



## AOA Critical Issues in Education

# Academic Faculty Demonstrate Weak Agreement in Evaluating Orthopaedic Surgery Residents

Taylor R. Bradley, BS, Cale A. Jacobs, PhD, ATC, and Ryan D. Muchow, MD, FAOA

*Investigation performed at University of Kentucky Department of Orthopedic Surgery and Sports Medicine, Lexington, Kentucky*

**Background:** As objective metrics fade, subjective elements of orthopaedic surgery applicants carry increasing importance during recruitment. Academic orthopaedic surgeons believe they can select for high-performing orthopaedic residents. However, can they agree? The purpose of this study was (1) to analyze an academic orthopaedic surgery department to determine whether they agree on which residents performed best during residency and; (2) to correlate preresidency and intraresidency factors with postresidency evaluations of resident performance.

**Methods:** With Institutional Review Board [IRB] approval, an orthopaedic surgery department completed surveys to evaluate residency performance for 10 years of graduated residents (2012-2022). Faculty determined (1) Post-Residency Class Rank (PRCR)—ranked from the highest (1) to lowest performing resident (5) relative to their class based on faculty perspective of performance—and (2) Rank List Score (RLS)—ranked based off the 5-point AOA SLOR used during recruitment interviews. RLS assessed how likely the department would have graduates in the program again based on residency performance. Free marginal Cohen's kappa statistics assessed faculty inter-rater agreement. Preresidency metrics (United States Medical Licensing Exam [USMLE] 1 and 2 scores, research publications, etc) were correlated with Orthopaedic In-Training Exam (OITE) scores, research productivity, American Board of Orthopaedic Surgery (ABOS) scores, and faculty-derived rankings. Linear regressions with forward variable entry ( $p < 0.05$ ) were used to determine factors associated with excellent resident performance.

**Results:** Eighteen of 25 faculty members (72%) provided survey responses evaluating 46 residents. Faculty agreed 37% and 38% of the time for PRCR (kappa 0.26) and RLS (kappa 0.23), respectively. Step 2 score was the only preresidency factor significantly associated with PRCR ( $p = 0.03$ ,  $r^2 = 0.15$ ) and RLS ( $p = 0.02$ ,  $r^2 = 0.3$ ). The only intraresidency factor significantly correlated with PRCR ( $p = 0.002$ ,  $r^2 = 0.50$ ) and RLS ( $p = 0.01$ ,  $r^2 = 0.39$ ) was PGY-4 OITE score.

**Conclusions:** An academic orthopaedic surgery department is able to come to a consensus on evaluations of residency performance relative to peers in the same year of training (PRCR) and an objective standard (RLS). Step 2 and Post-Graduate Year (PGY)-4 OITE scores were the only preresidency and intraresidency factors with significant association to higher postresidency, faculty-derived performance scores.

**Level of Evidence:** III.

**Disclosure:** The **Disclosure of Potential Conflicts of Interest** forms are provided with the online version of the article (<http://links.lww.com/JBJSOA/A574>).

Copyright © 2023 The Authors. Published by The Journal of Bone and Joint Surgery, Incorporated. All rights reserved. This is an open access article distributed under the terms of the [Creative Commons Attribution-Non Commercial-No Derivatives License 4.0](https://creativecommons.org/licenses/by-nc-nd/4.0/) (CCBY-NC-ND), where it is permissible to download and share the work provided it is properly cited. The work cannot be changed in any way or used commercially without permission from the journal.

## Introduction

Orthopaedic surgery remains one of the most competitive and challenging fields to attain residency match. According to National Resident Matching Program data, the average orthopaedic surgery match rate is around 70% over the past 8 years<sup>1-8</sup>. Factors such as research productivity, USMLE scores, volunteer and leadership activities, AOA membership, and medical school rank have all been associated with increased likelihood to match<sup>9-21</sup>. However, there is limited evidence that these factors predict successful performance during residency<sup>16,22,23</sup>.

The changing landscape in medical school education additionally affects recruitment efforts of orthopaedic surgery residency programs. Residencies have traditionally used objective metrics such as medical school grades or USMLE Step 1 scores to screen applicants<sup>16,24,25</sup>. The current shift toward pass/fail evaluations of knowledge transfers the screening process to more subjective variables in a student's application. While thoughtful arguments exist for whether this enhances the recruitment process, it potentially allows programs to identify application elements that more directly correlate with performance of an orthopaedic surgery resident, namely the subjective elements of performance such as professionalism, integrity, and character that drive a significant portion of a resident's measure of success<sup>26</sup>.

