Numerical Versus Pass/Fail Scoring on the USMLE: What Do Medical Students and Residents Want and Why?

Catherine E. Lewis, MD
Jonathan R. Hiatt, MD
LuAnn Wilkerson, EdD
Areti Tillou, MD
Neil H. Parker, MD
O. Joe Hines, MD

Abstract

Background Although the primary purpose of the US Medical Licensing Examination (USMLE) is assessment for licensure, USMLE scores often are used for other purposes, more prominently resident selection. The Committee to Evaluate the USMLE Program currently is considering a number of substantial changes, including conversion to pass/fail scoring.

Methods A survey was administered to third-year (MS3) and fourth-year (MS4) medical students and residents at a single institution to evaluate opinions regarding pass/fail scoring on the USMLE.

Results Response rate was 59% (n = 732 of 1249). Reported score distribution for Step 1 was 30% for <220, 38% for 220–240, and 32% for >240, with no difference between MS3s, MS4s, and residents ($P = .89$). Score distribution for Step 2 Clinical Knowledge (CK) was similar. Only 26% of respondents agreed that Step 1 should be pass/fail; 38% agreed with pass/fail scoring for Step 2 CK. Numerical scoring on Step 1 was preferred by respondents who: (1) agreed that the examination gave an accurate estimate of knowledge (odds ratio [OR], 4.23; confidence interval [CI], 2.41–7.43; $P < .001$); (2) scored >240 (OR, 4.0; CI, 1.92–8.33; $P < .001$); and (3) felt that acquisition of knowledge might decrease if the examination were pass/fail (OR, 10.15; CI, 3.32–31.02; $P < .001$). For Step 2 CK, numerical scoring was preferred by respondents who: (1) believed they gained a large amount of knowledge preparing for the examination (OR, 2.63; CI, 1.52–4.76; $P < .001$); (2) scored >240 (OR, 4.76; CI, 2.86–8.33; $P < .001$); (3) felt that the amount of knowledge acquired might decrease if it were pass/fail (OR, 28.16; CI, 7.31–108.43; $P < .001$); and (4) believed their Step 2 CK score was important when applying for residency (OR, 2.37; CI, 1.47–3.84; $P < .001$).

Conclusions Students and residents prefer the ongoing use of numerical scoring because they believe that scores are important in residency selection, that residency applicants are advantaged by examination scores, and that scores provide an important impetus to review and solidify medical knowledge.

Introduction

Since its implementation in the early 1990s, the United States Medical Licensure Examination (USMLE) program[1] has undergone relatively little change aside from transitioning to computer-based testing in 1999 and adding the Step 2 Clinical Skills examination in 2004. In contrast, major changes have occurred in medical education as well as the regulatory and practice environments. Recognizing this, the Composite Committee that governs the USMLE called for a review of the program to determine whether its current design, structure, and format continue to support expectations of the state medical licensing boards and the broader profession.[2] The 2 cosponsors of the examination, the National Board of Medical Examiners (NBME) and the Educational Commission on Foreign Medical Graduates, convened the Committee to Evaluate the USMLE Program (CEUP) to review and provide recommendations on the exam.[3] The initial draft recommendations included a proposed change in scoring from numerical to pass/fail. In contrast, the final report released in June 2008 excluded this proposed scoring change and noted that the committee felt that "the implications of [its] other recommendations… need[ed] to be further defined before USMLE would be in a position to consider this reporting issue."[3]

All authors are at University of California, Los Angeles. **Catherine E. Lewis, MD,** is a general surgery resident; **Jonathan R. Hiatt, MD,** is Professor of Surgery, Vice Chair for Surgical Education, and Chief, Division of General Surgery; **LuAnn Wilkerson, EdD,** is Professor of Medicine and Senior Associate Dean for Medical Education, David Geffen School of Medicine; **Areti Tillou, MD,** is Assistant Professor of Surgery and Associate Director of the Surgical Residency Program, Division of General Surgery; **Neil H. Parker, MD,** is Professor of Medicine and Senior Associate Dean for Student Affairs and Graduate Medical Education, David Geffen School of Medicine; and **O. Joe Hines, MD,** is Professor of Surgery and Director of the Surgical Residency Program, Division of General Surgery.

