

Moving the United States Medical Licensing Examination Step 1 After Core Clerkships: An Outcomes Analysis

Daniel Jurich, PhD, Michelle Daniel, MD, MHPE, Miguel Paniagua, MD, Amy Fleming, MD, MHPE, Victoria Harnik, PhD, Arnyce Pock, MD, Aubrie Swan-Sein, PhD, EdM, Michael A. Barone, MD, MPH, and Sally A. Santen, MD, PhD

Abstract

Purpose

Schools undergoing curricular reform are reconsidering the optimal timing of Step 1. This study provides a psychometric investigation of the impact on United States Medical Licensing Examination Step 1 scores of changing the timing of Step 1 from after completion of the basic science curricula to after core clerkships.

Method

Data from four schools that recently moved the examination were analyzed in a pre–post format using examinee scores from three years before and after the change. The sample included scores from 2008 through 2016. Several confounders

were addressed, including rising national scores and potential differences in cohort abilities using deviation scores and analysis of covariance (ANCOVA) controlling for Medical College Admission Test (MCAT) scores. A resampling procedure compared study schools' score changes versus similar schools' in the same time period.

Results

The ANCOVA indicated postchange Step 1 scores were higher compared with prechange (adjusted difference = 2.67; 95% confidence interval: 1.50–3.83, $P < .001$; effect size = 0.14) after adjusting for MCAT scores and rising

national averages. The average score increase in study schools was larger than changes seen in similar schools. Failure rates also decreased from 2.87% ($n = 48$) pre change to 0.39% ($n = 6$) post change ($P < .001$).

Conclusions

Results suggest moving Step 1 after core clerkships yielded a small increase in scores and a reduction in failure rates. Although these small increases are unlikely to represent meaningful knowledge gains, this demonstration of "noninferiority" may allow schools to implement significant curricular reforms.

Medical education in the United States needs a fundamental redesign,¹ yet disruptive innovation can be fraught with uncertainty.² The authors of a recent seminal article on medical education, commissioned by the Carnegie Foundation to commemorate the centennial of the Flexner Report, suggested that "ossified curricular structures" and "archaic assessment

practices" are challenges to institutions seeking to enact meaningful curricular reform.³ As medical schools undergo curricular revisions that break down the traditional 2 + 2 (two-year basic science, two-year clinical) paradigm, questions have arisen concerning the optimal timing for students to take Step 1 of the United States Medical Licensing Examination (USMLE).

Historically, students have taken Step 1 immediately following completion of their basic science curricula.⁴ The majority of U.S. allopathic medical schools require students to pass this examination before advancing to clinical training.⁵ A small but growing number of schools, however, have changed the timing of Step 1 to take place after students complete the core clerkships.⁶ This change can facilitate earlier entry into clinical environments and allow for increased innovation surrounding health systems science and other nonbasic science curricula in the preclinical time frame. Furthermore, the change may promote longer-term retention of foundational concepts, by using a

national standardized assessment (i.e., USMLE) to encourage integrated basic science learning in clinical contexts. Repositioning the Step 1 examination may also help promote the development of clinicians with stronger foundational knowledge by using the motivation of assessment to drive learning.^{6,7}

Given the role Step 1 scores can play in determining residency placements, administrators, faculty, and students are hesitant to embrace a change in its timing in the curriculum in the absence of clear outcomes data. At a minimum, institutions contemplating a change in the placement of Step 1 within the curriculum would want to know that moving the examination after clerkships will not adversely affect their students' performance (i.e., the change is, at a minimum, "noninferior" to placement before the clerkships). In a previous article, Daniel et al⁶ describe modest but promising gains in Step 1 scores when the examination is taken after core clerkships. The data, however, were reported in aggregate without controlling for potential confounders, such as

Please see the end of this article for information about the authors.

Correspondence should be addressed to Michelle Daniel, University of Michigan Medical School, 6123 Taubman Health Sciences Library, 1135 Catherine St., SPC 5726, Ann Arbor, MI 48109; telephone: (401) 525-0251; e-mail: micdan@med.umich.edu; Twitter:@Emergdoc1975.

