

Internal medicine resident perspectives on scoring USMLE as pass/fail

Sara L. Wallach^{a,b,c}, Christopher Williams^d, Robert T. Chow^{e,f}, Nagesh Jadhav^{g,h}, Sapna Kuehl^{f,i}, Jaya M. Raji and Richard Alweis^{b,h,k,l}

^aDepartment of Medicine, St. Francis Medical Center, Trenton, NJ, USA; ^bDepartment of Medicine, Hackensack Meridian School of Medicine, Nutley, NJ, USA; ^cMedicine, Drexel University College of Medicine, Philadelphia, PA, USA; ^dDepartment of Behavioral and Community Health, School of Public Health, University of Maryland, College Park, MD, USA; ^eDepartment of Medicine, The University of Maryland Medical Center Midtown Campus, Baltimore, MD, USA; ^fMedicine, University of Maryland School of Medicine, Baltimore, MD, USA; ^gDepartment of Medicine, Rochester General Hospital Internal Medicine Residency Program, Rochester, NY, USA; ^hMedicine, University of Rochester School of Medicine and Dentistry, Rochester, NY, USA; ⁱDepartment of Medicine, Ascension Saint Agnes Hospital, Baltimore, MD, USA; ^jDepartment of Medicine, Creighton University School of Medicine-Phoenix, St. Joseph's Hospital and Medical Center, Phoenix, AZ, USA; ^kEducation, Rochester Regional Health, Rochester, NY, USA; ^lHealth Sciences, Rochester Institute of Technology, Rochester, NY, USA

ABSTRACT

Background: The scoring rubric on the USMLE Step 1 examination will be changing to pass/fail in January 2022. This study elicits internal medicine resident perspectives on USMLE pass/fail scoring at the national level.

Objective: To assess internal medicine resident opinions regarding USMLE pass/fail scoring and examine how variables such as gender, scores on USMLE 1 and 2, PGY status and type of medical school are associated with these results.

Methods: In the fall of 2019, the authors surveyed current internal medicine residents via an on-line tool distributed through their program directors. Respondents indicated their Step 1 and Step 2 Clinical Knowledge scores from five categorical ranges. Questions on medical school type, year of training year, and gender were included. The results were analyzed utilizing Pearson Chi-square testing and multivariable logistic regression.

Results: 4012 residents responded, reflecting 13% of internal medicine residents currently training in the USA. Fifty-five percent of respondents disagreed/strongly disagreed with pass/fail scoring and 34% agreed/strongly agreed. Group-based differences were significant for gender, PGY level, Step 1 score, and medical school type; a higher percentage of males, those training at the PGY1 level, and graduates of international medical schools (IMGs) disagreed with pass/fail reporting. In addition, high scorers on Step 1 were more likely to disagree with pass/fail reporting than low scoring residents

Conclusion: Our results suggest that a majority of internal medicine residents, currently training in the USA prefer that USMLE numerical scoring is retained and not changed to pass/fail.

ARTICLE HISTORY

Received 15 June 2020
Accepted 13 July 2020

KEYWORDS

Graduate medical education; recruitment; United States medical licensing examination scoring

1. Introduction

The USA Medical Licensure Examination (USMLE), a three-step standardized examination administered by the Federation of State Medical Boards (FSMB) and the National Board of Medical Examiners (NBME), is one of the key components used by state medical boards in the USA to assess candidacy for licensure [1]. Since its inception, many stakeholders in medical education have adapted the examination for their own purposes and needs [2,3]. Medical schools have increasingly forgone reporting grades and GPA, especially in the preclinical years, and shifted to using USMLE for evaluation of individual student performance [3] as well as the assessment of the overall medical school curriculum [2,4–6]. Additionally, in the last decade, graduate medical education (GME) training programs in the US have seen an exponential rise in applications without a commensurate increase in staffing or resources, leading many to

increasingly utilize USMLE results as objective and standardized measures of the medical knowledge of applicants [2,7]. USMLE scores are often used as an initial screen to differentiate amongst applicants from an increasingly diverse undergraduate medical education training milieu [2,8]. Critics cite a number of problems in this extrapolated use of the examination, particularly the USMLE Step 1 examination: lack of correlation with physicians' clinical abilities, emphasis on the importance of these examinations in medical school curricula at the expense of other essential competencies, emphasis by GME programs on scores over other metrics, and potential cultural bias [9]. Amidst considerable controversy in the value of reporting USMLE scores, in August 2019, the decision was made to change Step 1 to a pass/fail score rather than the traditional three-digit numeric score [10].

