fail (11)
		Honors, Pass, No Pass
		Honors, Satisfactory with Commendation, Satisfactory
		Honors, Satisfactory, Unsatisfactory
		Letter of Distinction, Pass, Fail
		Superior, High Satisfactory, Satisfactory
4	43 (33)	A, B+, B, C
		A, B, C, F (3)
		Honors with Distinction, Honors, Pass, Unsatisfactory
		Honors, 93–90, 89–85, <85
		Honors, Commendable, Satisfactory, Unsatisfactory (2)
		Honors, Excellent, Satisfactory, No Pass
		Honors, High Pass, Pass, Conditional
		Honors, High Pass, Pass, Fail (21)
		Honors, High Pass, Pass, Marginal Pass
		Honors, High Pass, Pass, Pass with Remediation
		Honors, Near Honors, Pass, Fail
		Honors, Pass, Near Pass, Fail
		Honors, Pass, Conditional Pass, Fail
		Honors, Pass, Conditional Pass, Withdraw
		Honors, Pass, Conditional Pass/Pass, No Pass
		Honors, Pass, Fail, Withdraw
		Outstanding, Advanced, Proficient, Unsatisfactory
		Quartiles
5	33 (25)	A, A-, B+, B, B-A, B+, B, C+, C (3)
		A, B, C, D, F (5)
		Excellent with Honors, Excellent, Good, Marginal, Unsatisfactory
		GPA
		Honors, A, B, C, F
		Honors, High Pass, Pass, Below Pass, Fail
		Honors, High Pass, Pass, Conditional Pass, Fail (2)
		Honors, High Pass, Pass, Conditional, Fail (3)
		Honors, High Pass, Pass, Fail, Incomplete
		Honors, High Pass, Pass, Low Pass, Fail (3)
		Honors, High Pass, Pass, Marginal Pass, Fail
		Honors, High Pass, Pass, Marginal, Fail (2)
		Honors, High Pass, Pass, Pass with Remediation, Fail
		Honors, High Pass, Pass, Satisfactory, Fail
		Honors, High Satisfactory, Satisfactory, Low Satisfactory, Unsatisfactory
		Honors, High Satisfactory, Satisfactory, Marginal Unsatisfactory, Unsatisfactory
		Honors, Near Honors, Satisfactory, Marginal, Fail
		Outstanding, Above Expected, Expected, Below Expected, Fail
		Outstanding, Good, Satisfactory, Fail, Incomplete
6	6 (5)	A, AB, B, BC, C, F
		A, B+, B, C+, C, Incomplete
		Honors, High Pass, Pass, Fail/Pass (FP), Fail, Withdraw
		Honors, High Pass, Pass, Fail, Incomplete, Withdraw
		Honors, High Pass, Pass, Low Pass, Fail, Incomplete
		Honors, High Pass, Pass, Partial Competency/Pass, Partial Competency/Fail, Fail
11	1 (1)	Honors, A/Pass, A-, B+, B, C+, C, C-, D/C, D, F

Note: 3 universities were missing information on number of tiers and terminology; 4 universities were missing information on terminology.
[a] Numbers indicate how many schools reported the grading terminology; if no number is listed, only one school reported using that scheme.

these grades context. From analyzing the MSPEs of the majority of medical schools in the US, the present study demonstrates the persistent wide variability associated with core surgery clerkship grading, ranging from the grading tiers and terminology used to the composition and weight of grading components. Furthermore, the distribution of students in the top tier is disproportionately high, with the majority of medical schools reporting at least a third or more of their students having the highest grade, while essentially no students are placed in the bottom tiers. It is also important to note that a sizeable number of schools are not yet adhering to the most recent AAMC content guidelines for the MSPE with regards to providing information about the various components that determine the grades, which makes interpreting the grades even more challenging.

