

Available online at [www.sciencedirect.com](http://www.sciencedirect.com)

ScienceDirect

journal homepage: [www.JournalofSurgicalResearch.com](http://www.JournalofSurgicalResearch.com)

## Association for Academic Surgery

# Step 1 is Pass/Fail, Now What? Can Clinical Clerkship Grades be Used as a Reliable Metric to Screen General Surgery Residency Applicants?



Kristen M. Quinn, MD,<sup>a,\*</sup> Laura Campbell, BS,<sup>b</sup> Rupak Mukherjee, PhD,<sup>a</sup>  
 Andrea M. Abbott, MD MSCR,<sup>a</sup> and Christian J. Streck, MD<sup>a</sup>

<sup>a</sup> Department of Surgery, Medical University of South Carolina, Charleston, South Carolina

<sup>b</sup> College of Medicine, Medical University of South Carolina, Charleston, South Carolina

## ARTICLE INFO

## Article history:

Received 24 December 2021

Received in revised form

30 May 2022

Accepted 13 June 2022

Available online 1 August 2022

## Keywords:

Clerkship performance

Clerkship scores

Medical student education

Residency application

Step 1 pass/fail

## ABSTRACT

**Introduction:** For decades, the three-digit United States Medical Licensing Exam Step 1 score has been used to competitively evaluate and compare candidates during the residency application process. Starting in 2022, however, all Step 1 scores will be converted to pass/fail. A different quantitative measure will likely gain importance in its stead, one such being clerkship performance grades. This study aims to determine the consistency of class rank and distribution of clerkship grades reported by medical schools for applicants to a general surgery program.

**Methods:** Candidates' Medical Student Performance Evaluation letters from 141 unique US allopathic medical schools were reviewed for student overall class rank, the number of grading tiers in each clerkship, and the percent achieving honors criteria in each clerkship from the 2020 application cycle. Comparative analysis was performed by region and medical school prestige.

**Results:** Most medical schools rank students using a four-tier system (e.g., fail, pass, high pass, and honors). A third of schools do not provide an overall class rank of students (34.7% of schools); this was most prevalent in the Northeast and Western regions. Schools in the Central US more often rank their students in five tiers compared to the South ( $P < 0.01$ ). The percent of students that achieve the highest grading tier varies across the core clerkships (mean 37.1%, range 6.5%-78%); an average of 34.5% of students meet the highest honors tier in their Surgery clerkship. Students at US News and World Report Top 20 medical schools are more likely to receive the highest honors tier, across all core clerkships and overall class rank, than students at schools outside the Top 20 ( $P < 0.05$ ).

**Conclusions:** In the absence of the United States Medical Licensing Exam Step 1 score, the variability in clerkship grading tiers and overall class rank will likely pose a challenge to residency programs' ability to stratify desirable applicants. Further transparency and standardization may be required to compare students objectively and fairly from medical schools across the country.

© 2022 Elsevier Inc. All rights reserved.

\* Corresponding author. Department of Surgery, Medical University of South Carolina, 96 Jonathan Lucas Street Room CSB 417, Charleston, SC 29425. Tel.: 843-792-2300.

E-mail address: [quinnkr@musc.edu](mailto:quinnkr@musc.edu) (K.M. Quinn).

0022-4804/\$ – see front matter © 2022 Elsevier Inc. All rights reserved.

<https://doi.org/10.1016/j.jss.2022.06.047>

---

## Introduction

In February 2020, the National Board of Medical Examiners and Federation of State Medical Boards announced the decision to change the reporting of the United States Medical Licensing Exam (USMLE) Step 1 exam from scored to pass/fail effective January 2, 2022.<sup>1</sup> This change was met with mixed response among residency program directors (PDs)<sup>2</sup> and presents new challenges to the residency selection process that has heavily relied on the numerical Step 1 score to compare candidates.

