

# Are USMLE Scores Valid Measures for Chief Resident Selection?

Elaine R. Cohen, MEd  
 Joshua L. Goldstein, MD  
 Clara J. Schroedl, MD, MS

Nancy Parlapiano, BA  
 William C. McGaghie, PhD  
 Diane B. Wayne, MD

## ABSTRACT

**Background** The US Medical Licensing Examination (USMLE) Step 1 and Step 2 scores are often used to inform a variety of secondary medical career decisions, such as residency selection, despite the lack of validity evidence supporting their use in these contexts.

**Objective** We compared USMLE scores between non-chief residents (non-CRs) and chief residents (CRs), selected based on performance during training, at a US academic medical center that sponsors a variety of graduate medical education programs.

**Methods** This was a retrospective cohort study of residents' USMLE Step 1 and Step 2 Clinical Knowledge (CK) scores from 2015 to 2020. The authors used archived data to compare USMLE Step 1 and Step 2 CK scores between non-CR residents in each of the eligible programs and their CRs during the 6-year study period.

**Results** Thirteen programs enrolled a total of 1334 non-CRs and 211 CRs over the study period. There were no significant differences overall between non-CRs and CRs average USMLE Step 1 ( $239.81 \pm 14.35$  versus  $240.86 \pm 14.31$ ;  $P = .32$ ) or Step 2 scores ( $251.06 \pm 13.80$  versus  $252.51 \pm 14.21$ ;  $P = .16$ ).

**Conclusions** There was no link between USMLE Step 1 and Step 2 CK scores and CR selection across multiple clinical specialties over a 6-year period. Reliance on USMLE Step 1 and 2 scores to predict success in residency as measured by CR selection is not recommended.

## Introduction

The US Medical Licensing Examination (USMLE) Step 1 and Step 2 are designed as norm-referenced measures of acquired knowledge intended to inform medical licensure decisions.<sup>1</sup> However, USMLE Step 1 and Step 2 scores are also used to inform a variety of secondary medical career decisions, predominantly postgraduate residency selection by program directors, with almost two-thirds of postgraduate medical education programs reporting use of minimum cutoff scores.<sup>2,3</sup> Such secondary use may stem from convenience, the ability to compare applicants across schools,<sup>4</sup> and the belief that USMLE performance is correlated with success in residency. The medical education community recognizes that performance on USMLE Step 1 evaluates medical knowledge and predicts performance on other standardized tests, such as in-training and certification examinations.<sup>5-10</sup> However, there are few data to suggest that USMLE Step 1 and 2 scores predict performance across other critical competency domains, such as patient care, communication, and professionalism.<sup>11-14</sup>

The Association of American Medical Colleges, American Medical Association, Educational Commission for Foreign Medical Graduates, Federation of

State Medical Boards, and National Board of Medical Examiners convened a multistakeholder Invitational Conference on USMLE Scoring (InCUS) in March 2019 to address issues concerning USMLE score reporting, primary and secondary uses, and the transition from undergraduate medical education to graduate medical education (GME).<sup>15</sup> InCUS addressed such themes as the unintended consequences of USMLE Step 1 scoring, including its impact on medical student well-being and concerns about racial differences in USMLE performance. After the conference, a summary statement and a list of recommendations were published, including the need to decrease the emphasis on USMLE scores in residency selection and to pursue additional research on the relationship of USMLE scores with actual measures of residency performance.<sup>15</sup> In February 2020, the NBME announced that USMLE Step 1 scoring would be changed to pass/fail no earlier than January 2022.<sup>15,16</sup>

Selection as chief resident (CR) by program faculty or peers is a marker of exceptional achievement during residency education.<sup>17,18</sup> Medical historian Kenneth Ludmerer, MD, asserts that such recognition is the "crown jewel of graduate medical education."<sup>19</sup> While little published literature is available that reports how chief residents are selected in a wide range of programs, anecdotal information suggests

DOI: <http://dx.doi.org/10.4300/JGME-D-19-00782.1>

that these positions are offered to residents with exceptional leadership, clinical, and academic skills. These residents are likely highly regarded by their peers, coworkers, and program administration. Therefore, we used chief resident status as a proxy for a highly successful resident. Assuming this is the kind of candidate a program director would want to select when reviewing applications, the aim of this study was to compare USMLE scores between non-CRs and CRs at a US academic medical center that sponsors a wide range of GME programs.