Faculty regularly assess resident skill and expertise in clinical and operative settings based on subjective qualities<sup>26</sup>. The most successful residents often distinguish themselves through enhanced patient care and interpersonal, professional, and communication skills. These subjective metrics often drive resident performance evaluations from attending orthopaedic surgeons<sup>27,28</sup>.

Ultimately, if entrance into orthopaedic surgery residency is increasingly gained using subjective measures and evaluation of resident performance is affected by faculty analysis of these soft skills, are subjective evaluations reliable and predictive? Advancement to fellowship, securing jobs, and even a portion of the ABOS certification rely significantly on attending evaluations of an individual. Therefore, is it possible for a group of academic orthopaedic faculty to agree on which residents were highest performing in residency? An evaluation among academic orthopaedic faculty to assess which residents were most successful would be of great value to the orthopaedic community at large. This study attempts to measure characteristics of resident performance that are challenging to quantify. In addition, correlation between preresidency and intraresidency factors that predict high performance in residency would substantially benefit orthopaedic surgery residency programs during recruitment. The purpose of this study was to (1) analyze a group of academic orthopaedic surgeons to determine whether they agree on which residents performed best during residency and (2) assess correlation between preresidency applicant factors, intraresidency markers of performance, and postresidency faculty-derived rankings.

## Materials and Methods

### Study Design

The project is an IRB-approved retrospective study of academic orthopaedic surgeons at a large, public university

hospital. A survey was completed by faculty within an academic orthopaedic surgery department to evaluate residency performance of prior residency graduates over the past 10 years (classes 2012-2013 to 2021-2022). The survey consisted of 2 faculty ratings, Post-Residency Class Rank (PRCR) and Rank List Score (RLS), depicted in Figure 1. For PRCR, faculty at our institution were asked to rerank each graduating class (best resident in the class = 1, lowest performance = 5) based on the faculty's individual perspective of resident performance relative to their respective class. For RLS, faculty were asked to score each graduate on a 5-point scale based off the AOA SLOR SRS that is used at our institution during recruitment interviews. The RLS scale is a measure of how likely the department would have the resident in the program again based on performance over the 5 years. Scoring is as follows: 5—Guaranteed Match, 4—Above Average Resident, 3—Average Resident, 2—Below Average Resident, and 1—Do Not Rank. The primary outcome of the survey was to assess whether orthopaedic surgeons can agree on which residents were highest performing when compared with peers of the same year of training and with an objective standard.

Secondary outcomes investigated whether there was a correlation between (1) preresidency factors and PRCR and RLS, (2) intraresidency factors and PRCR and RLS, and (3) preresidency factors and intraresidency metrics. Preresidency metrics included USMLE Step 1 and 2 scores, applicant interview scores, AOA distinction, final match rank, and research productivity defined by number of publications. Intraresidency metrics were OITE performance, research productivity, ABOS Part 1 and 2 score, and ABOS prior failing scores.

### Statistical Analyses

Faculty agreement on PRCR and RLS was assessed using free marginal Cohen's kappa statistics for analysis<sup>29</sup>. Linear regressions with forward variable entry ( $p < 0.05$ ) were used to determine preresidency and intraresidency factors associated with excellent resident performance.

### Source of Funding

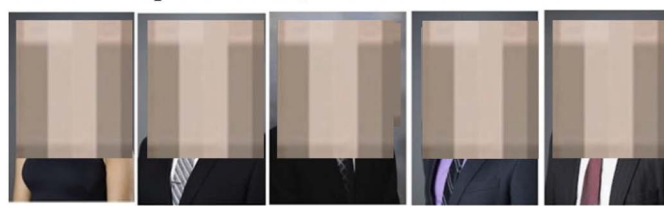
Funding was not provided for this study.

## Results

Of the 25 surveyed faculty, 18 (14/15 ACGME CORE faculty) provided evaluations of 45 total residents. There were 3 faculty who did not complete the survey because they had <5 years' experience at the institution (trauma, sports, and foot/ankle), 1 faculty abstained for personal reasons (hand), and 3 faculty did not complete the survey for unknown reasons (sports, shoulder/elbow, general/trauma). There were 4 pediatric orthopedic surgeons, 4 trauma, 3 total joint arthroplasty, 2 hand, 2 sports, 1 spine, and 1 MSK oncology surgeon that participated. There were 15 male (84%) and 3 female (16%) faculty, and 15 White and 3 Asian faculty members completed the survey.