Corresponding author: O. Joe Hines, MD, David Geffen School of Medicine at UCLA, 10833 Le Conte Avenue, 72-170 CHS, Box 956904, Los Angeles, CA 90095-6904, 310.206.0441, joehines@mednet.ucla.edu

Received June 26, 2010; revision received September 8, 2010; accepted October 11, 2010.

DOI: 10.4300/JGME-D-10-00121.1

Although the USMLE was designed to meet the needs of state licensure boards, it has significant secondary uses, including curriculum assessment,[4] promotion and graduation decisions,[5] and residency selection.[6] Using scores for nonlicensure purposes has been called an abuse by some educators and has prompted an ongoing debate over score reporting.[7] Arguments against numerical scoring include assertions that students study only content represented on the USMLE examinations, thus encouraging faculty to teach to the test, and that through this mechanism the NBME determines a significant portion of the medical school curriculum. In addition, although USMLE scores have a low to moderate correlation with future clinical and examination performance, there is little evidence that small differences in scores sometimes used to distinguish among students or residency applicants relate to subsequent performance.[6] The NBME acknowledges these concerns, and in response has performed regular reviews of its score reporting policy.[8,9]

This survey-based study was undertaken to evaluate medical students' and residents' opinions regarding the implications of a potential shift to pass/fail scoring of the USMLE Steps 1 and 2 Clinical Knowledge (CK). Survey questions focused on USMLE scoring as it relates to medical knowledge assessment and applications for residency training.

Methods

This study was reviewed and certified exempt by the Office for Protection of Research Subjects, Institutional Review Board, University of California, Los Angeles (UCLA). Participants included third-year medical students (MS3s), fourth-year medical students (MS4s), and residents at the David Geffen School of Medicine, UCLA. Web-based surveys were distributed to all residents and graduating MS4s by e-mail invitation in the spring of 2008. Nonrespondents were sent 3 follow-up e-mail reminders during the next 3 weeks to increase response rate. A paper-based survey was administered to MS3s during orientation sessions for the fourth year. MS3s attending orientations at outside institutions (combined UCLA/Drew and UCLA/UC Riverside students) were not surveyed. Participants were blinded to any specific hypotheses of the study. Responses were confidential, and participation in the survey was voluntary.

Survey Development

The survey was developed by 2 authors (C.E.L. and O.J.H.) using available literature, their own expertise, and information gained from a lecture given at the Keck School of Medicine by Peter Scoles, senior vice president for assessment programs, NBME (Feb. 20, 2008). The questions were refined and finalized in consultation with the other authors.

Questions on All Surveys We collected Step 1 performance information by asking respondents to place their score within a range (<180, 180–199, 200–219, 220–240, or >240). Respondents then answered questions regarding examination preparation, content, scoring, and relevance on a 5-point Likert scale, with the middle value being neutral or no effect (TABLE 1), and they indicated their current or planned specialty training. We also asked respondents to rank the following items in terms of importance in obtaining a desired residency: Alpha Omega Alpha membership, Medical School Performance Evaluation (MSPE, also known as the dean's letter), extracurricular activities, letters of recommendation, personal statement, prestige of medical school, research experience, and USMLE scores. Finally, respondents were invited to provide any additional comments.

Questions for MS4s and Residents Because MS3s had not yet taken the Step 2 CK examination, nor had they gone through the interview and matching process for residency, some survey questions were provided to MS4s and residents only (TABLE 1). Questions for this group included Step 2 CK performance information, opinions regarding Step 2 CK preparation and relevance, and opinions regarding effects of Step 1 scoring on number of residency interviews and ranking by residency programs.

Questions for Residents We asked residents who take a yearly specialty in-training examination to estimate the degree of correlation between these scores and Step 1 scores based on their personal performance.