## Methods

### Setting and Participants

This is a retrospective cohort study<sup>20</sup> of Step 1 and Step 2 Clinical Knowledge (CK) scores of residents in their final year of training conducted at the McGaw Medical Center of Northwestern University from 2015 to 2020. The McGaw Medical Center is a consortium of urban, suburban, specialized, and general hospitals sponsoring GME programs in metropolitan Chicago. McGaw currently has 26 residency programs accredited by the Accreditation Council for Graduate Medical Education. For this study, programs (and their residents) were included if (1) the training programs operated throughout the study period, and (2) the programs designated some, but not all, of their upper-level residents to be CRs. For example, programs in which all of the final-year residents were given the title of CR were excluded. All data were collected as part of routine operational activities, and results were reported in aggregate with no individual examinee identified.

### Outcome Measures

We used data archived in the McGaw GME office to compare USMLE Step 1 and Step 2 CK scores between non-CR residents in each of the eligible programs and their CRs during the 6-year study period. Each resident was included only once for the analysis; although, he or she may have completed multiple years of training. All of the program directors from the eligible programs agreed with the statement that their CRs were selected based on academic and clinical performance.

The study was deemed exempt from review by the Northwestern University Institutional Review Board.

### Analysis

In 2017, the average USMLE scores for first takers from the United States and Canada were 229 (SD = 20) for Step 1 and 242 (SD = 17) for Step 2 CK.<sup>21</sup> If we hypothesized an estimated 5-point

#### What was known and gap

The US Medical Licensing Examination (USMLE) Step 1 and Step 2 are designed as measures of acquired knowledge intended to inform medical licensure decisions, but they are often used to inform a variety of secondary medical career decisions, for which they lack validity evidence, such as chief resident selection.

#### What is new

A comparison of USMLE scores between non-chief residents and chief residents at a US academic medical center.

#### Limitations

Study was conducted at a single institution over a 6-year period with one method of selecting chief residents, limiting generalizability.

#### Bottom line

There was no link between USMLE Step 1 and Step 2 Clinical Knowledge scores and chief resident selection across multiple clinical specialties over a 6-year period.

difference between non-CRs (based on the national averages) and CRs, a sample size of 1356 (1159 non-CRs, 197 CRs) has 90% power to detect a difference between a non-CR Step 1 score of 229 and a CR Step 1 score of 234, assuming a common SD of 20 using an independent 2-sided *t* test with an  $\alpha$  of .05. Similarly, for Step 2, a sample size of 978 (836 non-CRs, 142 CRs) has 90% power to detect a difference between a non-CR Step 2 CK score of 242 and a CR Step 2 CK score of 247, assuming a common SD of 17 using an independent 2-sided *t* test with an  $\alpha$  of .05. This 5-point difference was meant as a conservative estimate for sample size. Given data from the USMLE website,<sup>21</sup> the 5-point difference approximately represents the difference between the 47th and 56th percentiles for Step 1 and the 42nd and 54th percentiles for Step 2. A larger point differential would require a smaller sample size in order to detect differences. We compared overall mean USMLE Step 1 and 2 scores between non-CRs and CRs using independent samples *t* tests. Multiple mean comparisons were also made by training programs using independent samples *t* tests with a Bonferroni correction. All statistical analyses were performed using Stata 14 (StataCorp LLC, College Station, Texas).

## Results

Thirteen of 26 current McGaw residency programs (57%) and their trainees qualified to be part of the analysis (TABLE 1). Five programs (family medicine [2], interventional radiology, thoracic surgery–integrated, and vascular surgery–integrated) were excluded because they did not exist for the entire study period. The remaining 8 programs (child neurology, family medicine, neurological surgery, ophthalmology, otolaryngology, plastic surgery, radiation oncology, and

**TABLE 1**  
Residency Program Characteristics

Residency Program	No. of Non-CR Residents	No. of CRs	No. of US Medical Schools	No. of International Medical Schools
Anesthesiology	163	19	69	0
Dermatology	16	15	24	1
Diagnostic radiology	73	12	45	4
Emergency medicine	108	23	63	1
General surgery	64	12	48	2
Internal medicine	292	26	70	3
Neurology	43	12	28	4
Obstetrics and gynecology	92	18	54	2
Orthopaedic surgery	56	14	39	0
Pathology	44	12	35	5
Pediatrics	241	18	90	1
Physical medicine and rehabilitation	82	18	56	1
Psychiatry	60	12	44	0
Total	1334	211	144	23

Abbreviation: CR, chief resident.

urology) were excluded because program leadership did not select CRs or used a hybrid approach. In these programs, each of the graduating residents served as a CR for part of his or her final year.