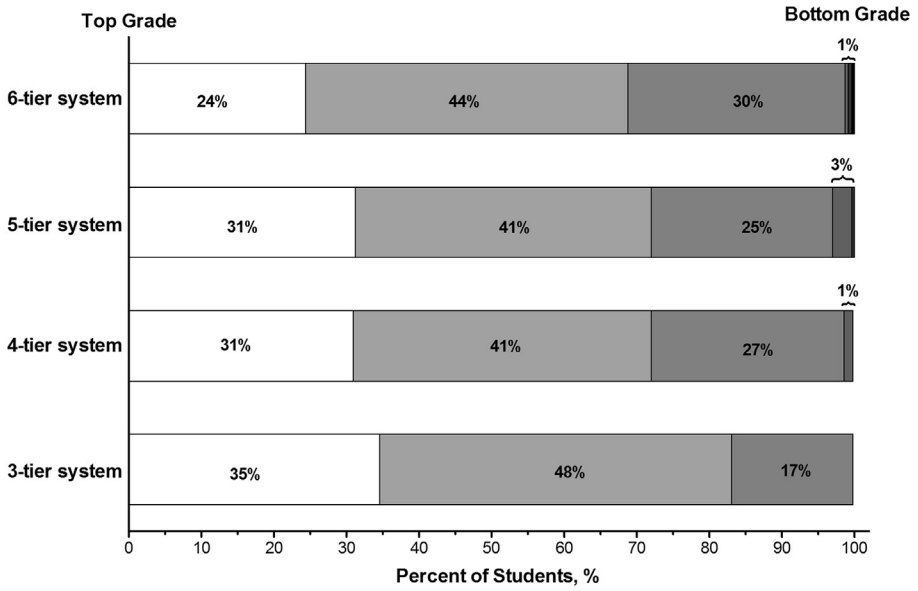

Fig. 1. Distribution of medical students' surgical clerkship grades, stratified by grading tier systems.

The number of grading schema ranged from two to eleven grading tiers in this study, which has been consistent at least since 2004.[2,3] Many residency programs will convert these grading tiers to an internal point system and given the difference in grade distribution based on the number of tiers, the conversion alone may introduce a source of bias.[3] The variation in grading terminology is similarly complicated for different reasons. First, the same number of tiers can be labeled with vastly incongruent terms. For example, in a four-tier system, a "Pass" can mean either the second or third highest tier. Secondly, this variation in grading terminology also makes it difficult to determine whether the student is within the highest grading tier, especially for schools that do not report the grading distribution of their students. For most schools, "Honors" denotes the highest tier, but in one case, "Honors with Distinction" then "Honors" indicates the highest and second tier respectively. In other schools, "Superior" and "Outstanding," represented the highest tier. Given that not every school provides comparative data and that there were 2391 general surgery applicants in the 2017–2018 cycle,[11] this creates a challenge for residency programs to reasonably parse and understand each medical school's grading scheme when evaluating applicants.

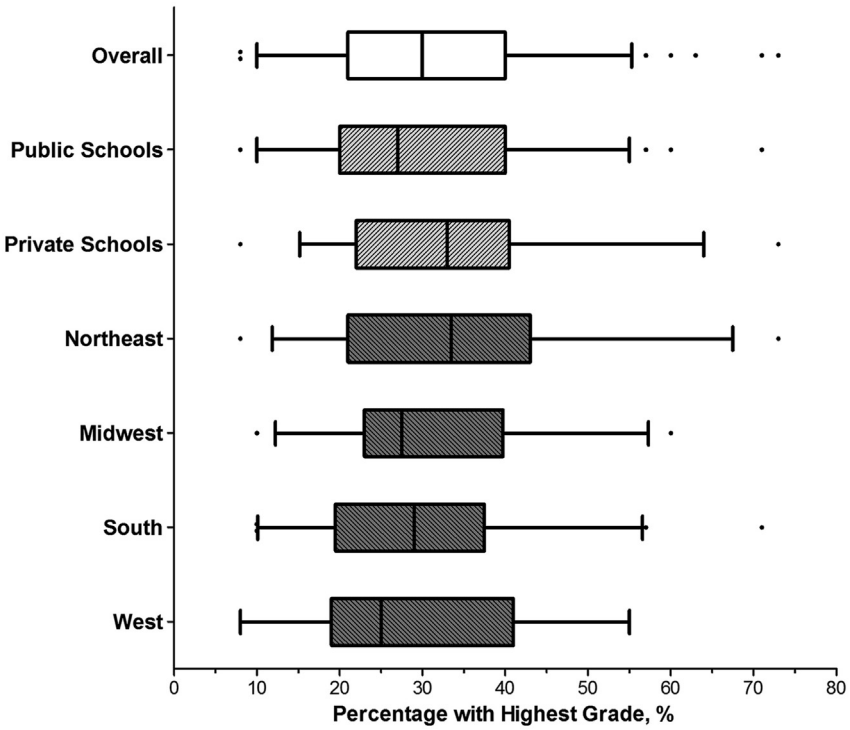

Fig. 2. Distribution of the proportion of medical students given the highest surgical clerkship grade, stratified by school type (public vs. private) and region.