Statistical Analysis

Descriptive statistics were used to summarize responses to all questions, and comments were grouped by theme based on consensus by authors. For statistical analysis, we combined the responses into 3 data sets: the lower 2 levels (eg, strongly disagree and disagree), the upper 2 levels (eg, strongly agree and agree), and the middle neutral or no effect level. We used the χ^2 statistic for univariate analysis of all survey questions to test for independence between question response and preference for pass/fail scoring. Questions pertaining to Step 1 were analyzed against scoring preference on the Step 1 examination, and likewise for Step 2 CK. We then used multivariate ordinal logistic regression to determine the relationship among variables and preference for numerical versus pass/fail scoring. Two regression models were created, one for each Step, with each model containing only questions pertaining to the specific examination. All relevant questions were entered into each model, and then stepwise regression with backward elimination of predictors was used to select significant variables. At each step, results of the Wald χ^2 test for individual parameters were examined, and the least significant effect that did not meet the study criteria was removed. Once an effect was removed from the model, it remained excluded. The process was repeated until no other

TABLE 1 **Opinions of Medical Students and Residents Regarding Numerical Versus Pass/Fail Scoring on United States Medical Licensing Examinations**

Question	Scale Anchors	Likert Scale, % of Time Response Was Selected[a]					P^b
		1	2	3	4	5	
All respondents (MS3, MS4, and residents)							
I would like Step 1 to be pass/fail	Strongly disagree to strongly agree	36	24	14	12	14	N/A
I would like Step 2 CK to be pass/fail	Strongly disagree to strongly agree	26	21	15	15	23	N/A
Compared with medical school classmates, I am	Far below to well above average	0	3	31	48	18	<.001
I believe Step 1 gives accurate estimation of knowledge	Strongly disagree to strongly agree	5	16	32	33	13	<.001
The amount of knowledge I gained/ relearned preparing for Step 1	None to a very large amount	0	10	30	37	23	<.001
If Step 1 was pass/fail, amount of knowledge gained/relearned would	Markedly decrease to markedly increase	18	36	42	3	1	<.001
For residency application, Step 1 score is	Unimportant to very important	0	4	20	33	43	<.001
For residency application, Step 2 score is	Unimportant to very important	14	28	27	22	9	.01
Relevance of Step 1 Content to future *residency training*	Not relevant to extremely relevant	6	29	41	19	5	<.001
Relevance of Step 1 Content to future *medical practice*	Not relevant to extremely relevant	11	33	38	15	3	<.001
MS4s and residents only							
The amount of knowledge I gained/ relearned preparing for Step 2 CK	None to a very large amount	3	18	40	30	9	<.01
If Step 2 was pass/fail, amount of knowledge gained/relearned would	Markedly decrease to markedly increase	12	31	55	2	0	<.001
Relevance of Step 2 Content to future *residency training*	Not relevant to extremely relevant	2	12	40	36	9	<.01
Relevance of Step 2 Content to future *medical practice*	Not relevant to extremely relevant	3	15	43	32	8	.02
If Step 1 were pass/fail, number of interviews granted to me would	Markedly decrease to markedly increase	7	27	46	15	6	<.001
If Step 1 were pass/fail I believe I would have been ranked more highly by residency programs	Strongly disagree to strongly agree	31	24	25	11	9	<.001
Residents only							
Compared with other residents, I am	Far below to well above average	0	3	44	45	8	0.26
Correlation between Step 1 score and yearly in-training exam scores	Very poor to very well	5	20	38	26	10	<.001

Abbreviations: CK, Clinical Knowledge; MS3s, third-year medical students; MS4s, fourth-year medical students.

[a] For statistical analysis, we reduced all responses from 5 to 3 levels by combining the lower 2 levels (1 and 2) and the upper 2 levels (4 and 5), and leaving the middle level (3) as neutral.

[b] Chi-square test for association between question response and preference for pass/fail scoring. Questions pertaining to Step 1 were analyzed against preference for scoring on Step 1 and likewise for Step 2 CK.