Despite being introduced in the 1990s as a pass/fail exam, Step 1 has become one of the most heavily weighted factors in determining candidates that would be granted interviews. Indeed, over 94% of residency program directors indicated that Step 1 was a significant factor in selecting residents to interview.<sup>3</sup> Consequently, the pressure associated with Step 1 performance has contributed to the culture of burnout, anxiety, depression, and suicidal ideation in trainees, and these factors are associated with the ultimate delivery of lower quality patient care.<sup>4-6</sup> Overemphasis on the importance of Step 1 scores in the residency application process created a “Step 1 Climate” and a parallel curriculum generated by a business ecosystem geared to maximizing Step 1 performance.<sup>7</sup> Some pre-clinical students chose to focus their study efforts on third-party, commercial board review resources, at times abandoning their institution’s curricula and possibly further reinforcing socioeconomic disparities noted in medical education.<sup>4</sup> Underrepresented minorities in medicine have lower Step 1 score averages, and have disproportionately been affected by Step 1 cut-off scores used as an applicant screening tool.<sup>8-10</sup> Therefore, the National Board of Medical Examiners stated reason for changing Step 1 score reporting was to minimize negative effects of this high-stakes exam on medical students’ well-being and medical education in general.<sup>3,11</sup>

Despite the anticipated benefits of changing Step 1 score reporting, drawbacks of this change have also become apparent. The standardized Step 1 exam helped to level the playing field between students at schools with greater name recognition versus state schools or smaller, lesser-known private medical schools. The elimination of numerical scores may lead to greater importance placed on medical school prestige and ranking, increasing the difficulty for International Medical Graduates or applicants from osteopathic schools, as examples. Furthermore, there is a perception this change simply “kicks the can down the road”, meaning the numerical score reporting of Step 2 or another quantitative ranking metric may hold emphasis the way Step 1 had. One grading scheme that may draw additional focus is student performance in core clerkships. Program directors from all specialties recently reported they will likely place more importance on clinical core clerkship performance evaluations, components found in applicants’ Medical Student Performance Evaluation (MSPE) letters.<sup>2</sup> The MSPE letter, formerly referred to as the Dean’s letter, summarizes a medical student’s performance and qualitatively evaluates the student’s potential as a future resident physician. The MSPE letter includes comparative data about individual medical

student’s pre-clinical and clinical performance and objective data comparing individual medical students to their peers. However, these letters are institution-dependent, and the information provided varies.<sup>12-15</sup> The present study sought to describe the consistency of student rank and the distribution of grades within core clerkships across US allopathic medical schools as reported in candidates’ MSPE letters. In this context, we can better understand the utility of the MSPE letter and its potential role in filling the void in the surgery residency selection process that changes in Step 1 score reporting have created.

---

## Methods

MSPE letters received from 141 US allopathic medical schools at a single general surgery residency during the 2020-2021 application cycles were reviewed. Information provided on the number of grading tiers, honors criteria, and grade distribution were extracted. This study was deemed exempt by our institutional review board.

The MSPE letters have a summary statement, either at the beginning or end, which highlights the applicant’s most notable achievements, traits, or abilities, and compares the student to their graduating class. This *overall rank* is often annotated as a level of recommendation from the Dean’s Office, as the applicant being, for example, an ‘outstanding’, ‘excellent’, or ‘very good’ candidate. Those three categories were taken to represent a three-tier *overall rank* system. Clerkship grading tiers were also defined as the number of different grades that a student could achieve in a clinical clerkship. For example, a school that placed students in three grading tiers represented all students receiving one of three qualifiers—fail, pass, or honors. Four tiers may represent fail, pass, high pass, and honors. Specific component criterion for achieving the highest honors were not routinely reported and not analyzed in this study. We collected instead for comparison the percent of students who were awarded the highest tier performance in each clerkship, referred hereafter as *Clerkship honors*. Medical schools were categorized by geographical region assigned by the AAMC and research rank assigned by the 2022 US News and World Report.<sup>16</sup>