Written work prepared by employees of the Federal Government as part of their official duties is, under the U.S. Copyright Act, a "work of the United States Government" for which copyright protection under Title 17 of the United States Code is not available. As such, copyright does not extend to the contributions of employees of the Federal Government.

Acad Med. 2019;94:371–377.

First published online September 11, 2018
doi: 10.1097/ACM.0000000000002458

Supplemental digital content for this article is available at <http://links.lww.com/ACADMED/A600>.

nationally rising Step 1 scores. To date, a strong psychometric investigation into the effects of delaying Step 1 has not been conducted. The objective of this study was to determine the effect of changing the timing of Step 1 in the curriculum on Step 1 scores, using data from four schools that moved the examination after core clerkships.

Method

Sample

We examined the change in Step 1 scores at four Liaison Committee on Medical Education (LCME)-accredited schools cited by Daniel et al⁶ where students currently sit for USMLE Step 1 after core clerkships. These four schools were chosen because they had recently moved

the exam and had an adequate number of years of data post change to evaluate. The other schools cited by Daniel et al⁶ either moved the exam several years ago or did not yet have enough data post change to study. Data three years prior to and three years after changing the timing of Step 1 were examined for each school, with the exception of one school that changed the timing of Step 1 in 2015, and thus only had two years of data post change (i.e., 2015 and 2016). The study sample included students within each school who took the test between 2008 and 2016. Our aim was to establish a baseline of performance without exceeding a reasonable time frame for comparison. Table 1 details the curricular and assessment characteristics of these four schools.

Across all four schools, the sample of students who first attempted USMLE Step 1 within the three years prior to the curricular change contained 1,668 examinees. The sample within the three years after the change contained 1,529 examinees. These two groups of examinees had similar demographic characteristics with regard to age, gender, and self-reported ethnicity. Medical Science Training Program students ($N = 209$) were removed from the analysis because these students had taken Step 1 prior to clerkships.

Analyses

All analyses were conducted on Step 1 scaled scores. Scaled scores are equated, a statistical process that maintains comparability of scores across time,

Table 1

Curricular and Assessment Characteristics of Four Schools Included in a Study of the Effects of Moving the United States Medical Licensing Examination Step 1 From After Basic Science Curricula to After Core Clerkships

	School 1	School 2	School 3	School 4
Class size	120/year	172/year	100/year	170/year
Year Step 1 moved after clerkships	2013	2014	2015	2012
Length of preclerkship (foundational science) curriculum before change	76 weeks	75 weeks	76 weeks	80 weeks
Length of preclerkship (foundational science) curriculum after change	60 weeks	57 weeks	49 weeks	57 weeks
Details of the new preclerkship curriculum	<ul style="list-style-type: none"> • Single pass • Normal taught simultaneously with abnormal 	<ul style="list-style-type: none"> • Single pass • Normal taught simultaneously with abnormal 	<ul style="list-style-type: none"> • Single pass • Normal taught simultaneously with abnormal 	<ul style="list-style-type: none"> • Single pass • Normal taught simultaneously with abnormal
Details of the new preclerkship assessment structure	<ul style="list-style-type: none"> • Pass / fail • Faculty developed • Essay and MCQ exams at the end of each module • No NBME exams • No cumulative exam 	<ul style="list-style-type: none"> • Honors / pass / fail • Faculty-developed MCQs supplement the NBME exams • Customized NBME exams used for each module, and on midterms and finals • No cumulative exam 	<ul style="list-style-type: none"> • Pass / fail • Faculty developed • Essay, short answer and laboratory practical exams at the end of each course • Customized NBME exams used in each course • No cumulative exam 	<ul style="list-style-type: none"> • Pass / fail • Faculty-developed MCQ exams at the end of each module • No NBME exams • No cumulative exam
Details of new science integration in the clerkships	<ul style="list-style-type: none"> • Two 1-week science intensives • Online basic and clinical integrated modules • One 4-week "selective" that integrates basic science and clinical medicine 	<ul style="list-style-type: none"> • Spaced education • Distance learning • Small-group discussions • Journal clubs 	<ul style="list-style-type: none"> • 1-week intensive science review • 6 interspersed daylong science sessions 	<ul style="list-style-type: none"> • Clerkship-specific basic science and clinical medicine integration
Length of Step 1 study period pre move	6–8 weeks	6 weeks ^a	6 weeks ^a	6–7 weeks
Length of Step 1 study period post move	5–6 weeks ^a	6 weeks ^a	6 weeks ^a	8 weeks ^a

Abbreviations: Step 1 indicates United States Medical Licensing Examination Step 1; MCQ, multiple-choice question; NBME, National Board of Medical Examiners.