Overall, McGaw programs enrolled a total of 1334 non-CRs and 210 CRs over the 6-year study period. These residents were drawn from 144 unique US medical schools and 23 international medical schools. Residency program size ranged from 11 (dermatology) to 118 (internal medicine) trainees annually. The number of CRs ranged from 2 to 5 per year.

USMLE Step 1 scores ranged from 185 to 273 for non-CRs and from 200 to 274 for CRs. USMLE Step 2 CK scores ranged from 199 to 286 for non-CRs and from 198 to 283 for CRs. The figure shows the spread, quartiles, and outliers of all non-CR and CR USMLE Step 1 and Step 2 CK scores. There were no significant differences overall between non-CRs' and CRs' average USMLE Step 1 ( $239.81 \pm 14.35$  versus  $240.86 \pm 14.31$ ;  $P = .32$ ) or Step 2 CK ( $251.06 \pm 13.80$  versus  $252.51 \pm 14.21$ ;  $P = .16$ ) scores.

USMLE comparisons within each program revealed similar results. Using Bonferroni adjusted  $\alpha$  levels of .0019 per test (.05/26 comparisons), scores between groups were similar overall (TABLE 2).

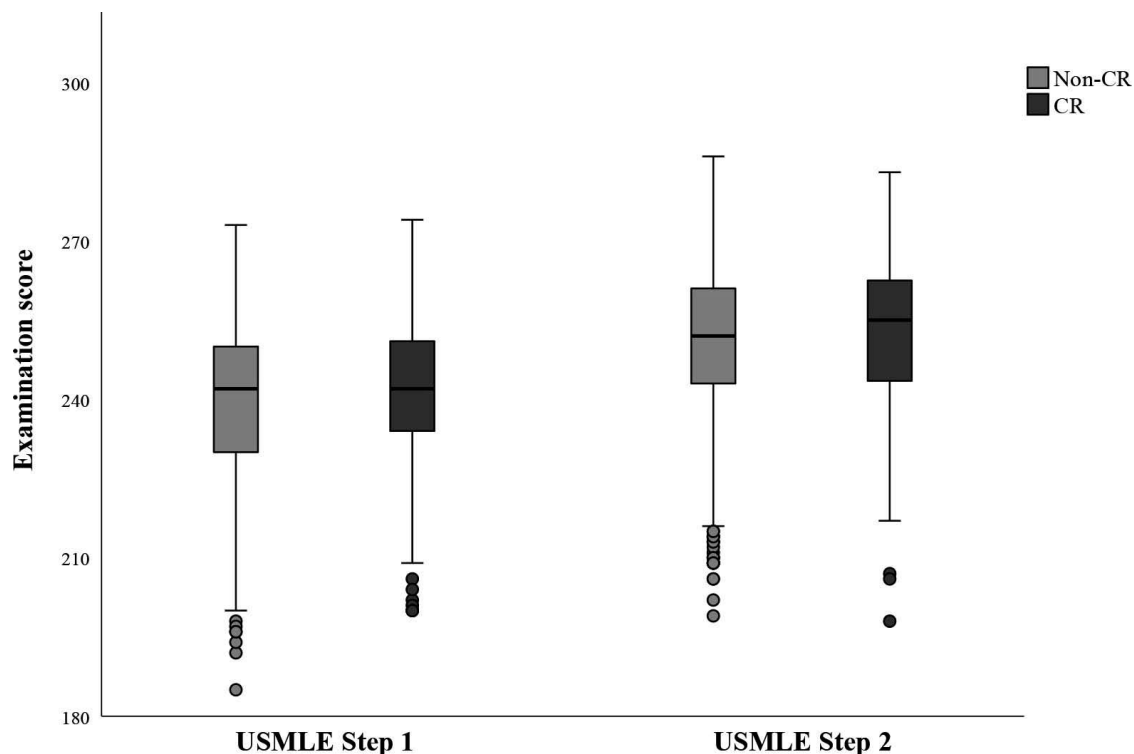
## Discussion

This study shows that there was no link between USMLE Step 1 and Step 2 CK scores, and selection as a CR across multiple clinical specialties over a 6-year period. Because of the lack of a link between USMLE scores and CR selection, this study confirms findings

from earlier research<sup>11</sup> that standardized examination scores are not a proxy for excellence in competencies such as communication and patient care.

