

Facilitator: To Be Determined
 Discussant: Leslee J. Martin, MA

Study Behaviors and USMLE Step 1 Performance: Implications of a Student Self-Directed Parallel Curriculum

Jesse Burk-Rafel, MD, MRes, Sally A. Santen, MD, PhD, and Joel Purkiss, PhD

Abstract

Purpose

To determine medical students' study behaviors when preparing for the United States Medical Licensing Examination (USMLE) Step 1, and how these behaviors are associated with Step 1 scores when controlling for likely covariates.

Method

The authors distributed a study-behaviors survey in 2014 and 2015 at their institution to two cohorts of medical students who had recently taken Step 1. Demographic and academic data were linked to responses. Descriptive statistics, bivariate correlations, and multiple linear regression analyses were performed.

Results

Of 332 medical students, 274 (82.5%) participated. Most students ($n = 211$; 77.0%) began studying for Step 1 during their preclinical curriculum, increasing their intensity during a protected study period during which they averaged 11.0 hours studying per day (standard deviation [SD] 2.1) over a period of 35.3 days (SD 6.2). Students used numerous third-party resources, including reading an exam-specific 700-page review book on average 2.1 times (SD 0.8) and completing an average of 3,597 practice multiple-choice questions (SD 1,611). Initiating study prior to the designated study period, increased review book usage,

and attempting more practice questions were all associated with higher Step 1 scores, even when controlling for Medical College Admission Test scores, preclinical exam performance, and self-identified score goal (adjusted $R^2 = 0.56$, $P < .001$).

Conclusions

Medical students at one public institution engaged in a self-directed, "parallel" Step 1 curriculum using third-party study resources. Several study behaviors were associated with improved USMLE Step 1 performance, informing both institutional- and student-directed preparation for this high-stakes exam.

The United States Medical Licensing Examination (USMLE) Step 1 ("Step 1") is a crucial milestone for medical

J. Burk-Rafel is an intern, Department of Medicine, New York University School of Medicine, New York, New York. He was previously a medical student, University of Michigan Medical School, Ann Arbor, Michigan; ORCID: <http://orcid.org/0000-0003-3785-2154>.

S.A. Santen is assistant dean, Educational Research and Quality Improvement, and clinical professor of emergency medicine and learning health sciences, University of Michigan Medical School, Ann Arbor, Michigan; ORCID: <http://orcid.org/0000-0002-8327-8002>.

J. Purkiss is assistant dean, Evaluation, Assessment, and Education Research, and assistant professor of internal medicine, Baylor College of Medicine, Houston, Texas. He was previously director, Evaluation and Assessment, University of Michigan Medical School, Ann Arbor, Michigan.

Correspondence should be addressed to Jesse Burk-Rafel, Office of Graduate Medical Education, NYU School of Medicine, 545 First Ave., Greenberg Hall, SC1-081, New York, NY 10016; e-mail: jesse.rafel@nyumc.org; telephone: (206) 334-2187; Twitter: @jbrafel.

Acad Med. 2017;92:S67-S74.

doi: 10.1097/ACM.0000000000001916

Copyright © 2017 by the Association of American Medical Colleges

Supplemental digital content for this article is available at <http://links.lww.com/ACADMED/A479>.

students. Passing Step 1 is necessary for progression through most medical schools, and for U.S. medical licensure. Moreover, residency program directors cite Step 1 score as the most commonly used factor when selecting applicants to interview, rating it more important than core clerkship grades, leadership/volunteerism, and the Medical Student Performance Evaluation.¹ Thus, while there are convincing arguments for reduced reliance on Step 1 scores in residency candidate selection,^{2,3} optimizing student performance on this high-stakes exam—even above the passing score threshold—remains critical for students and medical schools alike.

Towards this end, numerous studies have identified predictors of improved Step 1 performance, including prior academic performance (e.g., increased undergraduate grade point average,^{4,5} Medical College Admission Test [MCAT] scores,^{6,7} medical school preclinical performance,⁸⁻¹⁰ USMLE practice exam scores^{11,12}); student demographics (e.g., younger age,¹³ male gender,¹⁴ white

or Asian race,⁷ decreased premedical debt¹⁵); and curricular factors (e.g., problem-based learning,^{16,17} system-based education,¹⁸ content integration^{19,20}). While such factors are helpful in predicting students' expected Step 1 performance and identifying those at risk of exam failure,^{9,21,22} in our experience, medical students beginning their preparations for Step 1 take little comfort in such nonmodifiable factors. Rather, many students understandably desire advising on modifiable behaviors, such as study strategies or use of specific resources, to improve their Step 1 performance.

There is a comparative paucity of evidence regarding student-modifiable exam preparation approaches: Practice question usage,²³⁻²⁵ spaced repetition activities,^{26,27} and increased study time^{23,28} have correlated with improved Step 1 scores, while exam timing,²⁹ use of commercial preparation courses,³⁰ and process-oriented preparation³¹ have shown no clear association with scores. Unfortunately, many of these

studies are compromised by dated samples, low power, or inadequate control of confounding covariates (e.g., student demographics, prior academic performance), further limiting evidence-based advising.

