

RESEARCH ARTICLE

Open Access

# Objective predictors of intern performance



Amanda C. Filiberto<sup>1</sup>, Lou Ann Cooper<sup>2</sup>, Tyler J. Loftus<sup>1</sup>, Sonja S. Samant<sup>3</sup>, George A. Sarosi Jr<sup>1</sup> and Sanda A. Tan<sup>1\*</sup>

## Abstract

**Background:** Residency programs select medical students for interviews and employment using metrics such as the United States Medical Licensing Examination (USMLE) scores, grade-point average (GPA), and class rank/quartile. It is unclear whether these metrics predict performance as an intern. This study tested the hypothesis that performance on these metrics would predict intern performance.

**Methods:** This single institution, retrospective cohort analysis included 244 graduates from four classes (2015–2018) who completed an Accreditation Council for Graduate Medical Education (ACGME) certified internship and were evaluated by program directors (PDs) at the end of the year. PDs provided a global assessment rating and ratings addressing ACGME competencies (response rate = 47%) with five response options: excellent = 5, very good = 4, acceptable = 3, marginal = 2, unacceptable = 1. PDs also classified interns as outstanding = 4, above average = 3, average = 2, and below average = 1 relative to other interns from the same residency program. Mean USMLE scores (Step 1 and Step 2CK), third-year GPA, class rank, and core competency ratings were compared using Welch's ANOVA and follow-up pairwise t-tests.

**Results:** Better performance on PD evaluations at the end of intern year was associated with higher USMLE Step 1 ( $p = 0.006$ ), Step 2CK ( $p = 0.030$ ), medical school GPA ( $p = 0.020$ ) and class rank ( $p = 0.016$ ). Interns rated as average had lower USMLE scores, GPA, and class rank than those rated as above average or outstanding; there were no significant differences between above average and outstanding interns. Higher rating in each of the ACGME core competencies was associated with better intern performance ( $p < 0.01$ ).

**Conclusions:** Better performance as an intern was associated with higher USMLE scores, medical school GPA and class rank. When USMLE Step 1 reporting changes from numeric scores to pass/fail, residency programs can use other metrics to select medical students for interviews and employment.

**Keywords:** Medical student, Internship, Performance, Residency, Medical education

## Introduction

In selecting medical students for interview and employment, residency programs seek candidates that will perform well as residents and become excellent physicians. The definition of a successful intern likely varies by specialty, but an individual's work competence, organizational acumen, social intelligence and personal characteristics are major contributors [1]. The Accreditation Council for Graduate Medical Education (ACGME) uses competency

based medical education and specialty specific Milestones as a systematic framework to evaluate resident performance and determine promotion to the next level of training [2]. Selection committees consider a variety of factors when offering interviews and creating rank lists, including but not limited to standardized test scores, class rank, clerkship performance, letters of recommendation, and personal statements [3–5]. Determining which objective measures are most predictive of residency performance could help guide selection committees and program directors (PDs).

Historically, residency programs have relied on United States Medical Licensing Examination (USMLE) scores

\* Correspondence: [Sanda.Tan@surgery.ufl.edu](mailto:Sanda.Tan@surgery.ufl.edu)

<sup>1</sup>Department of Surgery, University of Florida Health, 1600 SW Archer Ave, PO Box 100109, Gainesville, FL 32610, USA

Full list of author information is available at the end of the article



© The Author(s). 2021 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

as an objective and consistent measure of medical student performance across different medical schools [4]. Among the 2018 National Resident Matching Program (NRMP) survey of program directors, 94% of the ~ 1200 programs cited USMLE Step 1 score as an important factor for selecting candidates to interview, and more than half reported the use of a prescribed cutoff score, below which applicants are ineligible [6]. In addition to their known use as a screening tool, USMLE scores have been shown to correlate with the NRMP rank list [7] as well as with in service-training exams [4, 5, 8]. However, USMLE scores have had inconsistent relationships with intern performance as assessed by faculty evaluations for several medical specialties [3–5, 8–10]. Additionally, on February 12, 2020 the USMLE announced that Step 1 results will now be reported as a pass/fail outcome, as compared with the previously reported 3-digit numerical score [11]. Therefore, other objective predictors of intern performance may become increasingly important in the process of selecting medical students for interview and employment.