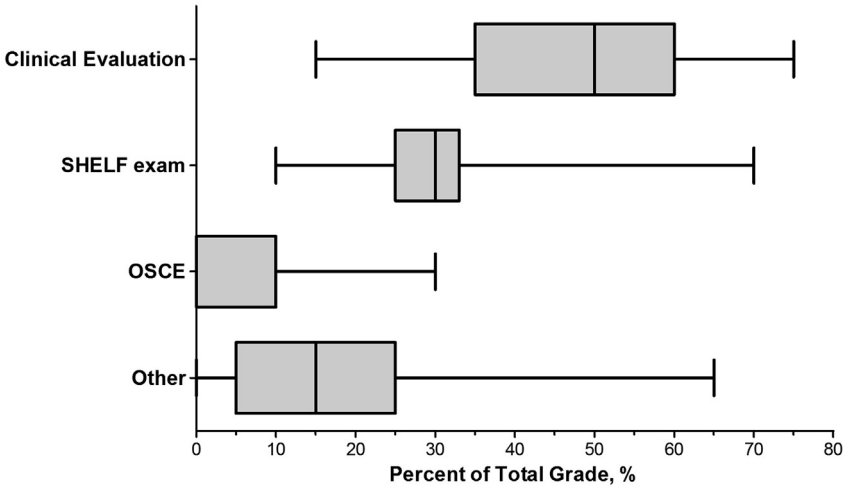

Fig. 3. Distribution of the weight given to grade components.

These results also demonstrate persistence of grade inflation with approximately one third of students in the highest grading tier. Conversely essentially no students are placed in the lowest tier. Our results align with findings from a previous study by Takayama et al., a retrospective cross-sectional study published in 2006 with similar methodology, which found that the percentage of students receiving the highest surgery grade in their institution ranged from 0% to 63% with an average of 27% of students in the highest tier.[3] Another study, analyzing clerkship enrollment during the 2009–2010 admission cycle, by Alexander et al. echo similar sentiments. In that study, approximately 97% of students were awarded one of the top three grades regardless of what grading scheme was used, with very few students in the lowest tier.[2] Other studies analyzing clerkship grading practices across specialties, including surgery, have also demonstrated similar findings of grade inflation.[1,12,13] A disproportionately high number of students in the highest tier also indicates that use of the surgery clerkship grade may be an unreliable measure of actual competency, decreasing its value in a student's application. What has yet to be determined by the graduate medical school community is the role grades play in assessing competence and preparedness to be successful in a residency training program. As long as grading tiers exist, having a consensus agreement across medical schools with respect to what defines “honors” or the top grade would be beneficial. Lipman and Schenarts have begun this work by using the Delphi technique to hone in on the top ten characteristics of an honor student in a core surgical clerkship.[13] Integrating this work into common practice has yet to occur. As the move towards competency-based medical education continues, the grading schema used will need to be consistent across schools to avoid confusion similar to what currently exists today.[14]

Low compliance of schools to AAMC guidelines for the MSPE has been an ongoing concern.[4,14] This is still true in light of the 2016 AAMC MSPE Task Force recommendations to encourage increased standardization and transparency of information presented by schools. In particular, these guidelines now suggest schools report the various components and their weights relative to the final grade.[5] To the authors' knowledge, this is the first paper to evaluate the extent to which medical schools are providing core surgical clerkship grading components in the MSPEs since the 2016 recommendations were released. The low adherence to the AAMC guidelines poses an obstacle for residency directors, as this comparative data is necessary to translate the content of the MSPE across schools. Without the complete data, the tendency often is to rely on standardized test scores as an objective measure of medical knowledge. However, this practice neglects assessing the other five core ACGME competencies which are more subjective in nature.[15]

Currently reported, the core surgical clerkship clinical evaluation is most commonly given the most weight with respect to the final grade, followed by the NBME shelf exam. This reality contributes to the difficulty in interpreting an applicant's surgery grade, as studies have shown clinical evaluations correlate poorly with objective performance and may not reflect clinical knowledge.[16] OSCEs were not frequently used to complement clinical evaluations. There may be an increased role for OSCEs in overall surgery clerkship grades, which may be able to provide additional information for clinical skills appraisal and subject knowledge.[17,18,19] Our findings also reiterate the need to be intentional about how competency-based medical education measures are defined and reported in the MSPE as we move towards that dogma.[14] Regardless, currently grading components are not consistently reported or weighted, making it difficult to compare students from different schools.

Conclusions

Wide variation in core surgery clerkship grading systems persists among U.S. medical schools and has not changed over time. In addition, while schools commonly share distribution of grades, there is low adherence to current AAMC guidelines endorsing clear communication as to how grades are calculated. The author's recommendations include further standardizing grading systems and increasing adherence to AAMC recommendations for the MSPE with regards to reporting comparative data and how grades are determined. These changes will be important as medical student education explores the future use of competency based measures of assessment.