In this student-initiated study, we sought to identify students' self-selected study tools and behaviors when preparing for the USMLE Step 1 and to examine their association with exam performance. On the basis of limited existing literature, we hypothesized that certain study behaviors and tools—such as earlier initiation of dedicated Step 1 studying, spending more time studying, and increased use of Step 1-focused review books and practice questions—are positively associated with Step 1 scores. Furthermore, we hypothesized that these exam-directed study behaviors predict Step 1 scores even when relevant covariates are statistically controlled.

Method

Local curriculum and USMLE Step 1

At the time of this study (2014, 2015), our medical school, a large public research institution, used a systems-based two-year pass/fail preclinical curriculum that did not overtly emphasize Step 1 preparation. Rather, students were encouraged to focus on their preclinical course work and examinations, which were created in-house by faculty. Students were then provided up to six weeks of protected Step 1 study time (the “study period”) at the conclusion of their second year, and our alumni association purchased *First Aid for the USMLE Step 1*³² (hereafter “review book”) for all students.

Participants

All medical students at the study institution completing their first Step 1 attempt within academic years 2014 or 2015 were included in the study, which received exemption from the University of Michigan Health Sciences Institutional Review Board. Among all students in this two-year cohort ($n = 332$), mean MCAT score was 34.7 (standard deviation [SD] 3.2), mean Step 1 score was 235.6 (SD 17.3), and overall Step 1 passing rate was 98.8%. To encourage response to our voluntary survey, a monetary incentive was provided to a randomly selected subset of respondents.

Study-behaviors survey

We created a study-behaviors survey (Supplemental Digital Appendix 1, <http://links.lww.com/ACADMED/A479>) based on input from diverse stakeholders (faculty, counselors, learning specialists), review of previous studies,^{25,28,31,33} and a local needs assessment of second-year students ($n = 121$) who had not yet taken the exam—all of which enhanced content validity. Respondents reported their approach to course work, timing of Step 1 study, use of study tools and resources, and “score goal”—the self-identified minimum score they wished to achieve when they began their protected study period. Practice-question usage was self-reported as a completion fraction of UWorld,³⁴ USMLE-Rx,³⁵ or Kaplan³⁶ question banks or self-assessments; unique questions (first-time completion) were differentiated from repeated questions. To ensure that we captured the scope of Step 1 study behaviors, we piloted the survey on five medical students who had recently completed Step 1, which promoted content and response process validity. Pilot respondents felt that score goal was an important determinant of overall study intensity, which led to its inclusion in this work. We distributed the survey electronically in September 2014 and June 2015 using Qualtrics software (Provo, Utah).

Demographic and academic performance data

We retrieved students' sex, most recent MCAT score, “preclinical score” (a weighted average of exam scores across preclinical organ-based sequences), study period duration, and first-attempt USMLE Step 1 score from internal institutional databases. Using National Institutes of Health guidelines,³⁷ students underrepresented in medicine (URiM) were identified on the basis of self-reported race and ethnicity. The lead investigator (J.P.) paired this information with survey respondent and nonrespondent cases, and deidentified data for subsequent analysis.

Statistical analysis

We performed statistical analysis using SPSS version 24.0 (SPSS Inc., Chicago, Illinois). Chi-square tests were used to assess differences in count data. Independent samples *t* tests were used to compare means. One-way analysis

of variance with contrasts was used to test linear trends in Step 1 score across categorical groups. Bivariate correlations used Pearson or point-biserial correlation. For bivariate data tables, predictor groupings were created at natural cut points for illustrative purposes and were not used for subsequent regressions. All *P* values were two sided, with the alpha level set at 0.05.

We implemented multiple linear regression analyses based on our a priori conceptual model. Control variables included sex, URiM status, MCAT score, preclinical score, and score goal. Variables of interest included “early study” (any dedicated Step 1 study before the protected study period), study period duration and hours per day studied, review book passes (number of times reading cover-to-cover), and unique and repeated questions. Variables with significant zero-order correlations were entered simultaneously in each block. Cases with missing data were excluded listwise. We assessed violations of linearity, normality, noncollinearity, and homoscedasticity.

Results

Respondents

Of 332 medical students, 274 (82.5%) responded to the study-behaviors survey. There were no significant differences in baseline characteristics between the 2014 and 2015 respondents (Table 1). Compared with respondents' scores, nonrespondents' Step 1 scores were 12.2 points lower ($P < .001$) and nonrespondents' preclinical scores were 2.2 points lower ($P < .001$). Nevertheless, 20.1% ($n = 55$) of respondents' scores fell within the lowest overall score quartile (≤ 224).

Study timing and intensity

While most respondents ($n = 235$; 90.4%) reported focusing primarily on their preclinical course work during the academic year, with Step 1 a secondary concern, 77.0% ($n = 211$) conducted dedicated Step 1 studying prior to their study period (“early study”). Some students ($n = 38$; 13.9%) began Step 1 studying in their first year, while most ($n = 173$; 63.1%) began in their second year. During the study period, students on average studied 11.0 hours per day (SD 2.1, range 4.5–17.5) over a period of 35.3 days (SD 6.2, range 11–98).