Internal medicine (IM) constitutes one-quarter of the total number of active residents and fellows among

specialties accredited by the Accreditation Council for Graduate Medical Education [11]. Internal medicine trainees represent the largest share of US allopathic, osteopathic, and international medical school graduates [12]. Although IM has 15,000 more residents than the second most populous specialty, little is known about the perspectives of IM residents concerning changes to USA Medical Licensing Examination (USMLE) reporting. To our knowledge, no study has systematically surveyed current IM trainees to understand perspectives on USMLE Step 1 policy changes. We sought to assess support for pass/fail reporting, attitudes regarding the value of the examination, and group-based differences due to gender, post-graduate year (PGY) level, or type of UME training.

2. Methods

The authors adapted, with permission, a previously published survey instrument that assesses medical student and resident opinions on pass/fail USMLE scoring [13]. In our study, eight questions, using a five-item Likert scale of agreement (Appendix A – Survey), were administered to residents training in internal medicine in the fall of 2019. Respondents indicated their Step 1 and Step 2 Clinical Knowledge scores from five categorical ranges. Two open-ended questions gathered feedback on the pass/fail scoring schema for Steps 1 and 2. Finally, questions on medical school type, year of training, and gender assessed representativeness.

The authors solicited resident survey participation using convenience sampling by posting on the single, largest discussion forum of internal medicine program directors, using a survey methodology that has been frequently cited [14,15]. The e-mail invitation included an anonymous hyperlink to the web-based survey (Survey Monkey, San Mateo, CA). An initial request was sent in October 2019, followed by five follow-up requests. The survey closed in December 2019. The Institutional Review Board of St. Francis Medical Center, Trenton, NJ approved this study.

3. Quantitative Data Analysis

Frequencies and percentages were used for descriptive statistics. Only PGY1 through PGY 3 resident responses were included in the analysis. Responses were not included in the analysis if the responder did not identify their PGY status or identified as PGY4. Likert scales were collapsed to three-item scales for analysis. The Pearson chi-square test was used to report P values for group-based differences between categorical variables. Population estimates for post-graduate year (PGY) level were available from the Accreditation Council for Graduate Medical Education (ACGME), and gender and medical school type were publicly available from the American Board of Internal Medicine [11,16]. Post-hoc comparisons were performed with Bonferroni correction for each combination of independent categories to control for type I error inflation when the Pearson chi-square comparisons were significant. We evaluated the probability values for each categorical value against the adjusted alpha. Gender, PGY, Step 1 score, Step 2 score and medical school were used to assess differences in agreement. Data were analyzed using IBM SPSS Statistics for Windows, version 26 (IBM Corp., Armonk, NY, USA).

We used a hierarchical modeling approach and performed a multivariate logistic regression for agreement with USMLE pass/fail reporting. Our model incorporates self-reported categorical Step 1 and Step 2 CK scores, attitudes on whether each respective exam reflected knowledge at the time of examination, and medical school type. Data were weighed for medical school type. The alpha level was set at 0.05.

4. Results

Four thousand twelve respondents participated in the survey; 265 were removed due to PGY exclusion criteria. Total respondents represented 13% (3,747/29,418) of all PGY-1, PGY-2, and PGY-3 IM residents training in the USA. The distribution of gender and PGY level did not differ from the overall population of IM residents (Table 1). The post-hoc analysis showed,

Table 1. Respondents versus total population based on gender, PGY, USMLE Step 1 and Step 2 CK scores, and medical school type.

	Male n (%)	Female n (%)	p-value	PGY1 n (%)	PGY2 n (%)	PGY3 n (%)	p-value
Total Population	15,497 (58)	11,315 (42)	0.095	11,114 (39)	8839 (31)	8467 (30)	0.491
Respondents	2089 (56)	1618 (44)		1469 (39)	1131 (30)	1142 (31)	
	US Allopathic			US Osteopathic		International	p-value
Total Population	12,860 (45)			4499 (16) ^{*a}		11,061 (39)	<0.00001*
Respondents	1756 (47)			366 (9.8) ^{*a}		1594 (43) ^{*a}	

* <0.05.

^aPost-analysis, Bonferroni correction.

Table 2. Agreement with pass/fail USMLE reporting and attitudes about USMLE based on applicant characteristics.