Funding sources

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Acknowledgements

None.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.amjsurg.2018.09.024.

References

1. Bowen RE, Grant WJ, Schenarts KD. The sum is greater than its parts: clinical evaluations and grade inflation in the surgery clerkship. *Am J Surg*. 2015;209(4):760–764. https://doi.org/10.1016/j.amjsurg.2014.10.023.
2. Alexander EK, Osman NY, Walling JL, Mitchell VG. Variation and imprecision of clerkship grading in U.S. medical schools. *Acad Med*. 2012;87(8):1070–1076. https://doi.org/10.1097/ACM.0b013e31825d0a2a.
3. Takayama H, Grinsell R, Brock D, Foy H, Pellegrini C, Horvath K. Is it appropriate to use core clerkship grades in the selection of residents? *Curr Surg*. 2006;63(6):391–396. https://doi.org/10.1016/j.cursur.2006.06.012.
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. https://doi.org/10.1097/acm.0000000000001034.
5. Addams, A. Recommendations for Revising the Medical Student Performance Evaluation (MSPE). https://www.aamc.org/download/470400/data/mspe-recommendations.pdf. Accessed 4 May 2018.
6. Catalanotti JS, Andolsek KM, Berger JS. The AAMC medical student performance evaluation task force recommendations: do they go far enough? *J Grad Med Educ*. 2017;9(1):1–4. https://doi.org/10.4300/jgme-d-16-00698.1.
7. AAMC. Participating Medical Schools and Deadlines. https://students-residents.aamc.org/applying-medical-school/article/participating-medical-schools-deadlines/. Accessed 4 May 2018.
8. AAMC. Regional Membership by State - Regions - Group on Student Affairs (GSA) - Member Center - AAMC. https://www.aamc.org/members/gsa/regions/309964/regionalmembershipbystate.html. Accessed 4 May 2018.
9. Daly SC, Deal RA, Rinewalt DE, et al. Higher clinical performance during a surgical clerkship is independently associated with matriculation of medical students into general surgery. *Am J Surg*. 2014;207(4):623–627. https://doi.org/10.1016/j.amjsurg.2013.07.037.
10. Saguil A, Balog EK, Goldenberg MN, et al. The association between specialty match and third-year clerkship performance. *Mil Med*. 2012;177(9 Suppl):47–52.
11. Bruce Alexander BC, Zaid Almarzooq BI, Keith Armitage BB, et al. *NRMP Results and Data: 2018 Main Residency Match*; 2017. http://www.nrmp.org/wp-content/uploads/2018/04/Main-Match-Result-and-Data-2018.pdf. Accessed May 3, 2018.
12. Grover S, Swisher-McClure S, Sosnowicz S, et al. Grade inflation in medical student radiation oncology clerkships: missed opportunities for feedback? *Int J Radiat Oncol Biol Phys*. 2015;92(4):740–744. https://doi.org/10.1016/j.ijrobp.2015.03.039.
13. Lipman JM, Schenarts KD. Defining honors in the surgery clerkship. *J Am Coll Surg*. 2016;223(4):665–669. https://doi.org/10.1016/j.jamcollsurg.2016.07.008.
14. Andolsek KM. Improving the medical student performance evaluation to facilitate resident selection. *Acad Med*. 2016;91(11):1475–1479. https://doi.org/10.1097/acm.0000000000001386.
15. 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. https://doi.org/10.5811/westjem.2016.10.32233.
16. Goldstein SD, Lindeman B, Colbert-Getz J, et al. Faculty and resident evaluations of medical students on a surgery clerkship correlate poorly with standardized exam scores. *Am J Surg*. 2014;207(2):231–235. https://doi.org/10.1016/j.amjsurg.2013.10.008.
17. Butler KL, Hirsh DA, Petrusa ER, et al. Surgery clerkship evaluations are insufficient for clinical skills appraisal: the value of a medical student surgical objective structured clinical examination. *J Surg Educ*. 2017;74(2):286–294. https://doi.org/10.1016/j.jsurg.2016.08.018.
18. Dong T, Zahn C, Saguil A, et al. The associations between clerkship objective structured clinical examination (OSCE) grades and subsequent performance. *Teach Learn Med*. 2017;29(3):280–285. https://doi.org/10.1080/10401334.2017.1279057.
19. Merrick HW, Nowacek G, Boyer J, Robertson J. Comparison of the objective structured clinical examination with the performance of third-year medical students in surgery. *Am J Surg*. 2000;179(4):286–288. https://doi.org/10.1016/S0002-9610(00)00340-8.