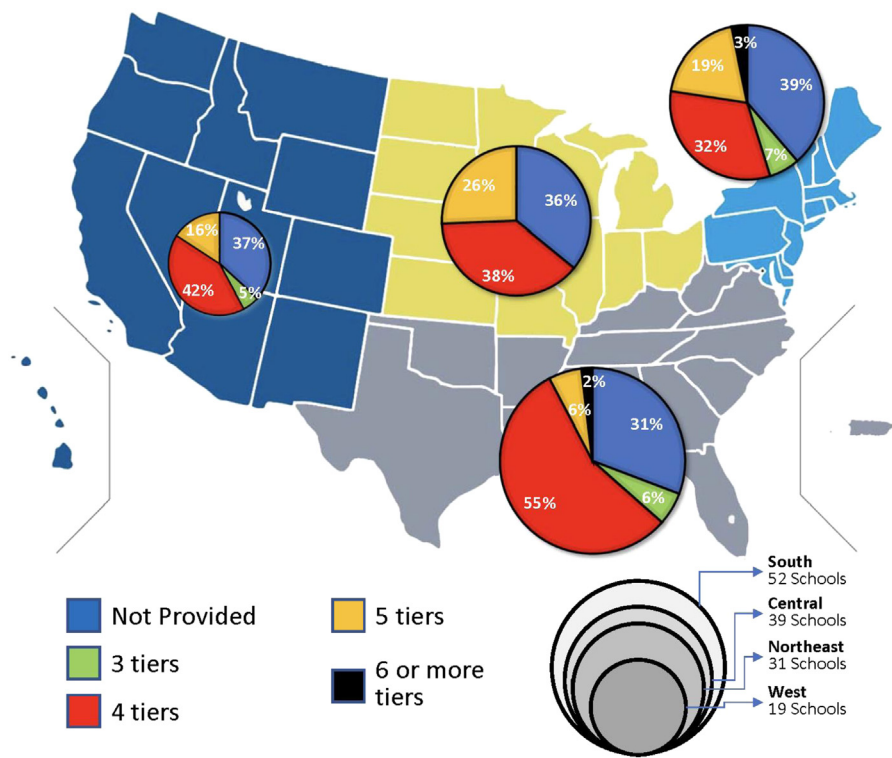
Aggregate data were analyzed using Microsoft Excel, GraphPad Prism (v9.0), and Stata (v16.1). Two-sample t-tests and Wilcoxon rank-sum (Mann–Whitney) tests were performed, as appropriate. *P* values of less than 0.05 were considered statistically significant.

---

## Results

### Overall rank

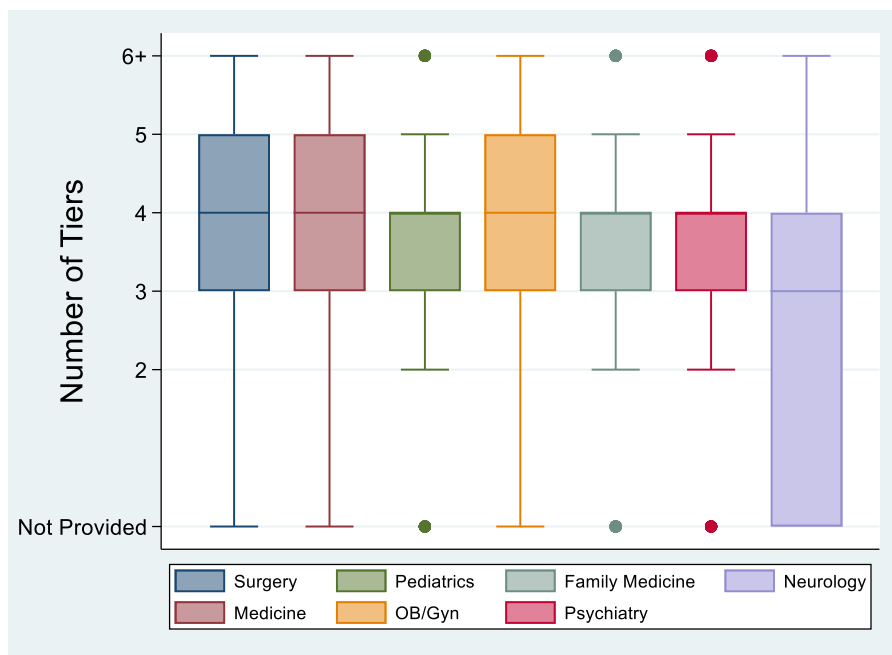
Overall class rank of the applicant was not always included, described consistently, or located in the same place in medical school MSPE letters. The rubric or transparency by which *overall rank* was calculated, the qualifiers by which it was denoted, and the number of *overall rank* tiers varied. *Overall rank* was most reported within four tiers (median = 4), but not