	Step 1 and Step 2 Should be Pass/Fail				USMLE or Standardized Exam Provides Accurate Estimate of Knowledge				Step 1 Accurately Reflected My Fund of Knowledge at Examination				Step 2 Accurately Reflected My Fund of Knowledge at Examination			
	Disagree	Neutral	Agree	p-value	Disagree	Neutral	Agree	p-value	Disagree	Neutral	Agree	p-value	Disagree	Neutral	Agree	p-value
Gender^a																
Male	1237 (59)**	190 (9)**	658 (32)**	<0.001*	710 (34)**	424 (20)	952 (46)**	<0.001*	700 (34)**	325 (16)	1062 (51)**	<0.001*	563 (27)	366 (18)	1152 (55)	0.432
Female	818 (51)**	210 (13)**	588 (36)**	0.008*	664 (41)**	332 (21)	620 (38)**	0.446	627 (39)**	271 (17)	717 (44)**	0.004*	430 (27)	310 (19)	871 (54)	
Residency																
PGY1	841 (57.3%)	126 (8.6%)**	500 (34.1%)	<0.001*	531 (36)	302 (21)	633 (43)	<0.001*	548 (37)	203 (14)**	716 (49)	0.004*	393 (27)	252 (17)	818 (56)	
PGY2	604(53.5%)	135 (11.9%)	391(34.6%)		446 (40)	219 (19)	465 (41)		412 (37)	210 (19)	507 (45)		304 (27)	222 (20)	601 (53)	
PGY3	628(55.2%)	142(12.5%)	368(32.3%)		413 (36)	240 (21)	486 (43)		382 (34)	188 (17)	570 (50)		307 (27)	207 (18)	622 (55)	
Step1^b																
<200	11 (14)**	11 (14)	59 (73)**	<0.001*	64 (79)**	10 (12)	7 (9)**	<0.001*	67 (83)**	7 (9)	7 (9)**	<0.001*	38 (48)**	20 (25)	22 (28)**	<0.001*
200–220	209 (30)**	85 (12)	398 (58)**		409 (59)**	138 (20)	146 (21)**		467 (67)**	109 (16)	117 (17)**		304 (44)**	136 (20)	249 (36)**	
221–240	685 (50)**	171 (13)	506 (37)**		571 (42)**	306 (23)	483 (36)**		576 (42)**	287 (21)**	499 (37)**		417 (31)**	289 (21)**	652 (48)**	
241–260	952 (76)**	105 (8)*	193 (15)**		240 (19)**	246 (20)	765 (61)**		137 (11)**	153 (12)**	961 (77)**		166 (13)**	175 (14)**	909 (73)**	
>260	146 (84)**	11 (6)	17 (10)**	<0.001*	20 (12)**	26 (15)	128 (74)**	<0.001*	8 (5)**	10 (6)**	156 (90)**	<0.001*	11 (6)**	10 (6)**	153 (88)**	<0.001*
Step2^c																
Less than or equal to 220	84(31)**	28(10)	162(59)**	<0.001*	171(62)**	47(17)	56(20)**	<0.001*	176(64)**	32(12)	66(24)**	<0.001*	185(67)**	40(15)	49(18)**	<0.001*
221–240	505(43)**	146(12)	523(44)**		561(48)**	245(21)	367(31)**		565(48)**	214(18)	396(34)**		470(40)**	279(24)**	426(36)**	
241–260	1011(63)**	167(10)	420(26)**		490(31)**	333(21)	776(48)**		455(28)**	255(16)	889(56)**		250(16)**	271(17)	1078(67)**	
>260	413(79)**	48(9)	59(11)**	<0.001*	79(15)**	107(21)	334(64)**	<0.001*	57(11)**	59(11)**	404(78)**	<0.001*	29(6)**	43(8)**	447(86)**	0.014*
Medical School																
US Allopathic	826 (47)**	220 (13)**	706 (40)**	<0.001*	759 (43)**	386 (22)	608 (35)**	<0.001*	691 (39)*	295 (17)	768 (44)*	<0.001*	477 (27)	335 (19)	942 (54)	
US Osteopathic	192 (53)	32 (9)	141 (39)		145 (40)	80 (22)	140 (38)		152 (42)*	65 (18)	147 (40)		117 (33)	67 (19)	171 (48)	
International	1040 (66)**	148 (9)	405 (25)**		481 (30)**	289 (18)	822 (52)**		495 (31)*	239 (15)	859 (54)*		404 (25)	277 (17)	911 (57)	
Overall	2073 (55)	403 (11)	1259 (34)		1390 (37)	761 (20)	1584 (42)		1342 (36)	601 (16)	1793 (48)		1004 (27)	681 (18)	2041 (55)	

* <0.05.
 ** Significant, post-analysis, Bonferroni correction.
^aExcludes 'non-binary', 'other', and 'I do not wish to respond'.
^bExcludes 'I prefer not to respond'.
^cExcludes 'I don't wish to respond'.

however, a small but significant shortfall of osteopathic and slight over-representation of international medical graduates in our respondent group ($p = 0.00001$) compared to the overall IM trainee population (Table 1).