Overreliance on USMLE scores in the residency selection process is the result of several factors. First, these standardized examination scores can be compared between applicants, while grades and rankings from various schools are challenging to compare. Additional factors include an increased number of medical students in the residency Match and the fact that applicants are applying to more programs than in prior years.<sup>22</sup> Even with the future change in scoring of Step 1 to pass/fail,<sup>16</sup> changing these patterns will not be easy. We believe that a good first step would be to develop an examination that measures the qualities educators seek in residency applicants rather than the ability to memorize facts. One possibility is to consider a score for the Step 2 Clinical Skills examination that measures a variety of skills, such as history taking, communication and interpersonal skills, and diagnostic reasoning. Validity evidence for this approach exists; a large study of internal medicine residents showed that Step 2 Clinical Skills communication and interpersonal skills scores predicted performance during the first year of training.<sup>23</sup>

This study has several limitations. First, it was conducted at a single institution over a 6-year period with 1 method of selecting CRs. These findings should be assessed at other centers of GME to determine generalizability. Not all residency programs were included, although the programs in our study reflect the results of 70% of US medical school seniors in the



**FIGURE**  
Box Plot of US Medical Licensing Examination (USMLE) Step 1 and 2 Score Variation, Quartiles (Medians), and Outliers for Chief Resident (CR) and Non-CR Groups

**TABLE 2**  
Comparison of USMLE Scores Between Non-CRs and CRs Across 13 Residency Programs<sup>a</sup>

Residency Program	USMLE Step 1 Mean (SD)			USMLE Step 2 Mean (SD)		
	Non-CR Residents	CRs	P Value	Non-CR Residents	CRs	P Value
Anesthesiology	239.39 (11.67)	241.32 (9.36)	.49	246.55 (12.93)	255.26 (14.77)	.007
Dermatology	248.75 (12.34)	250.13 (7.55)	.71	257.81 (12.46)	261.27 (9.17)	.39
Diagnostic radiology	244.16 (12.48)	248.25 (13.47)	.30	248.93 (15.59)	252.25 (12.60)	.49
Emergency medicine	241.03 (11.87)	241.78 (11.65)	.78	255.30 (12.22)	252.35 (11.60)	.29
General surgery	242.61 (15.06)	248.33 (7.92)	.21	250.48 (15.12)	255.75 (5.72)	.24
Internal medicine	243.75 (13.45)	244.46 (10.50)	.79	254.60 (13.56)	260.35 (8.82)	.035
Neurology	236.16 (14.60)	233.00 (15.66)	.52	246.05 (13.17)	247.50 (14.41)	.74
Obstetrics and gynecology	240.98 (12.53)	247.28 (11.71)	.05	255.39 (11.12)	260.56 (10.91)	.07
Orthopaedic surgery	244.02 (12.95)	243.29 (13.36)	.85	254.82 (11.83)	247.64 (19.35)	.08
Pathology	231.59 (17.01)	235.58 (15.05)	.46	239.19 (15.81)	241.33 (11.56)	.67
Pediatrics	237.79 (14.63)	236.28 (18.90)	.68	252.59 (11.59)	254.67 (10.07)	.46
Physical medicine and rehabilitation	232.57 (16.34)	28.25 (15.12)	.33	245.29 (13.93)	242.40 (17.65)	.49
Psychiatry	229.18 (15.81)	228.50 (16.05)	.89	242.45 (13.79)	236.83 (13.80)	.20

Abbreviations: CR, chief resident; USMLE, United States Medical Licensing Examination.

<sup>a</sup> Using Bonferroni adjusted alpha levels of .0019 per test (.05/26 comparisons).