Few studies have evaluated medical school metrics and intern performance across all specialties, and most of the existing studies are small, single institution and involve a single specialty, and thus are limited in generalizability [12, 13]. Due to the lack of consensus regarding the strength and validity of different applicant characteristics as predictors of residency performance, we collected objective data on graduating medical students from a single institution and obtained standardized evaluations from their respective PDs as a metric of intern performance. The aim of this study was to assess independent predictors of intern performance, so that these predictors can be used to accurately and precisely rank future applicants. This study tested the hypothesis that performance on USMLE scores, grade-point average (GPA), class rank, and class quartile would predict intern performance.

**Methods**

This single-institution retrospective cohort analysis included 244 graduates from the University of Florida College of Medicine during a four-year period (2015–2018) who subsequently completed an ACGME certified internship. Institutional Review Board approval was obtained. To assess intern performance, standardized evaluation forms were provided to residency PDs with a signed authorization form (Appendix A) at the conclusion of each year, surveyed annually for 4 years. The evaluation form was provided to PDs at the conclusion of the resident’s first year, and PDs were instructed to evaluate the resident as a first year intern. PDs provided a global assessment rating and ratings on specific items addressing the six ACGME core competencies (response

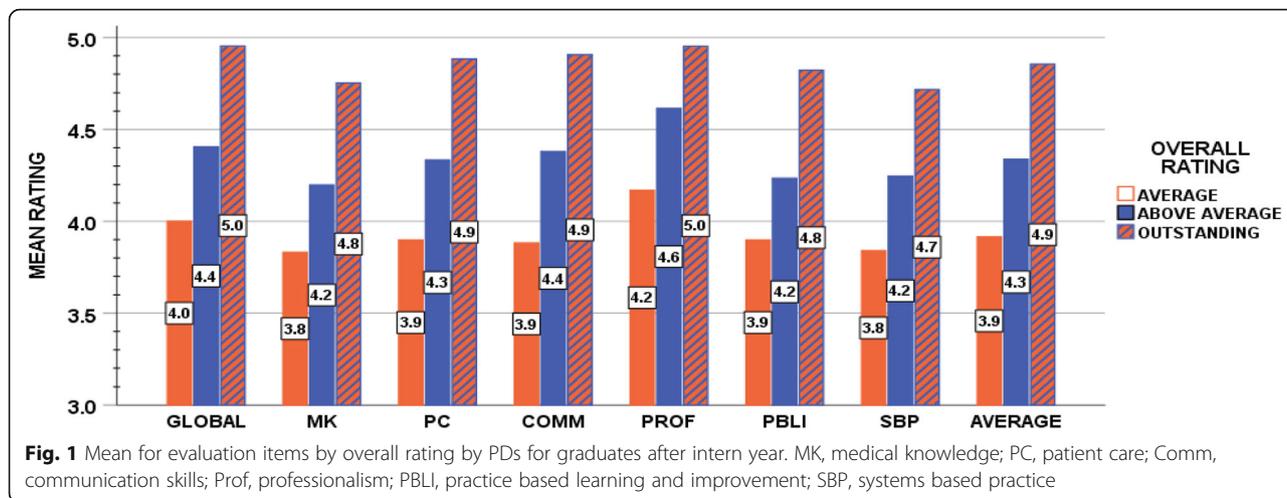
rate = 47%). Evaluation items had five response options: excellent = 5, very good = 4, acceptable = 3, marginal = 2, and unacceptable = 1. PDs were also asked to classify interns as outstanding (*n* = 85), above average (*n* = 95), average (*n* = 60), or below average (*n* = 4) relative to the other interns in their cohort (cohort comparison rating).

The medical school curriculum at our institution has had no significant changes during this period of time. Means for these metrics were compared between graduating classes, with subgroup analyses of interns who were evaluated by PDs and interns that were not evaluated by PDs. These analyses found no statistical differences (*p* < 0.05). We also found similar distributions of specialties in the two groups (Table 1); therefore, the data for all 4 years were analyzed in aggregate. Descriptive statistics and analyses were calculated using SAS software version 9.4 (SAS Institute Inc., Cary, NC); Fig. 1