Table 1

Respondent and Nonrespondent Medical Student Characteristics by Study Year, University of Michigan Medical School, 2014 and 2015

Characteristic	Respondents			P value ^a	Nonrespondents	
	2014 (n = 135)	2015 (n = 139)	All (n = 274)		All (n = 58)	P value ^b
Female sex, no. (%)	70 (51.9)	72 (51.8)	142 (51.8)	NS	24 (41.4)	NS
URiM, no. (%)	14 (10.4)	19 (13.7)	33 (12.0)	NS	13 (22.4)	NS
MCAT score, mean (SD)	35.0 (3.3)	34.7 (3.1)	34.9 (3.2)	NS	34.0 (3.1)	NS
Preclinical score, mean (SD)	90.7 (4.2)	90.7 (4.1)	90.7 (4.1)	NS	88.5 (4.4)	< .001
Score goal, mean (SD)	236.7 (14.8)	237.1 (14.3)	236.9 (14.5)	NS	—	—
Step 1 score, mean (SD)	238.9 (16.4)	236.7 (16.7)	237.8 (16.6)	NS	225.6 (16.9)	< .001

Abbreviations: NS indicates not significant at $\alpha = 0.05$; URiM, underrepresented in medicine; MCAT, Medical College Admission Test; SD, standard deviation; Step 1, United States Medical Licensing Examination Step 1.

^aComparison of 2014 and 2015 using chi-square tests (for counts) or independent samples *t* tests (for means).

^bComparison of nonrespondents and respondents with null hypothesis of no difference.

Study resources

A small number of students ($n = 49$; 17.9%) used lecture notes or video recordings from their preclinical courses to study for Step 1, using them lightly and finding them only somewhat useful (Table 2). In contrast, students reported predominant use of third-party Step 1 study resources: The review book and UWorld Step 1 question bank (“UWorld”) were used by over 99% of students. *Pathoma*,³⁸ a pathology

textbook with accompanying online lectures, was used heavily by many. Other common resources included Goljan audio lectures (a 30-hour series of lecture recordings of Dr. Edward Goljan, Professor of Pathology, Oklahoma State University Center for Health Sciences), the USMLE-Rx and Kaplan question banks, and Firecracker³⁹ spaced-repetition review. These latter resources were used less heavily and were identified as less useful to most

students compared with the review book, UWorld, and *Pathoma*. Finally, students reported taking an average of 2.8 self-assessment exams (sponsored by either UWorld⁴⁰ or the National Board of Medical Examiners⁴¹).

Students read the review book on average 2.1 times (SD 0.8; 95% confidence interval [CI]: 2.0–2.2), while completing 3,597 total practice questions (SD 1,611; 95% CI: 3,404–3,790), of which 2,961 completed questions were unique (SD 1,026; 95% CI: 2,838–3,084) and 636 were repeated (SD 871; 95% CI: 531–740).

Table 2

Top 10 Most Frequently Used USMLE Step 1 Study Resources by Medical Students (n = 274) at the University of Michigan Medical School, 2014 and 2015

Resource	Used resource, no. (%)	Heavy usage, ^a no. (%)	Very useful, ^b no. (%)	Recommend early use, ^c no. (%)
<i>First Aid</i> review book	271 (99.3)	256 (94.5)	244 (93.8)	181 (72.4)
UWorld Qbank	271 (99.3)	258 (95.2)	253 (98.1)	75 (30.1)
<i>Pathoma</i>	247 (90.5)	141 (57.1)	191 (83.0)	201 (88.2)
Goljan audio	145 (53.1)	25 (17.2)	28 (20.6)	50 (44.2)
USMLE-Rx Qbank	70 (25.6)	13 (18.6)	16 (20.8)	43 (76.8)
Firecracker	52 (19.0)	5 (9.6)	4 (8.7)	32 (100)
Lecture notes / videos	49 (17.9)	2 (4.1)	3 (5.7)	9 (36.0)
Kaplan Qbank	33 (12.1)	4 (12.1)	5 (11.9)	15 (65.2)
Self-assessments^d				
UWSA	221 (81.0)	—	116 (55.5)	4 (2.0)
NBME CBSSA	167 (60.9)	—	88 (54.0)	3 (2.0)

Abbreviations: USMLE indicates United States Medical Licensing Examination; *First Aid*, *First Aid for the USMLE Step 1*; UWorld Qbank, UWorld Step 1 question bank; *Pathoma*, *Pathoma Fundamentals of Pathology*; Goljan audio, audio lectures by Dr. Edward Goljan; USMLE-Rx Qbank, USMLE-Rx Step 1 question bank; Kaplan Qbank, Kaplan Step 1 question bank; UWSA indicates UWorld Step 1 self-assessment; NBME CBSSA, National Board of Medical Examiners Comprehensive Basic Science Self-Assessment.

^aOf those who used resource at all, number who used the resource “heavily.”

^bOf those who used resource at all, number who rated the resource “very useful.”

^cOf those who recommended resource use, number recommending use begin prior to the study period.

^dThe mean number of UWSA self-assessments students completed was 1.6 (SD 0.9), and the mean number of NBME CBSSA self-assessments students completed was 1.2 (SD 1.2).