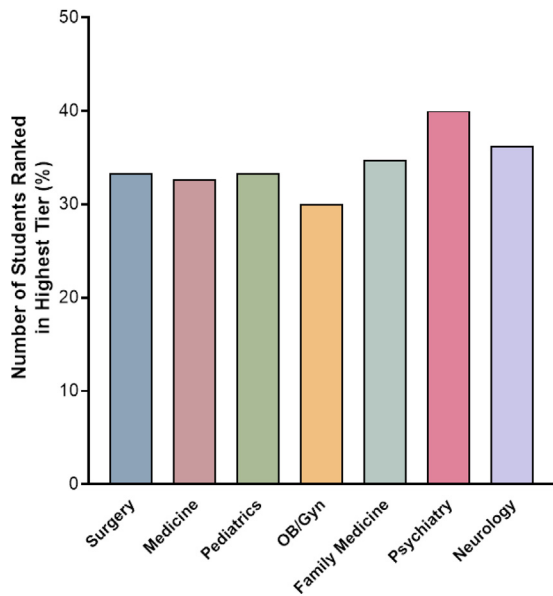
**Fig. 1 – Pie charts of the number of ranking tiers by region, including the proportion of medical schools in each region that did not provide information on the number of ranking tiers used.**

necessarily in equal quartiles. Second-most commonly, overall rank was not provided in about a third of the medical school MSPE letters (34.7% of schools). This trend appeared to be most

prevalent in schools from the Northeast and West. Schools in the Central region more often rank their students in five tiers compared to the South ( $P < 0.01$ ) (Fig. 1). Over half of the



**Fig. 2 – Box and whisker plots of grading tiers by each core clinical clerkship. The median is represented by the darker line within each box. The upper edge of each box represents the 75th percentile and the bottom edge the 25th percentile. The caps at the end of the lines represent the 10th and 90th percentiles, respectively. Data points outside the caps are individual values that are either less than the 10th percentile or greater than the 90th percentile.**



	Surgery	Medicine	Pediatrics	OB/Gyn	Family Medicine	Psychiatry	Neurology
% Ranked in Highest Tier	34.47%	35.03%	36.44%	34.26%	37%	43.60%	38.26%

**Fig. 3 – Percentage of students achieving honors criteria, or the top grading tier, by clerkship rotation displayed as means (bars and table).**

medical schools in the South (55.8%) provide the overall rank of their students in four tiers, while smaller percentages report using 5 (5.8%), 3 (5.8%), or 6 or more (0.9%) tiers. In the

Northeast, only 32.3% of schools use the four-tier system, and it is most common (38.7%) for Northeastern medical schools to not provide an overall rank. A small percentage (2%) of schools nationally stated they were unable to provide the overall rank of their students due to the COVID-19 pandemic-related disruptions to medical education.