2020 Match.<sup>24</sup> Additionally, we acknowledge that Northwestern residents performed higher than the average of all first-time USMLE test takers on Steps 1 and 2 CK. It is possible that a threshold score might differentiate residency performance, but that our residents exceeded this score. It is also not known if additional residents were approached to serve as CRs and declined. Although this is conceivable, the residents in this study were high performing, and no differences between these residents' and their peers' USMLE Step scores were found overall. Finally, we used selection for CR as an outcome measure and did not consider other variables, such as quality of clinical care provided, performance in simulated procedures and skills, and evaluations from clinical team members, such as nurses and patients. Further study is needed to assess these variables and their association with USMLE Step scores.

## Conclusions

USMLE scores were not associated with selection as CR across 13 programs over a 6-year period. Reliance on USMLE scores to predict markers of success in residency, such as CR selection, is not recommended due to the findings of this study.

## References

1. US Medical Licensing Examination. 2019 USMLE bulletin of information. <https://www.usmle.org/pdfs/bulletin/2019bulletin.pdf>. Accessed March 23, 2020.
2. Green M, Jones P, Thomas JX Jr. Selection criteria for residency: results of a national program director's survey. *Acad Med*. 2009;84(3):362–367. doi:10.1097/ACM.0b013e3181970c6b.
3. National Resident Matching Program. Results of the 2018 NRMP Program Director Survey. <http://www.nrmp.org/main-residency-match-data/>. Accessed March 23, 2020.
4. Prober CG, Kolars JC, First LR, Melnick DE. A plea to reassess the role of United States Medical Licensing Examination Step 1 scores in residency selection. *Acad Med*. 2016;91(1):12–15. doi:10.1097/ACM.0000000000000855.
5. Armstrong A, Alvero R, Neilsen P, Deering S, Robinson R, Frattarelli J, et al. Do U.S. Medical Licensure Examination Step 1 scores correlate with Council on Resident Education in Obstetrics and Gynecology in-training examination scores and American Board of Obstetrics and Gynecology written examination performance? *Mil Med*. 2007;172(6):640–643. doi:10.7205/milmed.172.6.640.
6. Nagasawa DT, Beckett JS, Lagman C, Chung LK, Schmidt B, Safae M, et al. United States Medical Licensing Examination Step 1 scores directly correlate with American Board of Neurological surgery scores: a single-institution experience. *World Neurosurg*. 2017;98:427–431. doi:10.1016/j.wneu.2016.11.091.
7. Welch TR, Olson BG, Nelsen E, Beck Dallaghan GL, Kennedy GA, Botash A. United States Medical Licensing Examination and American Board of Pediatrics certification examination results: does the residency program contribute to trainee achievement. *J Pediatr*. 2017;188:270–274.e3. doi:10.1016/j.jpeds.2017.05.057.
8. Caffery T, Fredette J, Musso MW, Jones GN. Predicting American Board of Emergency Medicine Qualifying Examination passage using United States Medical Licensing Examination step scores. *Ochsner J*. 2018;18(3):204–208. doi:10.31486/toj.17.0101.
9. Picarsic J, Raval JS, Macpherson T. United States Medical Licensing Examination Step 1 two-digit score: a correlation with the American Board of Pathology first-time test taker pass/fail rate at the University of Pittsburgh Medical Center. *Arch Pathol Lab Med*. 2011;135(10):1349–1352. doi:10.5858/arpa.2010-0240-EP.
10. Harmouche E, Goyal N, Pinawin A, Nagarwala J, Bhat R. USMLE scores predict success in ABEM initial certification: a multicenter study. *West J Emerg Med*. 2017;18(3):544–549. doi:10.5811/westjem.2016.12.32478.
11. McGaghie WC, Cohen ER, Wayne DB. Are United States Medical Licensing Exam Step 1 and Step 2 scores valid measures for postgraduate medical residency selection decisions? *Acad Med*. 2011;86(1):48–52. doi:10.1097/ACM.0b013e3181ffacdb.
12. Rifkin WD, Rifkin A. Correlation between housestaff performance on the United States Medical Licensing Examination and standardized patient encounters. *Mt Sinai J Med*. 2005;72(1):47–49.
13. Husain A, Li I, Ardolic B, Bond MC, Shoenberger J, Shah KH, et al. The standardized video interview: how does it affect the likelihood to invite for a residency interview? *AEM Educ Train*. 2019;3(3):226–232. doi:10.1002/aet2.10331.
14. Rutz M, Turner J, Pettit K, Palmer MM, Perkins A, Cooper DD. Factors that contribute to resident teaching effectiveness. *Cureus*. 2019;11(3):e4290. doi:10.7759/cureus.4290.
15. Barone MA, Filak AT, Johnson D, Skochelak S, Whelan A. Summary report and preliminary recommendations from the Invitational Conference on USMLE Scoring (InCUS). [https://www.usmle.org/pdfs/incus/incus\\_summary\\_report.pdf](https://www.usmle.org/pdfs/incus/incus_summary_report.pdf). Accessed March 23, 2020.
16. US Medical Licensing Examination. Change to pass/fail score reporting for Step 1. <https://www.usmle.org/inCus/#decision>. Accessed March 23, 2020.