Bivariate association between behaviors and Step 1 score

In bivariate analyses, the control variables male sex, MCAT score, preclinical score, and score goal were positively associated with Step 1 score (Table 3). URiM status was not significantly associated with Step 1 score. Study behaviors showing positive associations with Step 1 scores included early study, greater review book usage, and increased completion of unique or repeated practice questions. Study period hours per day studying and duration were not associated with Step 1 scores. Several variables associated with Step 1 score were also associated with preclinical score, a hypothesized confounder. Bivariate regression models (Table 4 [top]) identified the same significant associations noted in Table 3.

Multiple linear regression for Step 1 score

Our baseline control model including sex, MCAT score, preclinical score,

Table 3

Bivariate Associations: Preclinical Scores and USMLE Step 1 Scores by Medical Student Demographic and Study Behavior, University of Michigan Medical School, 2014 and 2015 (n = 274)

Variables ^b	No. (%)	Preclinical score ^a		Step 1 score ^a	
		Mean (SD)	P value ^c	Mean (SD)	P value
Sex			NS		.02
Male	132 (48.2)	90.6 (4.2)		240.3 (17.0)	
Female	142 (51.8)	90.8 (4.1)		235.5 (15.9)	
URiM			NS		NS
No	241 (88.0)	90.9 (4.0)		238.0 (17.0)	
Yes	33 (12.0)	89.4 (4.9)		236.2 (13.5)	
MCAT score			< .001		< .001
≤ 32	60 (22.3)	89.7 (3.5)		227.1 (15.5)	
33–37	133 (49.4)	90.2 (3.8)		237.5 (14.7)	
≥ 37	76 (28.3)	92.2 (4.7)		247.0 (15.2)	
Preclinical score					< .001
≤ 88	73 (26.6)	—		227.2 (16.4)	
88–94	137 (50.0)	—		237.1 (14.6)	
≥ 94	64 (23.4)	—		251.4 (10.3)	
Score goal			< .001		< .001
≤ 230	75 (31.9)	89.4 (4.5)		228.3 (17.9)	
230–245	90 (38.3)	90.2 (3.6)		237.6 (13.5)	
≥ 245	70 (29.8)	92.3 (4.1)		248.8 (11.9)	
Early study^d			NS		.03
No	63 (23.0)	91.3 (4.4)		233.8 (18.5)	
Yes	211 (77.0)	90.5 (4.0)		239.0 (15.8)	
Study period hours studied per day			NS		NS
≤ 10	92 (33.7)	90.4 (4.2)		237.7 (16.3)	
10–12	107 (39.2)	90.7 (4.1)		236.9 (17.7)	
≥ 12	74 (27.1)	91.2 (4.1)		239.4 (15.4)	
Study period duration			NS		NS
≤ 31 days	34 (12.4)	89.6 (3.4)		235.6 (18.4)	
32–36 days	162 (59.1)	91.2 (3.8)		238.1 (16.3)	
≥ 37 days	78 (28.5)	90.1 (4.8)		238.1 (16.6)	
Review book³² complete passes			< .001		< .001
≤ 2 passes	157 (59.0)	89.9 (4.2)		233.2 (16.4)	
> 2 passes	109 (41.0)	91.7 (3.8)		245.1 (14.4)	
Unique practice questions completed			NS		< .001
0–2,400	67 (24.8)	90.2 (3.3)		230.5 (17.2)	
2,401–3,200	134 (49.6)	91.0 (4.4)		238.3 (15.3)	
> 3,200	69 (25.6)	90.5 (4.2)		244.8 (14.5)	
Repeated practice questions completed			NS		.01
0	132 (48.9)	90.6 (4.3)		235.2 (17.3)	
1–1,199	62 (23.0)	91.1 (4.2)		240.4 (16.1)	
≥ 1,200	76 (28.1)	90.5 (3.8)		240.9 (14.1)	

Abbreviations: USMLE indicates United States Medical Licensing Examination; SD, standard deviation; NS, not significant at $\alpha = 0.05$; URiM, Underrepresented in Medicine; MCAT, Medical College Admission Test.

^aThe mean preclinical score for all students was 90.7 (SD 4.1), and the mean Step 1 score for all students was 237.8 (SD 16.6).

^bSex, URiM status, MCAT score, Preclinical score, and Score goal are all **control variables**; Early study, study period hours studied per day, Study period duration, review book³² complete passes, Unique practice questions completed, and Repeated practice questions completed are all **variables of interest**.

^cP value represents independent samples *t* test (two groups) or analysis of variance for trend (three groups). P value not corrected for multiple comparisons.

^dEarly study indicates any dedicated Step 1 studying prior to the protected study period.