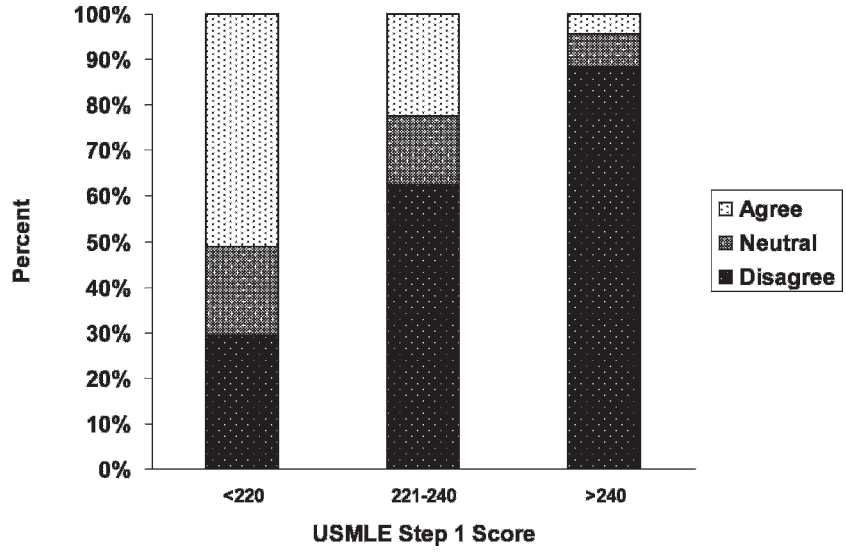

FIGURE **Medical Student and Resident Agreement With United States Medical Licensing Examination Step 1 Pass/Fail Scoring by Reported Step 1 Score**

Univariate χ^2 analysis demonstrated that respondents with Step 1 scores above 240 were more likely to prefer numerical scoring compared with respondents with scores between 220 and 240 and those with scores below 220 (88% versus 63% versus 29%; $P < .001$)

effect in the model met the specified level for removal. The significant level of the Wald χ^2 test for an effect to stay in the model was 0.05 in our model building.

Results

Pass/Fail Versus Numerical Scoring

Our overall response rate was 59% (732 of 1249). Response rate was 55% for residents (n = 501 of 911), 53% for MS4s (n = 109 of 205), and 93% for MS3s (n = 122 of 131). Results from the survey are summarized in TABLE 1.

Overall, only 26% of respondents (n = 189) agreed that Step 1 scoring should be pass/fail, whereas 60% (n = 437) were in favor of numerical scoring, and 36% (n = 265) favored numerical scoring strongly. Respondents were more in favor of pass/fail scoring on Step 2 CK compared with Step 1, with 38% (n = 281) preferring pass/fail and only 47% (n = 338) preferring numerical scores. Among respondents, MS4s were more in favor of pass/fail scoring for Step 1 than were MS3s or residents (38% [n = 41] versus 30% [n = 35] versus 22% [n = 113]; $P = .02$). MS3s were more in favor of pass/fail scoring for Step 2 CK than were MS4s or residents (54% [n = 65] versus 48% [n = 51] versus 33% [n = 165]; $P < .001$).

Examination Performance and Specialty

Reported score distribution for Step 1 was as follows: <220 for 30% of respondents (n = 211), 220 to 240 for 38% (n = 272), and >240 for 32% (n = 224); there was no difference among MS3s, MS4s, and residents ($P = .60$). Scores for Step 2 CK for MS4s and residents were similar, with 33% (n = 191) scoring <220, 38% (n = 220) scoring 221 to 240, and 29% (n = 164) scoring >240.

Respondents with higher Step 1 scores were more likely to prefer numerical scoring (FIGURE). Similarly, respondents with Step 2 CK scores >240 were more likely to prefer numerical scoring ($P < .001$). Multivariate analysis by ordinal logistic regression (TABLE 2) also demonstrated that respondents with scores >240 were more likely to prefer numerical scoring for both Step 1 and Step 2 CK. Respondents' Step 1 scoring preferences were not statistically different with respect to specialty choice. We performed further analysis by combining specialties into 3 categories based on competitiveness of individual specialties (TABLE 3). Competitiveness was estimated using data from the 2008 National Residency Matching Program (NRMP), including the number of applications from US senior medical students per position and the unmatched fraction for each specialty position. The authors determined that ophthalmology should be placed in the most competitive category, because this specialty does not participate in the NRMP. When specialties were combined by competitiveness, scoring preferences were significantly different, with respondents in more competitive specialties preferring numerical over pass/fail scoring ($P < .001$).