Two thousand seventy-three (55%) either 'strongly disagree' or 'disagree' that Step 1 and Step 2 should be reported as pass/fail while 34% (1,259) agreed, and 11% (403) gave neutral responses. Group-based differences were significant for gender, step 1 score, step 2 score, and medical school type. A higher percentage of males and IMG disagreed with the pass/fail decision. In addition, high scorers on Step 1 and Step 2 were more likely to disagree with pass/fail reporting than low scoring residents. In post-hoc analysis, each gender category and most categories for Step 1, Step 2 and medical school types were significant.

Most respondents agreed that 1) the USMLE accurately estimates knowledge (1,584, 42%) and that 2) Step 1 (1793, 48%) and 3) Step 2 CK (2041, 55%) reflected their fund of knowledge at the time of examination (Table 2). Significant between-group differences for Step 1 scores, Step 2 scores, and medical school type were found across all three of these questions upon initial analysis, and later confirmed by post-hoc analysis. Post-hoc comparisons were significant for Step 1 score, Step 2 score categories, and many medical school types. Gender differences were significant for USMLE knowledge estimation and knowledge reflection of Step 1, but not for Step 2 CK. After Bonferroni correction, agreement, and disagreement for USMLE overall and Step 1 had statistically significant counts (Table 2).

Our stepwise logistic regression model (Table 3) showed that the lowest performing groups for USMLE step 1 and USMLE Step 2 CK were 2.4 times and 1.6 times more likely to agree with pass/fail reporting compared to the top performing groups in the same category, respectively. All analyses were significant at the 0.05 significance level except for US osteopathic school applicants. In comparison to respondents who agreed that Step 1 or Step 2 CK reflected knowledge at the time of examination, those who disagreed were, respectively, 4.6 and 2.3 times more likely to agree with pass/fail reporting. US allopathic medical school graduates were two times more likely to agree with pass/fail reporting than international medical school graduates.

5. Discussion

This is the first study to assess systematically IM resident perspectives about changes in USMLE reporting from continuous to pass/fail. A majority of resident respondents (55%) disagreed with the decision to change USMLE Step 1 to pass/fail scoring. Residents' preferences for USMLE reporting

Table 3. Multivariate logistic regression with agreement with USMLE pass/fail reporting as explanatory outcome^a

	Model ($R^2 = 0.24$)		
	OR	95% CI	P
USMLE Step 1			<0.001**
≤220	2.37	1.90–2.95	<0.001**
221–240	1.57	1.32–1.87	<0.001**
≥241 (reference)	–	–	–
USMLE Step 2 CK			<0.001**
≤220	1.64	1.28–2.12	<0.001**
221–240	1.31	1.12–1.53	0.001**
≥241 (reference)	–	–	–
Step 1 Reflects Knowledge			<0.001**
Disagree	4.56	3.79–5.50	<0.001**
Neutral	2.07	1.68–2.56	<0.001**
Agree (reference)	–	–	–
Step 2 Reflects Knowledge			<0.001**
Disagree	2.25	1.87–2.70	<0.001**
Neutral	1.89	1.56–2.28	<0.001**
Agree (reference)	–	–	–
Medical School			<0.001**
US Allopathic	2.06	1.79–2.37	<0.001**
US Osteopathic	1.10	0.88–1.38	0.394
International (reference)	–	–	–

^aWeighted for medical school.

* $p < 0.05$

** $p < 0.01$.

correlated with examination performance, attitudes about the test's ability to reflect knowledge, and medical school type in the regression model. Low Step 1 scores, US allopathic status, and disagreement that the USMLE is an accurate assessment of knowledge were significant predictors for agreement with pass/fail reporting. The majority (66%) of international medical graduates responded that they disagree with pass/fail scoring.