**Post-Residency Class Rank**

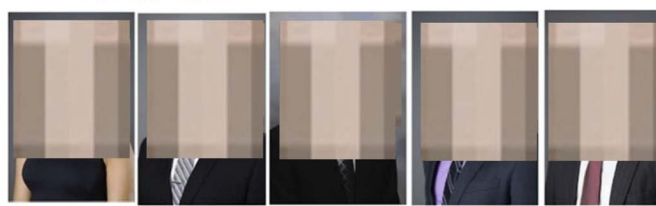
Please rank each graduate from highest to lowest performing within each class based on residency performance. Each resident should be evaluated relative to their residency class and given a unique score. Score: 1 = top of the class, 5= bottom of the class



	1	2	3	4	5
Resident 1	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Resident 2	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Resident 3	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Resident 4	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Resident 5	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

**Rank List Score**

5 Guaranteed Match at Our Program  
 4 Above Average Resident  
 3 Average Performing Resident  
 2 Below Average Resident  
 1 Do Not Rank



	1	2	3	4	5
Resident 1	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Resident 2	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Resident 3	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Resident 4	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Resident 5	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Fig. 1 Post-Residency Class Rank and Rank List Score surveys completed by academic faculty for 10 classes of graduated residents.

There were 41 male residents (89%) and 5 female residents (11%) evaluated, with a racial/ethnic breakdown consisting of 39 White, 2 African American, 2 Latino/Hispanic/Spanish origin, and 3 Asian residents. Their chosen specialties by fellowship were 9 spine, 9 sports, 7 peds, 6 trauma, 6 total joint arthroplasty, 4 foot/ankle, 3 hand, and 2 shoulder/elbow.

**Post-Residency Class Rank**

Using the mean PRCR score, faculty were able to create a rank order to distinguish highest to lowest performers within each graduating class of residents. There was considerable faculty variability for PRCR with an average 0.8 SD. Owing to the range of scores each resident received, faculty agreed only 37% of the time on the ranking order captured by PRCR with a 0.26 kappa statistic. Two residents attained unanimous agreement among faculty, which were scored for being the lowest performing residents in their respective class (Table I).

**Rank List Score**

The faculty-derived RLS differed from PRCR in that the residents were compared with a theoretical resident with standard grading on a 5-point scale. Thus, multiple residents in the same class could receive the same rating. The mean RLS for all resident graduates was 3.6, indicating that each resident was considered average performing and a good fit for the program based on residency performance compared

with the objective standard. Like PRCR, the RLS yielded significant faculty variability with an average SD of 0.8. Faculty registered 38% agreement with a 0.23 kappa statistic. There were no residents who yielded unanimous faculty agreement for RLS. The lowest SD achieved was 0.4, which interestingly was attained for 2 of the highest scoring residents (Table I).

**Relationship of Preresidency Factors to PRCR and RLS**

USMLE Step 2 score was the only preresidency factor with significant association to PRCR ( $p = 0.03$ ,  $r^2 = 0.15$ ) and significant correlation to RLS ( $p = 0.02$ ,  $r^2 = 0.3$ ). Preresidency research productivity (defined as number of peer-reviewed publications), AOA distinction, USMLE Step 1 scores, interview scores, and final match rank all had no correlation to PRCR and RLS ( $p > 0.05$ ). Fifty percent of the top-ranked residency applicants remained in the top 1 or 2 spots for postresidency rankings, yet preresidency rank list position had no correlation to PRCR ( $p = 0.43$ ) (Table II).

**Intraresidency and Postresidency Factors' Relationship to PRCR and RLS**

PGY-4 OITE score was the only intraresidency factor with significant correlation to PRCR ( $p = 0.002$ ,  $r^2 = 0.50$ ) and RLS ( $p = 0.01$ ,  $r^2 = 0.39$ ). Mean OITE score, research productivity, ABOS Part 1 score, and ABOS Part 1 pass rate did not yield significant correlation to PRCR and RLS ( $p > 0.05$ ) (Table II).