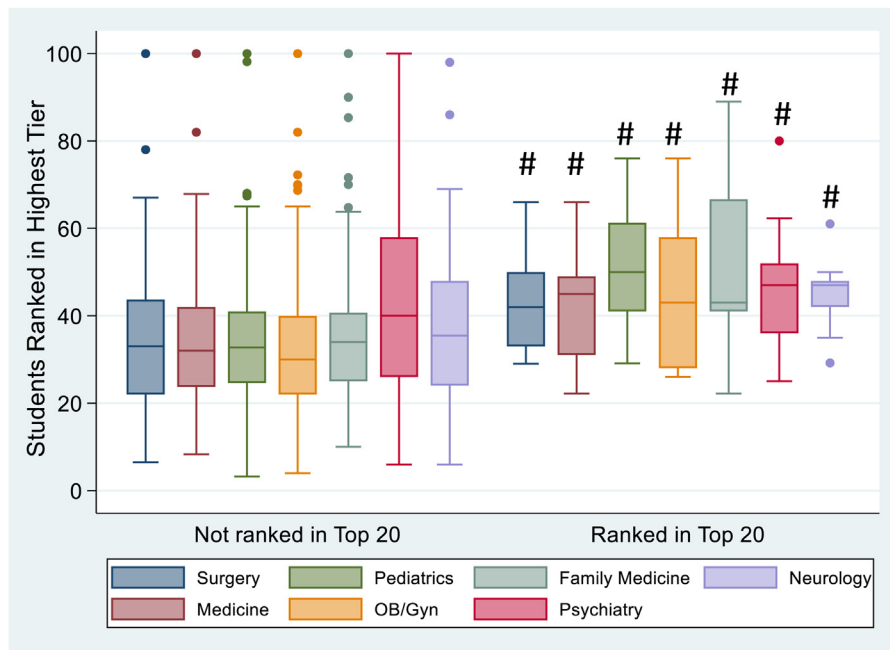
The percent of students who met the highest tier criteria in overall rank was on average the top 23.83% of the medical school class. Considerable variability was observed with a range from the top 0.5% to top 50% of students meeting their respective schools' top tier overall rank.

**Number of grading tiers in clerkships**

While there was large variability in the number of grading tiers used by individual US medical schools, the median number of tiers across the country was four, despite region. There were no significant differences found between specific core rotations (Fig. 2 and Supplementary Fig. 1) or medical school rank on the number of grading tiers used for core clerkship grading.

**Clerkship honors**

Performance in core clerkships may be overemphasized by reporting a greater percent of students achieving honors or the top grading tier. Schools establish their own criteria for honors, typically based on a combination of clinical performance on the wards, aptitude on a standardized clerkship exam, and evaluations by residents and faculty. Many of these criteria are subjective, with the clerkship standardized exam



**Fig. 4 – Box and whisker plots of the percentage of students reported as achieving honors criteria, or the highest grading tier, for each core clinical clerkship compared between medical schools ranked as US News Report Research Top 20 and those not ranked within the top 20. For each of the clinical clerkships, the percentage of students achieving honors criteria was higher for applicants from the Top 20 schools compared to those from schools not ranked in the top 20. #P < 0.05 for corresponding clerkships.**

representing the only true objective criteria. On average, 37.13% of students meet the highest honors in their clinical clerkship; however, this ranges by clerkship and school. Nationally, fewer students are awarded honors in their obstetrics/gynecology (34.26%) and surgery (34.47%) rotations, while 43.6% of students receive honors in their psychiatry rotation (Fig. 3). The percent of students achieving honors varied dramatically by school, with a range from 6.5 to 78% receiving honors in their surgery across the country.

Stratifying medical schools by those listed in US News and World Report Top 20 Research institutions revealed a significant pattern with respect to the percentage of students accorded the highest tier across all core clinical clerkships. Medical schools ranked in the Top 20 had significantly more students receiving honors compared to all other schools ( $P < 0.05$ ) (Fig. 4).

## Discussion

Our study found high variability in national patterns of ranking overall medical student performance against their class and in core clerkship rotations. An increased percentage of northeastern schools do not rank their students overall medical school performance and clerkship honors are awarded at a higher-than-expected percentage in Top 20 medical schools. With the loss of a nationalized, objective measure like the Step 1 score, PDs are seeking other numerical metrics by which to categorize applicants.<sup>2</sup> The wide range in students' overall rank, reported grading tiers, and honors criteria challenge PD's ability to compare applicants from different schools, especially those of different geographic regions or research ranks. The initial goal of the MSPE letter was to support a student's candidacy for residency by highlighting their achievements and clerkship performance; however, the lack of standardized ranking and variability in percent achieving honors hinder PDs' ability to efficiently and accurately understand how the candidate trainee performed compared to their class and to those from other schools.