**Table 1** Medical school graduates by specialty

	All graduates	Graduates with PD evaluations
<b>Specialty, n (%)</b>		
Anesthesia	33 (6)	23 (9)
Dermatology	15 (3)	6 (2)
Diagnostic radiology	20 (4)	10 (4)
Emergency medicine	64 (12)	31 (13)
Family medicine	32 (6)	17 (7)
General surgery	30 (6)	8 (3)
IM/Peds	3 (1)	1 (0)
Internal medicine	99 (19)	46 (19)
IR	1 (0)	1 (0)
Neurology	8 (2)	4 (2)
Neurosurgery	11 (2)	1 (0)
OB/Gyn	47 (9)	25 (10)
OMFS	14 (3)	2 (1)
Ophthalmology	14 (3)	5 (2)
Orthopedic	16 (3)	5 (2)
Otolaryngology	6 (1)	2 (1)
Pathology	6 (1)	1 (0)
Pediatrics	58 (11)	35 (14)
Plastic surgery	3 (1)	1 (0)
PM&R	2 (0)	1 (0)
Psychiatry	19 (4)	12 (5)
Radiation oncology	9 (2)	4 (2)
Urology	9 (2)	4 (2)
Vascular surgery	0 (0)	0 (0)
Thoracic surgery	2 (0)	0 (0)

PD program director, IM internal medicine, IR interventional radiology, Peds pediatrics, OB/Gyn obstetrics and gynecology, OMFS oral maxillofacial surgery, PM&R physical medicine and rehabilitation



was constructed using SPSS version 25 (IBM Corp., Armonk, N.Y). Mean USMLE scores (Step 1 and Step 2CK), third-year GPA (required clinical clerkships), class rank, and core competency ratings were compared using Welch’s ANOVA and follow-up pairwise t-tests.

**Results**

There were a total of 521 graduates from the University of Florida College of Medicine from 2015 to 2018. Two hundred and forty-four graduates were evaluated by their PDs using a standardized evaluation form (response rate = 47%). Interns in the following specialties had the highest PD response rates: internal medicine (17%), emergency medicine (14%), pediatrics (14%), and obstetrics and gynecology (10%). There were no significant differences among graduates that received evaluations and those that did not (Table 1). Six percent of all University of Florida graduates matriculated into general surgery, which is similar to the 2020 national average of 4.7%. Subgroup analyses demonstrated that students matching in general surgery and other surgical specialties had baseline characteristics that were similar to students matching into non-surgical specialties. There were no significant differences among the four classes of medical student graduates (2015–2018) in terms of

performance metrics, core competencies or global assessment.

Descriptive statistics of the sample stratified by PD’s assessment of intern performance (relative to other interns from the same residency program) are shown in Table 2. Descriptive statistics are provided for the students evaluated as below average ( $n = 4$ ), but this group is not included in the statistical analyses. One-way ANOVA showed higher GPA ( $p = 0.02$ ) and class rank ( $p = 0.016$ ) were associated with better performance, as indicated by PD overall assessment of intern performance as average, above average or outstanding. Better intern performance was also associated with higher USMLE Step 1 ( $p = 0.006$ ) and Step 2CK ( $p = 0.030$ ) scores. Pairwise t-tests demonstrated that interns rated as average had lower scores, lower GPA, and were more likely to have a lower class rank than interns rated as above average or outstanding. Although on average the metrics were higher for interns rated outstanding than for those rated above average, these differences were not statistically significant.

Figure 1 shows the results of the mean scores for the evaluation items using the overall rating as the grouping variable. Higher rating in each of the core competencies was associated with better performance as an intern, as

**Table 2** Medical school performance metrics stratified by program director’s overall assessment of intern performance (relative to their cohort) at the end of intern year

	Below Average <sup>a</sup> ( $n = 4$ )	Average ( $n = 60$ )	Above Average ( $n = 95$ )	Outstanding ( $n = 85$ )	ANOVA $F$ $p$	Follow up t-tests for significant ANOVA					
						A vs. AA	A vs. O	AA vs. O	$t$	$p$	$t$
Performance metrics	M (SD)	M (SD)	M (SD)	M (SD)		$t$	$p$	$t$	$p$	$t$	$p$
USMLE Step 1	210.5 (24.0)	224.6 (19.3)	232.4 (18.6)	234.6 (17.4)	5.3 < 0.01	2.5	0.01	3.2	< 0.01	0.9	0.38
USMLE Step 2CK	241.5 (15.8)	243.4 (16.1)	249.1 (15.9)	250.4 (15.6)	3.6 0.03	2.1	0.04	2.6	< 0.01	0.6	0.54
MS3 GPA	3.2 (0.1)	3.4 (0.4)	3.5 (0.3)	3.6 (0.3)	4.0 0.02	2.4	0.02	3.0	< 0.01	0.7	0.48
MS3 Rank	105.8 (11.6)	76.8 (37.5)	63.1 (38.6)	58.5 (38.8)	4.2 0.02	-2.1	0.04	-2.8	< 0.01	-0.8	0.41