Table 4

Simple and Multiple Linear Regression Models for First-Attempt USMLE Step 1 Score, University of Michigan Medical School, 2014 and 2015

Simple linear regression	No. of students included in model	Variable	Coefficient (95% CI) ^a	Pearson <i>r</i>	<i>P</i> value ^b
Model 1	274	Female sex	-4.77 (-8.68, -0.85)	-0.14	.02
Model 2	269	MCAT score	2.16 (1.60, 2.73)	0.42	< .001
Model 3	274	Preclinical score	2.25 (1.85, 2.65)	0.56	< .001
Model 4	235	Score goal	0.45 (0.39, 0.65)	0.45	< .001
Model 5	274	Early study	5.17 (0.51, 9.82)	0.13	.03
Model 6	266	Review book ³² passes	6.60 (4.30, 8.89)	0.33	< .001
Model 7	270	Unique questions ^c	0.58 (0.41, 0.76)	0.37	< .001
Model 8	270	Repeat questions ^c	0.34 (0.12, 0.56)	0.18	< .01
Multiple linear regression	No. of students included in model	Variable	Coefficient (95% CI) ^a	Standardized β	<i>P</i> value ^b
Control model^d	232	(Constant)	-41.5 (-80.9, -2.1)		
		Female sex	-1.82 (-5.10, 1.46)	-0.06	NS
		MCAT score	0.95 (0.45, 1.51)	0.19	< .001
		Preclinical score	1.92 (1.53, 2.31)	0.49	< .001
		Score goal	0.31 (0.19, 0.43)	0.27	< .001
Master model^e	225	(Constant)	-39.5 (-76.7, -2.2)		
		Female sex	-0.90 (-4.00, 2.19)	-0.03	NS
		MCAT score	1.10 (0.61, 1.60)	0.21	< .001
		Preclinical score	1.96 (1.59, 2.33)	0.49	< .001
		Score goal	0.18 (0.06, 0.30)	0.16	< .001
		Early study	4.23 (0.55, 7.91)	0.11	.03
		Review book ³² passes	2.29 (0.32, 4.26)	0.11	.02
		Unique questions ^c	0.35 (0.18, 0.52)	0.21	< .001
Repeat questions ^c	0.01 (-0.17, 0.19)	0.01	NS		

Abbreviations: USMLE indicates United States Medical Licensing Examination; CI, confidence interval; MCAT, Medical College Admission Test.

^aUnstandardized beta coefficient, with 95% CI.

^b*P* value indicates coefficient significance. NS indicates not significant at $\alpha = 0.05$ level.

^cAll question variables are in 100-question units.

^dIn the control model, $R = 0.70$, adjusted $R^2 = 0.48$, and $P < .001$.

^eIn the master model, $R = 0.76$, adjusted $R^2 = 0.56$, $\Delta R^2 = 0.092$, and $P < .001$.

and score goal accounted for 47.7% of Step 1 performance variation (Table 4 [bottom], $P < .001$). In this model, sex was no longer significantly associated with Step 1 scores. Adding early study, review book passes, unique question usage, or repeat question usage separately to the control model accounted for an additional 1.9% to 7.1% of variance, with each variable remaining statistically significant (Supplemental Digital Table 1, <http://links.lww.com/ACADMED/A479>). When added simultaneously to the control model, these variables combined to account for an additional 9.2% of Step 1 performance variation (Table 4 [bottom], master model, $P < .001$). Early study, review book passes, and unique questions completed remained significant in the master model, while sex and repeat questions

were not significant. Control variable beta coefficients were largely unchanged from the baseline control model. When controlling for the other covariates, early study was associated with a 4.2-point Step 1 score increase (95% CI: 0.6–7.9), each additional cover-to-cover review book reading a 2.3-point increase (95% CI: 0.3–4.3), and completion of 286 additional unique questions a 1.0-point increase (95% CI: 0.5–1.5). URiM status, study period hours per day, and study period duration were not included in the models, as their zero-order correlations were not statistically significant; sensitivity analyses including these variables did not substantially alter the model (data not shown), nor did exclusion of self-identified score goal (Supplemental Digital Table 2, <http://links.lww.com/ACADMED/A479>).

Discussion

“Early” self-directed Step 1 studying the norm

Prior to the protected Step 1 study period, most students at our institution focused primarily on their preclinical course work, yet also engaged in self-directed Step 1 studying. This “early” study phenomenon—which was associated with improved Step 1 performance in multiple regression models—may be a response to the increasing importance of Step 1 in resident selection, in conjunction with a preclinical curriculum that does not “teach to the test.” The effect of early study on Step 1 outcomes has not been previously described, although similar effects were seen for the Comprehensive Osteopathic Medical Licensing Examination of the United States.³³

Minimal use of course work review, ubiquitous use of third-party Step 1 resources

Few students reviewed formal course work (e.g., notes or lectures) to prepare for Step 1. Rather, students preferred third-party resources at additional monetary cost. Survey comments noted that a nearly ubiquitous 700-page review book,³² while imperfect, served as a unifying “content outline” covering “high-yield” concepts across the otherwise overwhelming scope of Step 1. Students also used a practice-question-heavy study strategy, replicating the actual Step 1 exam with case-based vignettes requiring multistep reasoning.

Question banks also provide detailed feedback and facilitate spaced repetition and repeated testing that may enhance concept recall.⁴² On the basis of this work, prior studies,^{24,31} and conversations with peer programs, we suspect that intense use of third-party Step 1 preparation resources is common practice among medical students at most U.S. medical schools. Additional work is needed to understand the appeal of these resources and to explore the feasibility and appropriateness of formal incorporation of these or related tools into core medical curricula.