**TABLE I Faculty-Derived Post-Residency Class Rank and Rank List Score\***

Resident	Post-Residency Class Rank		Rank List Score	
A	1.4	0.6	4.4	1.2
B	2.1	0.8	3.9	1.0
C	4.9	0.3	1.6	0.6
D	2.6	0.8	3.5	1.0
E	3.9	0.4	2.5	1.0

\*Rank List Score and Post-Residency Class Rank (mean  $\pm$  SD) for each resident in one representative residency class (5 of 46 residents) is displayed with SD of faculty rankings.

### Negative and Positive Outliers

Five of 46 residents (10.9%) received greater than 4.4 RLS ratings, indicating that their performance earned them a “Guaranteed to Match” status if theoretically ranked again. Despite these results, no preresidency factors or pattern of factors that these residents possessed statistically predicted for high performance.

Three of 46 residents (6.5%) averaged RLS ratings lower than 2, indicating their residency performance earned them a “Do Not Rank” status. There were no preresidency factors that correlated with this postresidency performance ranking. In fact, all 3 residents with a RLS of less than 2 were ranked as highest, second highest, and third highest applicants of their matched class. Of these 3 residents, 2 also earned AOA status and performed well on their other preresidency metrics.

### Discussion

Matching into orthopaedic surgery residency continues to be incredibly challenging as evidenced by the most recent 2023 unmatched rate of nearly 40%<sup>1</sup>. Layered on top of this issue is the changing landscape whereby traditional objective application metrics associated with matching into orthopaedics are omitted altogether by medical institutions. Greater emphasis on soft skills in the areas of professionalism, integrity, compassion, and character has been proven to affect assessment in residency and beyond<sup>30-33</sup>. Although reliable tools exist to quantify resident surgical skills, there is currently no validated scale to evaluate soft skills desirable of the orthopaedic surgery resident<sup>34</sup>. Given the challenges of subjectivity, can a group of academic orthopaedic surgeons agree on which residents perform best? The results of this study demonstrate that PRCR and RLS have weak agreement among a group of academic faculty when evaluating residents over a 10-year period. These scoring tools allowed for differentiation of residents with reference to their graduating class and a theoretical “average performing resident” adapted from the AOA SLOR SRS tool. The utility of these ranking systems provides quantitative analysis of each individual's summative residency performance. Therefore, such ranking systems could serve as additive tools when residents are evaluated for fellowship, ABOS certification, and future employment.

The purpose of the RLS was to standardize individual resident performance to allow for applicability across institutions and time. The mean RLS was 3.6, indicating a value between the “Average” and “Above Average” designation, yet with significant variability leading to the 0.23 kappa. Faculty were able to unanimously agree on individuals who stood out for poor performance during their residency, although the tools used were not statistically predictive. Variability in scores may be expected given various faculty rankings of residents either a point above or below the mean based on individual perspective. The collective score could be a useful tool to illustrate how a resident interacts and performs among a group of faculty and is potentially predictive of future practice as a surgeon. Used in a similar fashion to the AOA SLOR SRS and its correlation to match success<sup>35</sup>, RLS also contains variability in the average score, yet could be a predictor of future success of the resident.

**TABLE II Relationship of Preresidency and Intraresidency Factors to Post-Residency Class Rank and Rank List Score**

Preresidency Factors		
	Post-Residency Class Rank	Rank List Score
Objective score	-0.16	0.30
Position on the rank list	0.18	-0.10
Interview score	0.01	0.20
Step 1 score	-0.25	0.18
Step 2 score	-0.20	0.38*
Publications	0.14	0.03
Intraresidency and Postresidency Factors		
	Post-Residency Class Rank	Rank List Score
OITE-1	-0.20	0.19
OITE-2	-0.31	0.26
OITE-3	-0.27	0.13
OITE-4	-0.28*	0.18*
OITE-5	-0.25	0.31
Publications	-0.13	0.23
Average OITE	-0.34	0.29
ABOS Part 1 score	-0.23	0.24
ABOS Part 1 percentage	-0.22	0.19
ABOS Part 1 prior fails	0.24	-0.04
ABOS 2 prior fails	0.13	-0.10

\*Indicates significant correlation ( $p < 0.05$ ). Displayed are calculated Pearson correlation values of preresidency, intraresidency, and postresidency factors to assess relationship to Post-Residency Class Rank and Rank List Score. Of note, objective score was a collective score for medical school grades, letters of recommendation, and research experiences.