<sup>a</sup>Descriptive statistics provided for the students evaluated as “below average”, but this group is not included in the statistical analyses reported in this table

indicated by PD overall assessment of intern performance as average, above average or outstanding. For global assessment, the items based on ACGME core competencies and the average of these seven items, the clear differences shown in the graph are significant. Pairwise *t*-tests demonstrated that interns rated as average had lower competency ratings than those rated as above average or outstanding ( $p < 0.01$ ). A chi square test of association between the global assessment rating and the overall cohort comparison rating demonstrated a strong relationship between these two measures,  $\chi^2(9) = 299.4$ ,  $p < 0.0001$ .

## Discussion

Intern performance evaluations were significantly associated with objective measures of academic performance in medical school. These findings were consistent when interns were compared to their peers as well as a global assessment by PDs. There were no significant differences between above average and outstanding interns, reflecting the variability and overlap of the distribution of performance metrics among graduates that are making successful progress in their training.

With the introduction of policy changes to USMLE score reporting, it is important to consider other metrics that can be used by PDs to evaluate the growing pool of residency applicants. In addition to Step 1, our study found a significant association between performance as an intern and GPA, class rank and Step 2 CK. As the Step 2 CK examination places greater emphasis on the clinical application of medical knowledge, it may represent a better predictive measure of clinical performance. Unlike the USMLE, preclinical coursework and grading schemes vary substantially from school to school, which make it difficult to compare applicants but may be a useful tool for PDs evaluating multiple applicants from a single institution. Quartile designation, based on class rank determined by third year GPA, is included in the MSPE (Dean's letter) and may offer a way to compare applicants.

As the number of applicants to training programs continues to climb while the number of positions available remains unchanged [14], residency programs continue to face the difficult task of selecting future residents. The level of scrutiny of academic variables is difficult to quantify as residency programs typically do not publish criteria used to interview or score applicants. USMLE scores appear to be related to success in the match process [15], as specialties with the highest percentages of unmatched U.S. seniors such as otolaryngology and neurosurgery [14] have higher USMLE Step 1 and Step 2 CK scores [16]. Further, several studies have shown an association between USMLE scores and performance on in service-training examinations across several specialties

[17–20]. However, few studies have evaluated which factors predict success as a resident.

The results of our study are consistent with findings from several other studies regarding both overall performance as an intern and academic measures in medical school. In a study of 338 medical student graduates from a single institution, Alexander et al. [13] found that higher GPA, and USMLE Step 1 and Step 2 were each associated with better PD assessments of overall performance as an intern. Paolo et al. [12] found similar results when surveying PDs of 382 interns, reporting that residents with higher GPA and USMLE Step 1 and Step 2 scores in medical school were rated higher than those with lower scores. A study by Andriole et al. [9] that involved 87 recent graduates from a single medical school who pursued surgical training found that Step 2 was the only significant predictor of intern performance.

In contrast, several specialty-specific studies have reported that medical school metrics are not significantly associated with resident performance. Fryer et al. were unable to consistently predict any type of objective metric with general surgery resident performance [21]. A small retrospective study of 57 neurology residents by Burish et al. [3] found that Step 1 scores did not correlate with overall neurology resident quality, although it may predict success on future standardized medical examinations. In a study of 69 pediatric house officers, Borowitz et al. [8] found that neither medical school grades or performance on standardized exams were predictors of clinical performance during pediatric residency. Discrepancies in previously reported results may be attributable to the smaller sample sizes in the specialty-specific studies, rendering them underpowered to detect significant associations between medical school metrics and resident performance.