This study has several limitations. The survey responses reflect the opinions of 13% of all Internal Medicine residents in the USA, and residents from only one specialty were surveyed. The r-squared was low at 0.24, meaning that it does not account for most of the variation in the data. Although the demographic distribution of respondents was similar to that of the total population of Internal Medicine residents in terms of gender, PGY level, and percentage of US allopathic graduates, international graduates were slightly over-represented and osteopathic graduates were under-represented. Future studies should endeavor to explore further the needs and perspectives of international and osteopathic trained residents, who constitute an essential part of our physician workforce.

6. Conclusion

The scoring rubric on the USMLE Step 1 examination will be changing to pass/fail in January 2022. Our aggregated results suggest that a majority of internal medicine residents may disagree with this decision. Further decisions regarding the scoring of USMLE Step 2 should consider these viewpoints.

Acknowledgments

The authors would like to thank Lisa L. Willet, D Craig Brater, and Sheila Costa for their invaluable advice and assistance in the preparation of this manuscript.

Disclosure statement

No potential conflict of interest was reported by the authors.

ORCID

Richard Alweis  <http://orcid.org/0000-0002-4747-8066>

References

- [1] O'Donnell MJ, Obenshain SS, Erdmann JB. Background essential to the proper use of results of step 1 and step 2 of the USMLE. *Acad Med.* 1993;68(10):734–739.
- [2] Angus SV, Williams CM, Steward EA, et al. Internal medicine residency program directors' screening practices and perceptions about recruitment challenges. *Acad Med.* 2020;95(4):582–589.
- [3] Hoffman KI. The USMLE, the NBME subject examinations, and assessment of individual academic achievement. *Acad Med.* 1993;68(10):740–747.
- [4] Kay C MD, Jackson JL MD, MPH, Frank M MD. The relationship between internal medicine residency graduate performance on the ABIM certifying examination, yearly in-service training examinations, and the USMLE step 1 examination. *Acad Med.* 2015 Jan;90(1):100–104.
- [5] Taylor ML, Blue MD, PhD AV, et al. The relationship between the national board of medical examiners' prototype of the step 2 clinical skills exam and interns' performance. *Acad Med.* 2005 May;80(5):496–501.
- [6] Williams RG. Use of NBME and USMLE examinations to evaluate medical education programs. *Acad Med.* 1993;68(10):748–752.
- [7] Green M, Jones P, Thomas JX Jr. Selection criteria for residency: results of a national program directors survey. *Acad Med.* 2009;84(3):362–367. Available from: <https://www.ncbi.nlm.nih.gov/pubmed/19240447>
- [8] Berner ES, Brooks CM, Erdmann JB. Use of the USMLE to select residents. *Acad Med.* 1993;68(10):753759.
- [9] Williams M, Kim EJ, Pappas K, et al. The impact of USA medical licensing exam (USMLE) step 1 cutoff scores on recruitment of underrepresented minorities in medicine: A retrospective cross-sectional study. *Health Sci Rep.* 2020 Apr 20;3(2):e2161. ECollection 2020 Jun. *Health Sci Rep.* 2020. PMID: 32318628.
- [10] Changes to USMLE score reporting in 2022; [cited 2020 Mar 22]. Available from: <https://www.usmle.org/usmlescoreing/>
- [11] ACGME data resource book 2018-2019, table C.5: number of active residents by specialty and subspecialty, position type, and resident year; [cited 2020 May 23]. Available from: <https://www.acgme.org/About-Us/Publications-and-Resources/Graduate-Medical-Education-Data-Resource-Book>
- [12] The match, advance data tables: 2020 main residency match, "table 2: matches by specialty and applicant type; 2020 [cited 2020 Mar 22]. Available from: <https://mk0nrmp3oyqui6wqfm.kinstacdn.com/wp-content/uploads/2020/03/Advance-Data-Tables-2020.pdf>
- [13] Lewis CE, Hiatt JR, Wilkerson L, et al. Numerical versus pass/fail scoring on the USMLE: what do medical students and residents want and why? *J Grad Med Educ.* 2011;3(1):59–66.
- [14] Finn KM, Zaas AK, McDonald FS, et al. Misinterpretation of the American board of internal medicine leave policies for resident physicians around parental leave. *Ann Intern Med.* 2019. DOI:10.7326/M19-2490
- [15] Williams CM, Alweis RL, O'Connor AB, et al. Inappropriate communication during internal medicine fellowship recruitment: a mixed-methods analysis. *Am J Med.* 2019;132(6):770–775. .
- [16] "Number of programs and residents," and "percentage of third-year residents by gender and type of medical school"; [cited 2020 Apr 12]. Available from: <https://www.abim.org/about/statistics-data/resident-fellow-workforce-data.aspx>