The variability noted in the PRCR additionally supports the RLS mean. If a group of faculty ranked “average” to “above average” performing residency classes per the RLS tool, one may assume that variability would likewise be present for PRCR. Consistent with this reasoning, the greatest agreement among faculty was achieved on negative outliers—faculty agreed most on individuals who stood out for poor performance with increased disagreement among average and high performers. The weak agreement with a kappa of 0.26 is taken in the context of numerous average to above average performers and lends itself to the varied faculty opinions. The PRCR tool may serve as an adjunct to RLS in reinforcing high, mid, and low scores when paired with the class rank and agreement of faculty contributing to that rank.

The gold standard of residency performance evaluation would be to construct a metric that correlated preresidency factors with intraresidency performance and predicted for success after residency. In our secondary analysis, the only preresidency factor that had significant association with PRCR and RLS was USMLE Step 2 score, a finding consistent with Raman et al.<sup>16</sup>. However, clerkship grades, AOA distinction, and letters of recommendation did not generate statistical correlation in our study<sup>16,23</sup>. The only intraresidency factor to yield significant correlation to high overall resident performance was PGY-4 OITE score, a finding consistent with Spitzer et al.<sup>23</sup>. No preresidency or intraresidency factors correlated with ABOS Part 1 score or pass rate. Similarly, PRCR and RLS did not have any correlation with ABOS Part 1 score or pass rate, although this may be due to low ABOS Part 1 failure rates. Thus, our study was not able to link RLS or PRCR with any measurable markers of professional success.

There were several limitations to this study. First, while generally consistent year over year, we could not control for slight annual adjustments made to the department's algorithm to best capture preresident applicant scores. Although improvements enhanced the algorithm, data collection may include subtle differences in reported preresidency objective scores that could have influenced the study's results. Second,

the study is also limited by potential recall bias of the faculty who assessed residency performance for individuals who graduated from the program up to 10 years ago. In addition, future applicability of these study results may be limited secondary to USMLE Step 2 routinely being taken before applications are submitted compared with historically being taken at varied time points throughout the fourth year of medical school. Another limitation is the paucity of research on the accuracy and reliability of subjective assessments in quantifying orthopaedic resident performance. Many studies have used OITE scores to assess overall resident performance; however, few have investigated the merit and utility of scoring systems that quantify subjective characteristics that often drive performance evaluations<sup>15,27,36-42</sup>. In this study, we used the limited existing data as a scaffold in creation of PRCR. In addition, RLS was generated based off the AOA SLOR SRS used during our institution's applicant interviews, which warrants further research to validate its credibility.

In conclusion, an academic orthopaedic surgery department is able to come to a consensus on evaluations of residency performance relative to peers in the same year of training (PRCR) and an objective standard (RLS). Step 2 and PGY-4 OITE scores were the only preresidency and intraresidency factors with significant association to higher postresidency, faculty-derived performance scores. Therefore, Step 2 score may aid in and influence residency program recruitment. Additional research on correlation between RLS, PRCR, and future career success would be helpful to identify whether any predictive elements exist with these performance measures. ■

Taylor R. Bradley, BS<sup>1</sup>  
Cale A. Jacobs, PhD, ATC<sup>1</sup>  
Ryan D. Muchow, MD, FAOA<sup>1</sup>

<sup>1</sup>University of Kentucky Department of Orthopedic Surgery and Sports Medicine, Lexington, Kentucky