MCAT, preclinical performance, and score goal are important covariates

Similar to previous studies,^{6–10} we found that MCAT scores and preclinical performance were associated with Step 1 performance. We extended prior work in demonstrating the significance of preclinical performance even when including multiple covariates, supporting our medical school’s emphasis that learning content throughout the preclinical years builds an important scientific foundation that also pays dividends in Step 1 performance.

We also described a novel covariate, students’ self-identified score goal (i.e., the score above which students hoped to score, as conceived before beginning the intensive study period). Students piloting our study emphasized that their score goal shaped their overall study intensity; its persistence as a significant covariate merits additional research and consideration when advising students.

Early study and test preparation resource use are associated with Step 1 performance

We confirmed our hypothesis that early study and use of exam-specific resources were positively associated with Step 1 scores, even when controlling for likely covariates. For example, our model estimates that a student who studied early, completed an additional review book pass, and went through an additional question bank in its entirety would score 13.5 points higher than a student who did not engage in these behaviors. These effect sizes are statistically meaningful, as a score increase greater than 10 points for an individual test taker is significant at 95% confidence.⁴³ Moreover, these effect sizes are meaningful in the context of an increasingly Step 1–reliant residency application process, where small score differences can put an applicant above a screening cutoff. Our effect size for unique question usage—a 1.0-point Step 1 score increase per 286 additional questions—is comparable to prior studies showing a 1-point increase per 200 to 445 questions.^{23,25,26}

Contrary to our hypothesis, amount studied during the dedicated study period was not significantly associated with Step 1 performance, suggesting that identifying approaches for “smarter” study may be more important than increased study amounts. It is possible that study efficiency or the amount of early study before the dedicated study period are associated with Step 1 performance, but we were unable to assess these behaviors accurately in this study.

We distinguished between unique and repeated study questions for the first time in the Step 1 literature. Emerging evidence suggests that time spent testing (e.g., question bank usage) is superior to repeated studying (e.g., reading review books) for long-term content retention.^{44,45} However, the distinction between testing on unique versus repeated identical questions is not well described. Here, repeated testing on identical questions was inferior to testing on new questions covering similar Step 1 content, suggesting an important area for further learning science research, as this approach is broadly applicable in medical education.

Finally, we showed that rigorously controlling for multiple likely

confounders—absent from many prior Step 1 studies—is critical for such cross-sectional studies. Our control model accounted for the majority of explained Step 1 performance variance, and the effect sizes of study behaviors were attenuated when considering control variables.

Limitations

This was a single-institution study, which may limit generalizability to institutions with differing curricular approaches, although studies suggest that students engage in similar behaviors at peer institutions.^{23–25} The study was cross-sectional, limiting conclusions on causality and directionality. Behaviors were self-reported retrospectively, raising the possibility of recall bias; however, pilot students reported that their Step 1 study behaviors were intentional and regimented, aiding accurate recall. Sampling bias is possible, as nonrespondents had lower Step 1 scores than respondents, although our high response rate and sampling of many students in the lowest score quartile attenuate the impact of this bias. One hourlong information session for students was conducted between study years detailing findings from the first year of our study, such as the most commonly used resources. We speculate that the effect of this session on behaviors in our second cohort was minimal, as we did not detect substantial heterogeneity across study years. Finally, as our analysis was planned a priori, we did not adjust *P* values for multiple comparisons; however, type I errors may exist, especially for *P* values near the significance threshold.

Implications and future work

This student-initiated study describes in detail the nature and impact of student self-directed USMLE Step 1 study behaviors, providing preliminary evidence to guide students and medical schools alike. Many medical schools, including our own, evidence an institutional culture of not “teaching to the test.” Importantly, student self-directed co-curricular Step 1 studying and use of third-party resources suggest a “parallel” Step 1 curriculum, with preclinical course work perceived as insufficient alone for preparation.

While the USMLE Step 1 serves an important purpose in licensure, it

correlates poorly with resident clinical skills,² predominantly assessing the “medical knowledge” competency.⁴⁶ Thus, the monetary and time investment of Step 1 preparation—in our sample averaging nearly five 80-hour weeks during the study period alone—represent a significant opportunity cost for both students and educators. Like others,³ we advocate reassessing the use of the USMLE Step 1 exam in residency candidate selection, which appears to be driving students to seek scores well above the pass/fail threshold. In the meantime, Step 1 performance remains an important outcome for students, and this work provides insight into self-directed study behaviors that may benefit performance. Aligning core preclinical curricula with students’ parallel Step 1 curriculum may yield benefits to students and educators, but should be balanced against overemphasizing Step 1 performance as a meaningful marker of overall trainee quality.

In the future, we plan to explore the relationship between Step 1 preparation behaviors and performance in core clinical clerkships, as we do not know how Step 1 study approaches interact with longer-term measures of medical knowledge retention. Additionally, we hope to conduct a prospective multi-institutional study incorporating local curricular factors to determine more broadly generalizable value-added study behaviors. Finally, the medical education research community should work toward identifying alternative approaches for screening residency applicants for interview, which might reign in students’ Step 1 performance stress and “parallel” curriculum.