TABLE 2 **Ordinal Logistic Regression Analysis of Survey Response Variables as Predictors for Numerical Scoring Preference on the United States Medical Licensing Examination (USMLE)**

	OR	95% CI	*P*
Step 1			
Examination score >240 versus 220–240 versus <220	4.0	1.92–8.33	<.001
Examination gives accurate estimation of candidates' knowledge	4.23	2.41–7.43	<.001
Knowledge would decrease if pass/fail scoring were used	10.15	3.32–31.02	<.001
Step 2 CK			
Examination score is important when applying for residency	2.37	1.47–3.84	<.001
A large amount of knowledge is gained studying for examination	2.63	1.52–4.76	<.001
Examination score >240 versus 220–240 versus <220	4.76	2.86–8.33	<.001
Knowledge would decrease if pass/fail scoring were used	28.16	7.31–108.43	<.001

Abbreviations: CI, confidence interval; CK, Clinical Knowledge; OR, odds ratio.

Knowledge Assessment and Step 1

Sixty percent of respondents indicated that they gained or relearned a large amount of knowledge in preparation for Step 1, and 23% rated it a very large amount; 54% believed that this amount would decrease or markedly decrease if the scoring were pass/fail. Univariate χ^2 (TABLE 1) and multivariate ordinal logistic regression (TABLE 2) analyses demonstrated that numerical scoring was preferred by respondents who believed that Step 1 had educational value. Numerical scoring was preferred significantly more often by respondents who believed that Step 1 gave an accurate estimation of knowledge, that a large amount of knowledge was gained in preparing for the examination, that knowledge would decrease if the exam were pass/fail, that content on Step 1 was relevant to future residency training, and that content was relevant to future medical practice. Multivariate analysis demonstrated that respondents who agreed that Step 1 gave an accurate estimate of knowledge were more likely to prefer numerical scoring, as were respondents who felt that the amount of knowledge learned in preparation for the examination would decrease if pass/fail scoring were used (odds ratios, 10.15 for Step 1 and 28.16 for Step 2 CK).

Knowledge Assessment and Step 2 CK

In general, respondents identified lower educational value for Step 2 CK than for Step 1, but they still identified a justification for numerical scoring. Although the amount of knowledge gained or relearned in preparation for Step 2 CK was rated as a large amount by only 39% of respondents and as little or none by 21%, 43% felt that the amount of knowledge gained would decrease or markedly decrease were Step 2 CK to become pass/fail. Similar to analyses for Step 1 scoring, univariate (TABLE 1) and multivariate (TABLE 2) analyses demonstrated that numerical scoring for Step 2 CK was preferred by respondents who believed that the examination had educational value.

Residency Selection Process

Seventy-five percent of respondents judged the Step 1 score to be important to the residency application process, and 31% judged the Step 2 CK score to be important. Were Step 1 scoring to become pass/fail, 46% of respondents believed that the number of invitations for interviews would remain the same, 34% believed the number would decline, and 55% believed that they would have been ranked lower by residency programs. Univariate (TABLE 1) and multivariate (TABLE 2) analyses demonstrated that respondents who believed that the examinations were important in resident selection were more in favor of numerical scoring for both Step 1 and Step 2 CK.

Comparing various items used in the residency application process (TABLE 4), MS3s and MS4s believed USMLE scores were most important. For all respondents, recommendation letters, USMLE scores, and the MSPE were ranked highest, followed by prestige of medical school, Alpha Omega Alpha status, research experience, personal statement, and extracurricular activities.