17. Department of Medicine, Duke University School of Medicine. Internal medicine residency. <https://medicine.duke.edu/education-and-training/internal-medicine-residency/>. Accessed March 23, 2020.
18. Yale School of Medicine. Choosing our chief residents. <https://medicine.yale.edu/news-article/14043/>. Accessed March 23, 2020.
19. Ludmerer KM. *Time to Heal: American Medical Education from the Turn of the Century to the Era of Managed Care*. New York, NY: Oxford University Press; 1999.
20. Shadish WR, Cook TD, Campbell DT. *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Boston, MA: Houghton Mifflin; 2001.
21. US Medical Licensing Examination. USMLE score interpretation guidelines. [https://www.usmle.org/pdfs/transcripts/USMLE\\_Step\\_Examination\\_Score\\_ Interpretation\\_Guidelines.pdf](https://www.usmle.org/pdfs/transcripts/USMLE_Step_Examination_Score_ Interpretation_Guidelines.pdf). Accessed March 23, 2020.
22. Weiner S. Should the USMLE be pass/fail? <https://news.aamc.org/medical-education/article/should-the-usmle-be-pass-fail/>. Accessed March 23, 2020.
23. Winward ML, Lipner RS, Johnston MM, Cuddy MM, Clauser BE. The relationship between communication scores from the USMLE Step 2 Clinical Skills Examination and communication ratings for first-year internal medicine residents. *Acad Med*. 2013;88(5):693–698. doi:10.1097/ACM.0b013e31828b2df1.
24. National Resident Matching Program. Results and Data: 2020 Main Residency Match. <https://>

mk0nrmp3oyqui6wqfm.kinstacdn.com/wp-content/uploads/2020/05/MM\_Results\_and-Data\_2020.pdf. Accessed May 12, 2020.



**Elaine R. Cohen, MEd**, is Research Associate, Department of Medicine, Northwestern University Feinberg School of Medicine; **Joshua L. Goldstein, MD**, is Associate Dean, Graduate Medical Education, and Associate Professor of Pediatrics, Medical Education and Neurology, Northwestern University Feinberg School of Medicine, and Designated Institutional Official, McGaw Medical Center of Northwestern University; **Clara J. Schroedl, MD, MS**, is Assistant Professor of Medicine and Medical Education, Northwestern University Feinberg School of Medicine, and Director of Institutional Accreditation, McGaw Medical Center of Northwestern University; **Nancy Parlapiano, BA**, is Executive Director, McGaw Medical Center of Northwestern University; **William C. McGaghie, PhD**, is Professor of Medical Education and Preventive Medicine, Northwestern University Feinberg School of Medicine; and **Diane B. Wayne, MD**, is Vice Dean for Education and Dr. John Sherman Appleman Professor of Medicine and Medical Education, Northwestern University Feinberg School of Medicine, and President, McGaw Medical Center of Northwestern University.

**Funding:** The authors report no external funding source for this study.

**Conflict of interest:** The authors declare they have no competing interests.

The authors would like to thank all of the dedicated residents at the McGaw Medical Center of Northwestern University.

**Corresponding author:** Elaine R. Cohen, MEd, Northwestern University Feinberg School of Medicine, Division of Hospital Medicine, 211 East Ontario Street, 7th Floor, Chicago, IL 60611, 312.926.5909, elaine.cohen@northwestern.edu

Received November 5, 2019; revision received January 31, 2020; accepted February 29, 2020.