The single-institution, retrospective design of this study limits its generalizability. The analyses performed in this study, however, have not been previously reported using national data, and the authors believe that findings from this study may be useful to educators and serve as framework for larger, multi-center studies. While our overall survey response rates were much higher than the NRMP PD survey response rates [6], data were missing for approximately 53% of the overall cohort. Further, among the known limitations associated with survey response bias, there is the possibility of the halo effect, where raters tend to rely on general perceptions even when they are asked to evaluate specific characteristics of individuals [22]. Evaluators may also exhibit central tendency, where ratings are limited to values near the midpoint of the rating scale, avoiding extreme ratings [23]. Grade inflation is also a well-recognized limitation in education literature. Evaluators across specialties may interpret survey questions differently, leading to inconsistencies and response bias. In addition to

a global assessment, interns were assessed by their PDs relative to their peers, introducing the possibility that evaluations were confounded by differences in the quality of interns in different specialties and at different hospitals. Ranking systems such as the *U.S. News & World Report* lack objective quality measurements [24], and do not provide data for all program specialty and subspecialties, and therefore were not included in our analysis.

Although our study found associations between both Step 1 and Step 2 CK and intern performance, with the introduction of the new score reporting system for Step 1, more of an emphasis may be placed on additional metrics such as GPA, class rank and Step 2 CK scores in the future and should be further evaluated. Additionally, although objective measures play a significant role in the evaluation of applicants for residency positions, the importance of subjective factors, such as baseline personal characteristics should be emphasized [25].

## Conclusions

Better performance as an intern was associated with higher USMLE scores, medical school GPA and class rank. When USMLE Step 1 reporting changes from numeric scores to pass/fail, residency programs can use other metrics to predict clinical performance. Although no single factor can be used to accurately predict performance in residency, these findings provide a framework for using objective metrics to select medical students for interview and employment, seeking candidates that will perform well as residents and become excellent physicians.

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12909-021-02487-0>.

**Additional file 1.** Appendix A: Program Director Evaluator Form.

## Abbreviations

USMLE: United States Medical Licensing Examination; GPA: Grade-point average; ACGME: Accreditation Council for Graduate Medical Education; PDs: Program directors

## Acknowledgements

Not applicable.

## Authors' contributions

A.C.F., L.A.C., T.J.L., and S.A.T. contributed to the literature review and study design. T.J.L., L.A.C., S.S.S. contributed to data collection and analysis. A.C.F., T.J.L., L.A.C., G.A.S. and S.A.T. interpreted the analysis. A.C.F. and T.J.L. drafted the manuscript. A.C.F., L.A.C., T.J.L., S.S.S., G.A.S., and S.A.T. provided critical revisions to the manuscript. All authors read and approved the final manuscript.

## Funding

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

## Availability of data and materials

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

## Ethics approval and consent to participate

The University of Florida Health Institutional Review Board approved this study (IRB #201802345). Informed consent was obtained. All methods were performed in accordance with the relevant guidelines and regulations.

## Consent for publication

Not applicable.

## Competing interests

The authors declare that they have no competing interests.

## Author details

<sup>1</sup>Department of Surgery, University of Florida Health, 1600 SW Archer Ave, PO Box 100109, Gainesville, FL 32610, USA. <sup>2</sup>Office for Educational Affairs, University of Florida College of Medicine, Gainesville, FL, USA. <sup>3</sup>University of Florida College of Medicine, Gainesville, FL, USA.