Respondents' Comments

A total of 158 comments were received from 132 respondents, most concerning USMLE scores in the residency application process. The most common theme was that scores allow students to distinguish themselves when applying for residency (34%; n = 54), particularly for students attending medical schools with pass/fail grading (25%; n = 40). The issue was of increased importance for students, with 27 of 52 comments from MS3s and MS4s, who expressed concern that they would not be able to

TABLE 3 **Medical Student and Resident Preference for Step 1 Pass/Fail Scoring by Competitiveness of Specialty and Number of Survey Respondents Per Specialty**

		I Would Like Step 1 to Be Pass/Fail		
Specialty	**No. of Respondents (%)**	**Disagree, %**	**Neutral, %**	**Agree, %**
Highly competitive	201 (28)	73	9	18
Dermatology	13 (2)			
General surgery	57 (8)			
Ophthalmology	19 (3)			
Orthopedic surgery	33 (5)			
Otolaryngology	27 (4)			
Plastic surgery	8 (1)			
Radiation oncology	3 (0)			
Radiology	41 (6)			
Competitive	170 (23)	61	14	25
Anesthesiology	72 (10)			
Emergency medicine	25 (3)			
Internal medicine/pediatrics	9 (1)			
Neurosurgical surgery	11 (2)			
Obstetrics and gynecology	37 (5)			
Urology	16 (2)			
Less competitive	361 (49)	52	17	31
Family medicine	37 (5)			
Internal medicine	134 (18)			
Neurology	21 (3)			
Pathology	22 (3)			
Pediatrics	81 (11)			
Physical medicine and rehabilitation	3 (0)			
Psychiatry	44 (6)			
Undecided/unmatched/preliminary	19 (3)			
Total	**732 (100)**	***P* < .001[a]**		

[a] *P* value for χ^2 analysis of preference for Step 1 pass/fail scoring based on competitiveness of specialty.

distinguish themselves with USMLE pass/fail scoring. Among all respondents, 21% (n = 33) identified the importance of an objective measure for residency applications and the USMLE scores as the only such measure at present. Other themes included: USMLE scores are a reflection of effort put into examination preparation and/or one's test-taking skills (11%; n = 18); test scores may not accurately represent a student's quality as a residency applicant (9%; n = 14); pass/fail scoring would decrease test preparation and hence the fund of knowledge (8%; n = 12); and pass/fail scoring would promote laziness (5%; n = 8).

Discussion

Numerical scores have been reported on all NBME examinations since 1916,[8] and prior to the CEUP, the NBME upheld this practice following reviews of its score-reporting policy in 1989[9] and 1997.[8] Reporting of numerical scores has prompted an ongoing debate because

TABLE 4 **Survey Opinions of Medical Students and Residents Regarding Rank of Importance of Residency Application Items (1 Is Most Important and 8 Is Least Important)**

	Item Rank (Average Rank)			
Application Item	**All**	**MS3**	**MS4**	**Resident**
Recommendation letters	1 (3.12)	2 (3.48)	2 (3.07)	2 (3.04)
USMLE scores	2 (3.15)	1 (2.61)	1 (2.95)	3 (3.33)
Dean's letter	3 (3.19)	3 (3.61)	3 (3.75)	1 (2.96)
Medical school prestige	4 (4.33)	4 (4.08)	4 (4.18)	4 (4.42)
Alpha Omega Alpha membership	5 (4.61)	5 (4.81)	5 (4.84)	5 (4.51)
Research experience	6 (5.36)	6 (5.05)	6 (5.02)	6 (5.51)
Personal statement	7 (5.89)	7 (6.00)	7 (5.93)	7 (5.85)
Extracurricular activities	8 (6.35)	8 (6.36)	8 (6.26)	8 (6.37)

Abbreviations: MS3s, third-year medical students; MS4s, fourth-year medical students; USMLE, United States Medical Licensing Examination.

some educators believe that scores should not be used for nonlicensure purposes, and thus advocate for pass/fail scoring to prevent this.[7] Others argue that use of scores for secondary purposes is justified because the examinations are standardized and highly reliable tools[1] that provide valid assessments of the content tested.[1,5,8]

Medical student preference for pass/fail scoring has varied over time. In the review conducted by the NBME in 1989,[9] fewer than half of all respondents favored pass/fail scoring for either Step 1 or 2. In its more recent review conducted in 1997,[8] 61% were in favor of pass/fail scoring, and although the CEUP did not provide specific figures, it reported that surveys of medical student leaders revealed a strong preference for pass/fail scoring.[3] In our study, all respondents (both medical students and residents) were more in favor of numerical scoring. As expected, and in agreement with prior reports,[8] performance on the examination was directly related to scoring preference, with high scores predicting preference for numerical scoring and low scores predicting preference for pass/fail scoring. Our students and residents reported using the USMLE, particularly Step 1, as an important tool to review, learn, and solidify knowledge during preparation for the examination. Prior studies have confirmed lower performance for examinations with pass/fail scoring.[10,11]