^aMore if vacation time is used.

allowing scores to be meaningfully aggregated across years. Prior to conducting statistical analyses examining pre–post differences, we addressed several potential confounding factors to better isolate the impact of changing the timing of Step 1, and to aid interpretation of the results. Specifically, we accounted for rising Step 1 scores nationally, adjusted for different initial years of implementation, and controlled for potential differences in cohort ability. We also sought to understand how the scores of examinees in our sample compared with students' scores at similar schools in the same time periods. Here we describe in detail how each of these issues was addressed.

Rising Step 1 scores nationally. Across the time frame of this study, scores on Step 1 increased nationally (Supplemental Digital Appendix 1, available at <http://links.lww.com/ACADMED/A600>). The equating process helps ensure that this increase reflects an increase in examinee ability across time. However, as noted by Daniel et al.,⁶ this increase in ability presents a confounding factor, as it would potentially inflate comparisons of Step 1 scores if this generic increase were conflated with gains related to changing the timing of Step 1 in the curriculum. To isolate ability increases resulting from the timing change as best as possible, we first computed the deviation between each school's average Step 1 score and the national average for each year. For this adjustment, the national average was computed as the mean performance for first-time Step 1 test takers from U.S. and Canadian medical schools within the given year, excluding the four study schools. This process accounts for the average gain seen across U.S. and Canadian schools when comparing scores across time by setting the average year-to-year growth of these schools to zero. For example, if a cohort from one of the study schools performed four points higher than the previous cohort, but the national average also increased by four points during the same time, the change in deviations would equal zero. Therefore, a positive increase in the deviation score indicates that the school's scores improved at a greater rate than the national average and would provide evidence that the Step 1 timing change could have had a greater impact on scores than the typical year-to-year changes occurring across schools.

Different initial years of implementation. After computing deviation scores for the study schools within each available year, the data were aggregated across the four schools by cohort relative to the exam timing change (three years prior, two years prior, etc.) because schools implemented the change during different years. This allowed the data from different schools to be pooled together in a meaningful manner despite the different implementation years. Step 1 scale scores for the studied years were equated and thus could be compared across time.

Potential differences in cohort ability. In addition to controlling for steadily increasing national scores on Step 1, this study used incoming individual student Medical College Admission Test (MCAT) scores as a covariate to control for potential differences among the initial test-taking ability of each cohort involved in this study. For this, we used analysis of covariance (ANCOVA) with deviation scores as the dependent variable and pre–post change as the independent factor. The ANCOVA helped ensure that potential differences we would attribute to the curricular change were not arising from an increase in the initial ability of students admitted to these schools over time. For instance, if the cohorts after the curricular change had happened to have higher initial ability, as measured by MCAT scores, we may have falsely attributed a difference in Step 1 scores to the curricular change rather than the disparity of cohort ability. In essence, the ANCOVA creates a level baseline for comparison in terms of initial knowledge.

Comparison with similar schools. Another comparison of interest was how scores increased in the study schools compared with similar schools in the same time periods. To address this question, we employed a resampling procedure to create a distribution of change score from similar schools that were not included in the study group. This process involved randomly matching each of the four examinee cohorts used in our study schools to an LCME-accredited school from a subset of schools with similar incoming MCAT scores during the year of the Step 1 timing change and deriving the average gain score estimate that aligned with the one computed in this study (see Supplemental Digital Appendix 2 at <http://links.lww.com>

com/ACADMED/A600 for additional description of the procedure). This process was replicated with replacement 200,000 times to construct a distribution of gain scores for various combinations of the matched schools from the subset of schools with similar MCAT scores. This distribution can be used to make statements regarding the likelihood that we would observe a particular increase from any random assortment of four other schools with similar MCAT scores had they been sampled from the same period as the cohorts in our study. All data analyses in this study were conducted via R 3.4.0 (R Foundation, Vienna, Austria).