Acknowledgments: The authors wish to thank Dr. Linda Li and Mr. Joey Linzey for their helpful review of the survey instrument; Dr. Eric Middleton, Ms. Amy Tschirhart, and Ms. Tania Reis for input on academic advising factors; Dr. James Cranford for helpful discussions of the statistical approach; Dr. Paula Ross for critical reading; and the numerous students who participated in the study.

Funding/Support: Monetary incentives to promote survey responses for this study were provided through faculty development funds from the Department of Learning Health Sciences at the University of Michigan Medical School, Ann Arbor, Michigan.

Other disclosures: None reported.

Ethical approval: This study was approved as exempt by the University of Michigan

Health Sciences Institutional Review Board (HUM00090072).

Previous presentations: The authors’ needs assessment and survey methods were presented as an oral presentation at the Association of American Medical Colleges (AAMC) Central Group on Educational Affairs regional spring meeting in April 2016 in Ann Arbor, Michigan, and the AAMC Research in Medical Education annual meeting in November 2016 in Seattle, Washington.

References

- National Residency Matching Program. Results of the 2016 NRMP Program Director Survey. <http://www.nrmp.org/wp-content/uploads/2016/09/NRMP-2016-Program-Director-Survey.pdf>. Accessed July 19, 2017.
- McGaghie WC, Cohen ER, Wayne DB. Are United States Medical Licensing Exam Step 1 and 2 scores valid measures for postgraduate medical residency selection decisions? *Acad Med.* 2011;86:48–52.
- 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;91:12–15.
- Julian ER. Validity of the Medical College Admission Test for predicting medical school performance. *Acad Med.* 2005;80:910–917.
- Basco WT Jr, Way DP, Gilbert GE, Hudson A. Undergraduate institutional MCAT scores as predictors of USMLE step 1 performance. *Acad Med.* 2002;77(10 suppl):S13–S16.
- Donnon T, Paolucci EO, Violato C. The predictive validity of the MCAT for medical school performance and medical board licensing examinations: A meta-analysis of the published research. *Acad Med.* 2007;82:100–106.
- White CB, Dey EL, Fantone JC. Analysis of factors that predict clinical performance in medical school. *Adv Health Sci Educ Theory Pract.* 2009;14:455–464.
- Haight SJ, Chibnall JT, Schindler DL, Slavin SJ. Associations of medical student personality and health/wellness characteristics with their medical school performance across the curriculum. *Acad Med.* 2012;87:476–485.
- Coumarbatch J, Robinson L, Thomas R, Bridge PD. Strategies for identifying students at risk for USMLE Step 1 failure. *Fam Med.* 2010;42:105–110.
- Gohara S, Shapiro JI, Jacob AN, et al. Predictors of success on the United States Medical Licensing Examinations (USMLE). *Learn Assist Rev (TLAR).* 2011;16(1):11–20.
- Morrison CA, Ross LP, Fogle T, Butler A, Miller J, Dillon GF. Relationship between performance on the NBME Comprehensive Basic Sciences Self-Assessment and USMLE Step 1 for U.S. and Canadian medical school students. *Acad Med.* 2010;85(10 suppl):S98–S101.
- Sawhill A, Butler A, Ripkey D, et al. Using the NBME self-assessments to project performance on USMLE Step 1 and Step 2: Impact of test administration conditions. *Acad Med.* 2004;79(10 suppl):S55–S57.
- Kleshinski J, Khuder SA, Shapiro JI, Gold JP. Impact of preadmission variables on USMLE Step 1 And Step 2 performance. *Adv Health Sci Educ Theory Pract.* 2009;14:69–78.
- Cuddy MM, Swanson DB, Clauser BE. A multilevel analysis of examinee gender and USMLE step 1 performance. *Acad Med.* 2008;83(10 suppl):S58–S62.
- Andriole DA, Jeffe DB. Prematriculation variables associated with suboptimal outcomes for the 1994–1999 cohort of US medical school matriculants. *JAMA.* 2010;304:1212–1219.
- Hecker K, Violato C. How much do differences in medical schools influence student performance? A longitudinal study employing hierarchical linear modeling. *Teach Learn Med.* 2008;20:104–113.
- Hoffman K, Hosokawa M, Blake R Jr, Headrick L, Johnson G. Problem-based learning outcomes: Ten years of experience at the University of Missouri—Columbia School of Medicine. *Acad Med.* 2006;81:617–625.
- Lieberman SA, Ainsworth MA, Asimakis GK, et al. Effects of comprehensive educational reforms on academic success in a diverse student body. *Med Educ.* 2010;44:1232–1240.
- Wilkerson L, Wimmers P, Doyle LH, Uijtdehaage S. Two perspectives on the effects of a curriculum change: Student experience and the United States Medical Licensing Examination, Step 1. *Acad Med.* 2007;82(10 suppl):S117–S120.
- Yoshida H, Sims KL. Education initiatives for improved United States Medical Licensing Examination performance. *Med Sci Educ.* 2013;23(4):637–647.
- Holtman MC, Swanson DB, Ripkey DR, Case SM. Using basic science subject tests to identify students at risk for failing step 1. *Acad Med.* 2001;76(10 suppl):S48–S51.
- Burns ER, Garrett J. Student failures on first-year medical basic science courses and the USMLE Step 1: A retrospective study over a 20-year period. *Anat Sci Educ.* 2015;8:120–125.
- Kumar AD, Shah MK, Maley JH, Evron J, Gyftopoulos A, Miller C. Preparing to take the USMLE Step 1: A survey on medical students’ self-reported study habits. *Postgrad Med J.* 2015;91:257–261.
- Bonasso P, Lucke-Wold B, Reed Z, et al. Investigating the impact of preparation strategies on USMLE Step 1 performance. *MedEdPublish.* 2015;4(1):5.
- Johnson JH, Jordan EJ, Burton W, Silbiger S. Are questions the answer? The effect of popular study resources on USMLE Step 1 performance. Oral abstract presented at: Association of American Medical Colleges Annual Meeting; November 1–7, 2012; San Francisco, CA. <https://members.aamc.org/eweb/upload/Are%20Questions%20the%20Answer%20PPT%2011-6-12%2010PM.pdf>. Accessed July 19, 2017.
- Deng F, Gluckstein JA, Larsen DP. Student-directed retrieval practice is a predictor of medical licensing examination performance. *Perspect Med Educ.* 2015;4:308–313.
- Kerfoot BP, DeWolf WC, Masser BA, Church PA, Federman DD. Spaced education improves the retention of clinical knowledge by medical students: A randomised controlled trial. *Med Educ.* 2007;41:23–31.
- Thadani RA, Swanson DB, Galbraith RM. A preliminary analysis of different approaches to preparing for the USMLE step 1. *Acad Med.* 2000;75(10 suppl):S40–S42.
- Pohl CA, Robeson MR, Hojat M, Veloski JJ. Sooner or later? USMLE Step 1 performance