Our survey respondents identified the importance of USMLE scores, particularly the Step 1 score, in the residency application process. This is in agreement with prior studies,[8] which have shown that medical students believed that their USMLE scores helped them obtain a desired residency. The importance of USMLE scores in the residency application process has been one of the major reasons the NBME has upheld numerical scoring.[8] A survey of program directors by the NRMP in 2008 revealed that Step 1 scores are required by 97.7% of all programs, Step 2 scores are required by 74.6% of all programs, and Step 1 scores received the third highest ranking among factors considered in applicant selection.[12] In an earlier study by Bowles et al,[8] 82% of program directors favored numerical over pass/fail scoring for both Steps 1 and 2, and many reported relying heavily on USMLE scores as a standardized measure of applicants. In addition, many program directors state they would look to other objective instruments, such as the Medical College Admission Test (MCAT) or SAT examinations, or require a new standardized test to be constructed if pass/fail scoring were established for the USMLE.[8]

Program director comments from the Bowles et al study[8] reflected that USMLE scores are a nationally standardized, objective report of information in the student profile, and that other measures, such as dean's letters and clerkship grades, are far more subjective and difficult to interpret. Dean's letters are often advocacy documents that fail to provide residency programs with negative information on a student. A 1999 study by Edmond et al[13] revealed that the evaluated variables were omitted from dean's letters in up to 50% of cases in which they were present on the transcripts. Omissions included failing or marginal grades, and even the requirement to repeat an entire year. Since this time, the dean's letter has been transformed into the MSPE; however, in a recent study by Shea et al,[14] only 70% to 80% of MSPEs stated grades clearly, fewer than 70% indicated whether the student had had any adverse actions, and only 17% provided comparative data in the summary paragraph. Medical school grades may be predictive of future performance, particularly in the specialty of interest[15–18]; however, there is significant variability in grading systems used in medical schools, and an even greater range of possible scores within those various systems. The percentage of highest grades given among different institutions and

among clerkships within a single institution is also highly inconsistent. Furthermore, almost 10% of schools do not use any grading system, and some fail to provide interpretable data regarding their grading system in the MSPEs.[14,19]

In order to support the use of USMLE scores in the resident application process, scores must provide useful predictive information regarding future performance. Substantial literature exists on this topic, and most studies have demonstrated that USMLE scores correlate with performance during residency and future certifying examinations.[6,8,9,17,20–23] The degree of correlation varies depending on which USMLE Step/Part is evaluated and the outcomes of interest, with the highest correlation generally between performance on the USMLE and future examinations, either during or at conclusion of training. Still, even when evaluating which measures best predict resident ratings by faculty or performance in practice, USMLE scores have consistently been the most significant and accurate variable.

One limitation to the current study may be selection bias, because surveys were administered at a single institution, which may not be representative of all institutions. Our medical school and residency programs have traditionally been rated highly,[24,25] and may therefore select more competitive applicants. As our analysis showed, those in more competitive specialties were more in favor of numerical scoring, and this may also translate to respondents in more competitive medical schools or residency programs. Another consideration is that our institution uses a pass/fail grading system throughout all 4 years of medical school. Lack of formal grades and class rank may cause medical students to judge the importance of USMLE examination scores more highly, and therefore make respondents more likely to desire numerical scoring. An additional limitation to our survey is the inaccuracy of self-reported data. Most respondents rated themselves as average or above average, and only 3% of participants reported themselves to be below average. Also, although we verified that the reported examination score distribution appeared to be accurate based on known actual performance by our medical students, there was no way to verify this for our residents.