Lack of transparency and uniformity may inadvertently permit increased bias in evaluating applicant merit. PDs may be most comfortable and familiar with MSPE letters from medical schools whose applicants have historically matched at their program. This may give an advantage to students who come from a medical school that is either a regional, research rank, or academic/community equivalent to the residency program. Greater weight on MSPE letters may ultimately limit the mobility of students attempting to transition their training from one geographic or academic setting to another. Especially in the recent context of virtual interviews, applicants and PDs have reported decreased ability to demonstrate interest and make lasting impressions.<sup>17</sup> We found Top 20 research rank medical schools award more students honors than schools ranked outside of the top 20 ( $P < 0.05$ ). This finding potentially hinders the applicant from a less prestigious school, who our research shows is already less likely to receive honors, to appear as competitive as an individual student from a Top 20 school. Clerkship scores, by lacking transparency and uniformity, may hamper a PD's ability to sort applications and to allow exceptional candidates to stand out.<sup>9</sup>

We encourage the holistic review of general surgery applicants. The selection of trainees with experiences, characteristics, and achievements that align with a program's mission strengthens both parties. A recent survey of general surgery PD's revealed that the pass/fail score change in combination with a holistic review process may encourage a greater volume of applicants in general surgery.<sup>18</sup> PD's also felt clerkship grades, medical school reputation, personal statements, letters of recommendation, and other accomplishments in research or service will now have a higher impact.<sup>18</sup> The authors feel students building a more well-rounded application is a strength, but increased weight on less standardized metrics may present an opportunity for increased bias. We encourage medical schools to adopt a uniform grading system and a uniform distribution of grades and provide overall class rank. Our data demonstrated that most schools use four clerkship grading tiers and we recommend a consistent distribution of honors to increase transparency (e.g., 25% honors, 25% high pass, 50% pass).

This study is limited in its inclusion of only US allopathic medical schools of general surgery applicants to our single institution residency program. Our program received at least one application from 141 of the 155 accredited US allopathic medical schools (91%); those schools not included were primarily in a different geographic region than our institution. Osteopathic and international medical schools were not examined. The application cycle analyzed was disrupted by COVID-19. A small percentage (2%) of programs endorsed changes in their grading and curriculum due to the pandemic and some programs may have experienced a grading skew towards not ranking students or placing them all in a one-tier passing category.

Future directions include additional research on clerkship performance and correlation with Step 2 score, the quality or number of letter of recommendations, and reflection on what makes the most effective resident for each residency program. Perhaps only with an enhanced understanding and definition of success in residency, can PDs make the most informed decision on candidates. Each program and specialty will likely develop a different formula/algorithm to score candidates based on criteria they believe will determine success during residency. More evidence is needed to determine what matters the most in producing competent physicians. Reverse engineering the characteristics meritorious individuals displayed as applicants may allow for a more evidence based and effective selection into residency programs.

## Conclusions

A historic over-reliance on the numerical Step 1 score to stratify applicants is easily translated to a sudden increased importance on MSPE letters and clerkship grades. This study revealed that most US allopathic medical schools provide an overall class rank of students using four-tiers, but many schools (39% northeast, 37% western, 36% central, and 31% southern) do not list an overall rank of their students in the MSPE letter. Clerkship grades are most reported using a four-tier system and on average 34.5% of students achieve honors in surgery. Clerkship honors and top-tier class rank is accorded to more students at Top 20 medical schools than schools

not in the Top 20. The purpose of making Step 1 pass/fail may have been to move away from granular, strict, grade-based rankings of students; however, the wide variability in overall class rank and clerkship performance may permit increased bias if greater transparency or uniformity among medical schools is not achieved.

---

## Supplementary Materials

Supplementary data related to this article can be found at <https://doi.org/10.1016/j.jss.2022.06.047>.