- and test administration date at the end of the second year. *Acad Med.* 2002;77(10 suppl): S17–S19.
- 30 Werner LS, Bull BS. The effect of three commercial coaching courses on Step One USMLE performance. *Med Educ.* 2003;37:527–531.
- 31 Strowd RE, Lambros A. Impacting student anxiety for the USMLE Step 1 through process-oriented preparation. *Med Educ Online.* 2010;15(1). <http://www.tandfonline.com/doi/full/10.3402/meo.v15i0.4880>. Published February 24, 2010. Accessed July 19, 2017.
- 32 Le T, Bhushan V, Sochat M, eds. *First Aid for the USMLE Step 1 Exam*, 2015 edition. New York, NY: McGraw-Hill Education; 2015.
- 33 Vora A, Maltezos N, Alfonzo L, Hernandez N, Calix E, Fernandez MI. Predictors of scoring at least 600 on COMLEX-USA Level 1: Successful preparation strategies. *J Am Osteopath Assoc.* 2013;113:164–173.
- 34 UWorld. USMLE Step 1 Qbank. https://www.uworld.com/step1/step1_qbank.aspx. Accessed July 19, 2017.
- 35 USMLE-Rx. Step 1 QMax. <https://www.usmle-rx.com/content/step-1-qmax>. Accessed July 19, 2017.
- 36 Kaplan Medical. USMLE Step 1 Qbank. <https://www.kaptest.com/medical-prep/usmle/usmle-prep-course/step-1-qbank>. Accessed July 19, 2017.
- 37 National Institutes of Health. Notice of NIH's interest in diversity. <http://grants.nih.gov/grants/guide/notice-files/NOT-OD-15-053.html>. Published January 12, 2015. Accessed July 19, 2017.
- 38 Sattar HA. Pathoma: Fundamentals of pathology. <http://www.pathoma.com/>. Accessed July 19, 2017.
- 39 Firecracker. <http://www.firecracker.me/>. Accessed July 19, 2017.
- 40 UWorld. USMLE Step 1 self-assessment exam. https://www.uworld.com/step1/step1_exam.aspx. Accessed July 19, 2017.
- 41 National Board of Medical Examiners. National Board of Medical Examiners (NBME) self-assessment services. <https://nsas.nbme.org/home>. Accessed July 19, 2016.
- 42 Delaney PF, Verkoeijen PPJL, Spigel A. Spacing and testing effects: A deeply critical, lengthy, and at times discursive review of the literature. In: Ross B, ed. *The Psychology of Learning and Motivation*. Vol 53. Burlington, VT: Academic Press; 2010:63–148.
- 43 United States Medical Licensing Examination. USMLE score interpretation guidelines. http://www.usmle.org/pdfs/transcripts/USMLE_Step_Examination_Score_Interpretation_Guidelines.pdf. Updated June 1, 2016. Accessed September 1, 2016.
- 44 Roediger HL 3rd, Butler AC. The critical role of retrieval practice in long-term retention. *Trends Cogn Sci.* 2011;15:20–27.
- 45 Larsen DP, Butler AC, Roediger HL 3rd. Repeated testing improves long-term retention relative to repeated study: A randomised controlled trial. *Med Educ.* 2009;43:1174–1181.
- 46 United States Medical Licensing Examination. USMLE Step 1 content outline and specifications. <http://www.usmle.org/step-1/#content-outlines>. Accessed July 19, 2017.