Conclusions

The USMLE examination will likely undergo significant change within the next 5 years as the NBME reviews the recommendations made by CEUP and develops hypothetical new models for the USMLE program. The results of our survey show that our students and residents prefer the ongoing use of numerical scoring because they believe that Step 1 scores are important in residency selection, that residency applicants are advantaged by examination scores, and that scores provide an important impetus to review and solidify basic medical knowledge.

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 Scoles PV. Comprehensive review of the USMLE. *Adv Physiol Educ*. 2008;32(2):109–110.
3 Committee to Evaluate the USMLE Program. Comprehensive review of USMLE. Summary of the final report and recommendations. Available at: http://www.usmle.org/general_information/CEUP-Summary-Report-June2008.PDF. Accessed December 3, 2008.
4 Williams RG. Use of NBME and USMLE examinations to evaluate medical education programs. *Acad Med*. 1993;68(10):748–752.
5 Hoffman KI. The USMLE, the NBME subject examinations, and assessment of individual academic achievement. *Acad Med*. 1993;68(10):740–747.
6 Berner ES, Brooks CM, Erdmann JB. Use of the USMLE to select residents. *Acad Med*. 1993;68(10):753–759.
7 Muller S. Physicians for the twenty-first century: report of the Project Panel on the General Professional Education of the Physician and College Preparation for Medicine. *J Med Educ*. 1984;59(11, pt 2):1–208.
8 Bowles LT, Melnick DE, Nungester RJ, et al. Review of the score-reporting policy for the United States Medical Licensing Examination. *Acad Med*. 2000;75(5):426–431.
9 Nungester RJ, Dawson-Saunders B, Kelley PR, Volle RL. Score reporting on NBME examinations. *Acad Med*. 1990;65(12):723–729.
10 Weller LD. The grading nemesis: an historical overview and a current look at pass/fail grading. *J Res Devel Educ*. 1983;17(1):39–45.
11 Suddick DE, Kelly RE. Effects of transition from pass/no credit to traditional letter grade system. *J Exp Educ*. 1981;50:88–90.
12 National Resident Matching Program (NRMP). Results of the 2008 NRMP Program Director Survey. Available at: http://www.nrmp.org/data/programresultsbyspecialty.pdf. Accessed February 2, 2009.
13 Edmond M, Roberson M, Hasan N. The dishonest dean's letter: an analysis of 532 dean's letters from 99 U.S. medical schools. *Acad Med*. 1999;74(9):1033–1035.
14 Shea JA, O'Grady E, Morrison G, Wagner BR, Morris JB. Medical Student Performance Evaluations in 2005: an improvement over the former dean's letter? *Acad Med*. 2008;83(3):284–291.
15 Andriole DA, Jeffe DB, Whelan AJ. What predicts surgical internship performance? *Am J Surg*. 2004;188(2):161–164.
16 Boyse TD, Patterson SK, Cohan RH, et al. Does medical school performance predict radiology resident performance? *Acad Radiol*. 2002;9(4):437–445.
17 Hamdy H, Prasad K, Anderson MB, et al. BEME systematic review: predictive values of measurements obtained in medical schools and future performance in medical practice. *Med Teach*. 2006;28(2):103–116.
18 Gonnella JS, Erdmann JB, Hojat M. An empirical study of the predictive validity of number grades in medical school using 3 decades of longitudinal data: implications for a grading system. *Med Educ*. 2004;38(4):425–434.
19 Takayama H, Grinsell R, Brock D, Foy H, Pellegrini C, Horvath K. Is it appropriate to use core clerkship grades in the selection of residents? *Curr Surg*. 2006;63(6):391–396.
20 Naylor RA, Reisch JS, Valentine RJ. Factors related to attrition in surgery residency based on application data. *Arch Surg*. 2008;143(7):647–651.
21 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*. 2011;38(1):65–69.
22 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):1016–1019.
23 Armstrong A, Alvero R, Nielsen P, 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.
24 *US News and World Report*. Best Medical Schools (ranked in 2008). Available at: http://grad-schools.usnews.rankingsandreviews.com/grad/med/search. Accessed December 20, 2008.
25 *US News and World Report*. America's Best Hospitals. Available at: http://health.usnews.com/sections/health/best-hospitals. Accessed December 20, 2008.