E-mail address for R. Muchow: ryan.muchow@uky.edu

## References

1. National Resident Matching Program. Advance Data Tables: 2023 Main Residency Match®. Washington, DC: National Resident Matching Program; 2023.
2. National Resident Matching Program. Charting Outcomes in the Match: Senior Students of U.S. Medical Schools, 2022. Washington, DC: National Resident Matching Program; 2022.
3. National Resident Matching Program. Results and Data: 2021 Main Residency Match®. Washington, DC: National Resident Matching Program; 2021.
4. National Resident Matching Program. Charting Outcomes in the Match: Senior Students of U.S. Medical Schools, 2020. Washington, DC: National Resident Matching Program; 2020.
5. National Resident Matching Program. Results and Data: 2019 Main Residency Match. Washington, DC: National Resident Matching Program; 2019.
6. National Resident Matching Program. Charting Outcomes in the Match: U.S. Allopathic Seniors, 2018. Washington, DC: National Resident Matching Program; 2018.
7. National Resident Matching Program. Results and Data: 2017 Main Residency Match®. Washington, DC: National Resident Matching Program; 2017.
8. National Resident Matching Program. Charting Outcomes in the Match for U.S. Allopathic Seniors, 2016. Washington, DC: National Resident Matching Program; 2016.
9. Smolev ET, Coxe FR, Iyer S, Kelly AM, Nguyen JT, Fufa DT. Orthopaedic surgery residency match after an early-exposure research program for medical students. *J Am Acad Orthop Surg.* 2021;5(9):e21.00113.
10. Li NY, Gruppuso PA, Kalagara S, Eltorai A, DePasse JM, Daniels AH. Critical assessment of the contemporary orthopaedic surgery residency application process. *J Bone Joint Surg Am.* 2019;101(21):e114.
11. Trikha R, Keswani A, Ishmael CR, Greig D, Kelley BV, Bernthal NM. Current trends in orthopaedic surgery residency applications and match rates. *J Bone Joint Surg Am.* 2020;102(6):e24.
12. Holderread BM, Liu J, Craft HK, Weiner BK, Harris JD, Liberman SR. Analysis of current orthopedic surgery residents and their prior medical education: does medical school ranking matter in orthopedic surgery match? *J Surg Educ.* 2022;79(4):1063-75.
13. Collier M, Janney C. Are orthopaedic surgery residency applicants making an impact? A bibliometric evaluation of applicants. *J Surg Orthop Adv.* 2021;30(3):150-5.
14. Adams AJ, Sherman M, Purtill JJ. Analyses of orthopaedic surgery residency interviews. *JB JS Open Access.* 2023;8(1):e22.00084.