Received: 21 September 2020 Accepted: 14 December 2020

Published online: 26 January 2021

## References

- de Villiers M, van Heerden B, van Schalkwyk S. 'Going the extra mile': supervisors' perspectives on what makes a 'good' intern. *S Afr Med J*. 2018; 108:852–7.
- Eno CCR, Stewart NH, Lim J, Westerman ME, Holmboe ES, Edgar L. Milestones guidebook for residents and fellows. Accreditation Council for Graduate Medical Education (ACGME) [<https://www.acgme.org/Portals/0/PDFs/Milestones/MilestonesGuidebookforResidentsFellows.pdf>].
- Burish MJ, Fredericks CA, Engstrom JW, Tateo VL, Josephson SA. Predicting success: what medical student measures predict resident performance in neurology? *Clin Neurol Neurosurg*. 2015;135:69–72.
- Harfmann KL, Zirwas MJ. Can performance in medical school predict performance in residency? A compilation and review of correlative studies. *J Am Acad Dermatol*. 2011;65(5):1010–22 e1012.
- Sutton E, Richardson JD, Ziegler C, Bond J, Burke-Poole M, McMasters KM. Is USMLE step 1 score a valid predictor of success in surgical residency? *Am J Surg*. 2014;208(6):1029–34 discussion 1034.
- National resident matching program data release and research committee. Results of the 2018 NRMP Program Director Survey. [<https://www.nrmp.org/wp-content/uploads/2018/07/NRMP-2018-Program-Director-Survey-for-WWW.pdf>].
- Swanson WS, Harris MC, Master C, Gallagher PR, Mauro AE, Ludwig S. The impact of the interview in pediatric residency selection. *Ambul Pediatr*. 2005;5(4):216–20.
- Borowitz SM, Saulsbury FT, Wilson WG. Information collected during the residency match process does not predict clinical performance. *Arch Pediatr Adolesc Med*. 2000;154(3):256–60.
- Andriole DA, Jeffe DB, Whelan AJ. What predicts surgical internship performance? *Am J Surg*. 2004;188(2):161–4.
- Neely D, Feinglass J, Wallace WH. Developing a predictive model to assess applicants to an internal medicine residency. *J Grad Med Educ*. 2010;2(1): 129–32.
- United States Medical License Examination. USMLE program announces upcoming policy changes. [<https://www.usmle.org/announcements/>].
- Paolo AM, Bonaminio GA. Measuring outcomes of undergraduate medical education: residency directors' ratings of first-year residents. *Acad Med*. 2003;78(1):90–5.
- Alexander GL, Davis WK, Yan AC, Fantone JC. Following medical school graduates into practice: residency directors' assessments after the first year of residency. *Acad Med*. 2000;75(10 Suppl):S15–7.
- National Resident Matching Program. Results and data: 2019 main residency match [[https://mk0nrmp3oyqui6wqfm.kinstacdn.com/wp-content/uploads/2019/04/NRMP-Results-and-Data-2019\\_04112019\\_final.pdf](https://mk0nrmp3oyqui6wqfm.kinstacdn.com/wp-content/uploads/2019/04/NRMP-Results-and-Data-2019_04112019_final.pdf)].
- Gauer JL, Jackson JB. The association of USMLE step 1 and step 2 CK scores with residency match specialty and location. *Med Educ Online*. 2017;22(1): 1358579.

16. Association of American Medical Colleges. AAMC Table B1: test scores and experiences of first-year residents, by specialty [<https://www.aamc.org/data-reports/students-residents/interactive-data/table-b1-test-scores-and-experiences-first-year-residents-specialty>]. Accessed 1 March 2020.
17. McDonald FS, Zeger SL, Kolars JC. Associations between United States Medical Licensing Examination (USMLE) and Internal Medicine In-Training Examination (IM-ITE) scores. *J Gen Intern Med.* 2008;23(7):1016–9.
18. Spurlock DR, Holden C, Hartranft T. Using United States Medical Licensing Examination (®) (USMLE) examination results to predict later in-training examination performance among general surgery residents. *J Surg Educ.* 2010;67(6):452–6.
19. Thundiyil JG, Modica RF, Silvestri S, Papa L. Do United States Medical Licensing Examination (USMLE) scores predict in-training test performance for emergency medicine residents? *J Emerg Med.* 2010;38(1):65–9.
20. Fening K, Vander Horst A, Zirwas M. Correlation of USMLE step 1 scores with performance on dermatology in-training examinations. *J Am Acad Dermatol.* 2011;64(1):102–6.
21. Fryer JP, Corcoran N, George B, Wang E, Darosa D. Does resident ranking during recruitment accurately predict subsequent performance as a surgical resident? *J Surg Educ.* 2012;69(6):724–30.
22. Boet S, Sharma S, Goldman J, Reeves S. Review article: medical education research: an overview of methods. *Can J Anaesth.* 2012;59(2):159–70.
23. Williams RG, Klamen DA, McGaghie WC. Cognitive, social and environmental sources of bias in clinical performance ratings. *Teach Learn Med.* 2003;15(4):270–92.
24. Sehgal AR. The role of reputation in U.S. News & World Report's rankings of the top 50 American hospitals. *Ann Intern Med.* 2010;152(8):521–5.
25. Baimas-George M, Vrochides D. The sorting hat of medicine: why Hufflepuffs wear stethoscopes and Slytherins carry scalpels. *J Surg Educ.* 2020;77(4):772–8. <https://doi.org/10.1016/j.jsurg.2020.01.004>.

### Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