---

## Author Contributions

KQ and LC contributed to this work through study design, data collection, analysis, and manuscript preparation. RM contributed with statistical analysis, manuscript edits, and figure preparation. AA and CS contributed through study design, mentorship, and manuscript edits.

---

## Disclosure

None declared.

---

## Funding

None.

---

## REFERENCES

1. Kogan JR, Hauer KE. Sparking change: how a shift to Step 1 Pass/fail scoring could promote the educational and catalytic effects of assessment in medical education. *Acad Med*. 2020;95:1315–1317.
2. Wang A, Karunagan KL, Story JD, Ha EL, Braddock CH, Geffen D. Residency program director perspectives on changes to US medical licensing examination. *JAMA Netw Open*. 2021;4:e2129557.
3. Blamoun J, Hakemi A, Armstead T. Perspectives on transitioning Step 1 of the United States medical licensing examination to a pass/fail scoring model: defining new frameworks for medical students applying for residency. *Adv Med Educ Pract*. 2021;12:149–154.
4. Chen DR, Priest KC, Batten JN, Fragosa LE, Reinfeld BI, Laitman BM. Student perspectives on the “Step 1 climate” in preclinical medical education. *Acad Med*. 2019;94:302–304.
5. Brazeau CM, Shanafelt T, Durning SJ, et al. Distress among matriculating medical students relative to the general population. *Acad Med*. 2014;89:1520–1525.
6. Rotenstein LS, Ramos MA, Torre M, et al. Prevalence of depression, depressive symptoms, and suicidal ideation among medical students: a systematic review and meta-analysis. *JAMA*. 2021;316:2214–2236.
7. Burk-Rafel J, Santen SA, Purkiss J. Study behaviors and USMLE Step 1 performance: implications of a student self-directed parallel curriculum. *Acad Med*. 2017;92:S67–S74.
8. Rubright JD, Jodoin M, Barone MA. Examining demographics, prior academic performance, and United States medical licensing examination scores. *Acad Med*. 2019;94:364–370.
9. Jarman BT, Kallies KJ, Joshi ART, et al. Underrepresented minorities are underrepresented among general surgery applicants selected to interview. *J Surg Educ*. 2019;76:e15–e23.
10. Williams M, Kim E, Pappas K, et al. The impact of United States Medical Licensing Exam (USMLE) step 1 cutoff scores on recruitment of underrepresented minorities in medicine: a retrospective cross-sectional study. *Health Sci Rep*. 2021;3:e2161.
11. Monrad SU, Wolff M, Kurtz J, et al. What is the association between student well-being and high-stakes examination scores? *Med Educ*. 2021;55:872–877.
12. Kies SM, Roth V, Rowland M. Association of third-year medical students’ first clerkship with overall clerkship performance and examination scores. *JAMA*. 2010;304:1220–1226.
13. Lipman JM, Schenarts KD. Defining honors in the surgery clerkship. *J Am Coll Surg*. 2016;223:665–669.
14. Groth DJ, Magee RG, Rosenberg SE, Mingioni N. Gender disparity in evaluation of internal medicine clerkship performance. *JAMA Netw Open*. 2021;4:e2115661.
15. Seligman L, Abdullahi A, Teherani A, Hauer KE. From grading to assessment for learning: a qualitative study of student perceptions surrounding elimination of core clerkship grades and enhanced formative feedback. *Teach Learn Med*. 2021;33:314–325.
16. The Best Medical Schools for Research, Ranked. 2021. Available at: <https://www.usnews.com/best-graduate-schools/top-medical-schools/research-rankings>. Accessed July 26, 2021.
17. Spencer E, Ambinder D, Christiano C, et al. Finding the next Resident Physicians in the COVID-19 Global Pandemic: an Applicant Survey on the 2020 Virtual Urology Residency Match. *Urology*. 2021;157:44–50.
18. Aziz H, Khan S, Rocque B, Javed MU, Sullivan ME, Cooper JT. Selecting the next generation of surgeons: general surgery program directors and coordinators perspective on USMLE changes and holistic approach. *World J Surg*. 2021;45:3258–3265.