15. Horan DP, Baldwin KM, Purtill JJ, Namdari S. Predictors of success in an orthopaedic residency. *JBJS Rev.* 2021;9(6):e20.00180.
16. Raman T, Alrabaa RG, Sood A, Maloof P, Benevenia J, Berberian W. Does residency selection criteria predict performance in orthopaedic surgery residency? *Clin Orthop Relat Res.* 2016;474(4):908-14.
17. Hallan DR, Mikhail D, Lu K, Henry A, Chiang K, Patterson M, Sakya SM. Residency specialty and national resident matching program outcomes as predictors of academic vs non-academic position as an attending physician. *Cureus.* 2020;12(8):e9548.
18. Geissler J, VanHeest A, Tatman P, Gioe T. Aggregate interview method of ranking orthopedic applicants predicts future performance. *Orthopedics.* 2013;36(7):e966-e970.
19. O'Mary K, Tompane T, Perry N, Smith J, Janney C. Matching into orthopedic surgery in the U.S. Navy: a review of applicant selection criteria. *Mil Med.* 2023;188(11-12):3641-5.
20. Huebner C, Adnan M, Kraeutler MJ, Brown S, Mulcahey MK. Use of the United States medical licensing examination step-1 score as a screening tool for orthopaedic surgery away rotations. *J Bone Joint Surg Am.* 2019;101(20):e106.
21. Schrock JB, Kraeutler MJ, Dayton MR, McCarty EC. A comparison of matched and unmatched orthopaedic surgery residency applicants from 2006 to 2014: data from the national resident matching program. *J Bone Joint Surg Am.* 2017;99(1):e1.
22. Spitzer AB, Gage MJ, Looze CA, Walsh M, Zuckerman JD, Egol KA. Factors associated with successful performance in an orthopaedic surgery residency. *J Bone Joint Surg Am.* 2009;91(11):2750-5.
23. Su CA, Furdock RJ, Rascoe AS, Vallier HA, Liu RW, Voos JE, Gillespie RJ. Which application factors are associated with outstanding performance in orthopaedic surgery residency? *Clin Orthop Relat Res.* 2023;481(2):387-96.
24. Chen AF, Secrist ES, Scannell BP, Patt JC. Matching in orthopaedic surgery. *J Am Acad Orthop Surg.* 2020;28(4):135-44.
25. Gu A, Farrar J, Fassih SC, Stake S, Ramamurti P, Wei C, Wessel LE, Fufa DT, Rao RD. Effect of change in USMLE Step 1 grading on orthopaedic surgery applicants: a survey of orthopaedic surgery residency program directors. *J Am Acad Orthop Surg.* 2021;5(5):e20.00216.
26. Rekman J, Gofton W, Dudek N, Gofton T, Hamstra SJ, Gofton T, Hamstra SJ. Entrustability scales: outlining their usefulness for competency-based clinical assessment. *Acad Med.* 2016;91(2):186-90.
27. Crawford CH III, Nyland J, Roberts CS, Johnson JR. Relationship among United States Medical Licensing Step I, orthopedic in-training, subjective clinical performance evaluations, and American Board of Orthopedic Surgery examination scores: a 12-year review of an orthopedic surgery residency program. *J Surg Educ.* 2010;67(2):71-8.
28. Brown E, Rosinski EF, Altman DF. Comparing medical school graduates who perform poorly in residency with graduates who perform well. *Acad Med.* 1993;68(10):806-8.
29. Warrens MJ. Inequalities between multi-rater kappas. *Adv Data Anal Classification.* 2010;4:271-86.
30. Teman NR, Minter RM, Kasten SJ. Utility of factor analysis in optimization of resident assessment and faculty evaluation. *Am J Surg.* 2016;211(6):1158-63.
31. Debets MPM, Scheepers RA, Boerebach BCM, Arah OA, Lombarts KMJMH. Variability of residents' ratings of faculty's teaching performance measured by five- and seven-point response scales. *BMC Med Educ.* 2020;20(1):325.
32. Colletti JE, Flottesmesch TJ, O'Connell TA, Ankel FK, Asplin BR. Developing a standardized faculty evaluation in an emergency medicine residency. *J Emerg Med.* 2010;39(5):662-8.
33. Mitchell M, Srinivasan M, West DC, Franks P, Keenan C, Henderson M, Wilkes M. Factors affecting resident performance: development of a theoretical model and a focused literature review. *Acad Med.* 2005;80(4):376-89.
34. Bagley JJ, Piazza B, Lazarus MD, Fox EJ, Zhan X. Resident training and the assessment of orthopaedic surgical skills. *JB JS Open Access.* 2021;6(4):e20.00173.
35. Samade R, Balch Samora J, Scharschmidt TJ, Goyal KS. Use of standardized letters of recommendation for orthopaedic surgery residency applications: a single-institution retrospective review. *J Bone Joint Surg Am.* 2020;102(4):e14.
36. Dyrstad BW, Pope D, Milbrandt JC, Beck RT, Weinhoefl AL, Idusuyi OB. Predictive measures of a resident's performance on written Orthopaedic Board scores. *Iowa Orthop J.* 2011;31:238-43.
37. Swanson D, Marsh JL, Hurwitz S, DeRosa GP, Holtzman K, Bucak SD, Baker A, Morrison C. Utility of AAOS OITE scores in predicting ABOS Part I outcomes: AAOS exhibit selection. *J Bone Joint Surg Am.* 2013;95(12):e84.
38. Ponce B, Savage J, Momaya A, Seales J, Oliver J, McGwin G, Theiss S. Association between orthopaedic in-training examination subsection scores and ABOS Part I examination performance. *South Med J.* 2014;107(12):746-50.
39. Van Heest AE, Armstrong AD, Bednar MS, Carpenter JE, Garvin KL, Harrast JJ, Martin DF, Murray PM, Peabody TD, Saltzman CL, Saniei M, Taitsman LA, Marsh JL. American board of orthopaedic surgery's initiatives toward competency-based education. *JB JS Open Access.* 2022;7(2):e21.00150.
40. Dougherty PJ, Walter N, Schilling P, Najibi S, Herkowitz H. Do scores of the USMLE Step 1 and OITE correlate with the ABOS Part I certifying examination? A multicenter study. *Clin Orthop Relat Res.* 2010;468(10):2797-802.
41. Pacana MJ, Thier ZT, Jackson JB III, Koon DE Jr, Grabowski G. More than one-third of orthopaedic applicants are in the top 10%: the standardized letter of recommendation and evaluation of orthopaedic resident applicants. *Clin Orthop Relat Res.* 2021;479(8):1703-8.
42. Claessen FMAP, Beks RB, Schol I, Dyer GS. What predicts outstanding orthopedic residents among the program? *Arch Bone Joint Surg.* 2019;7(6):478-83.