Results

Step 1 descriptive statistics are presented in Table 2. In the first year of implementation, Step 1 scores increased by an average of 6.83 scaled score points compared with the year prior to implementation, a similar value to the estimate provided by Daniel et al.⁶ However, as this metric of growth is confounded with the general trend of increasing Step 1 scores nationally, it is more appropriate to interpret growth relative to the difference from the national average. This adjusted metric indicates that Step 1 scores increased by 4.09 scale score points relative to the national average for the first cohort after implementation, and by an average of 2.78 score points when comparing the three years pre change to the three years post change.

Table 2 shows that Step 1 failure rates in the study schools decreased from 2.71% ($n = 15$) to 0.74% ($n = 4$) for the first cohort implementation. The average fail rate decreased from 2.87% ($n = 48$) pre change to 0.39% ($n = 6$) post change. Fisher exact test indicates that the decrease of 2.48% was statistically significant ($P < .001$).⁸ The 95% confidence interval (CI) for the difference between fail rates using Newcombe's⁹ recommended method was –3.43% to –1.65%.

Figure 1 displays the mean difference from the national average by cohort, along with the 95% CI for each mean. The figure shows that the score increase between the cohort prior to the change and the cohort immediately after is considerable, with no overlap among the error bars.¹⁰ The study schools were trending toward the national average in each of the two

Table 2

Descriptive Statistics, Aggregated Across Four Study Schools, Relative to Implementation of a Curricular Change to Administer the United States Medical Licensing Examination Step 1 After Core Clerkships

Cohort relative to implementation ^a	No. of students	Step 1 score, mean (SD)	Mean difference from national average, (SD difference score)	Fail rate
-3	548	230.09 (20.94)	6.38 (20.84)	3.83%
-2	566	229.75 (19.79)	5.08 (20.04)	2.12%
-1	554	230.03 (19.71)	3.61 (20.25)	2.71%
1	538	236.86 (17.44)	7.70 (17.63)	0.74%
2	536	238.36 (17.09)	8.89 (17.09)	0.19%
3	455	236.25 (17.00)	6.64 (16.95)	0.22%

^aThe -3 cohort represents three years prior to exam timing requirement change, -2 represents two years prior, and so forth. Cohorts took the Step 1 exam between 2008 and 2016.

cohorts prior to the change in Step 1 timing. The subsequent two cohorts after implementation of the change showed an increase in scores after accounting for national averages, before a slight regression back toward the national average for the third cohort after implementation.

Results of the ANCOVA, including all six years of data, showed that after controlling for MCAT scores, students

in the cohorts following the timing change (adjusted mean = 7.89, standard error = 0.41) performed significantly better on Step 1 relative to the national average than students in the cohorts prior to the change (adjusted mean = 5.23, standard error = 0.43) by an average of 2.66 points (95% CI: 1.50–3.83; $F_{(1,3116)} = 20.10, P < .001$). Computing a standardized mean difference effect size from the covariate adjusted data¹¹ yielded

an effect size of 0.14. This value indicates that the postchange group scored 0.14 standard deviations above the prechange group after accounting for differences in MCAT scores, and represents a small-to-negligible effect according to conventional guidelines.¹²

Figure 2 shows the density distribution created by 200,000 replications in our resampling procedure. The average score gain was 0.88, with a standard deviation of 1.58. The distribution appears rather normally distributed, with perhaps a slight positive skew toward more extreme increases in Step 1 scores. Across our randomly drawn samples of four schools, 99.92% ($n = 199,845$) fell below the average gain for the study schools. In other words, only about 8 in every 10,000 sampled sets of schools had an average score gain of 6.83 or above in our resampling procedure.

Discussion

The primary reasons for moving Step 1 included facilitating earlier entry into clinical environments, promoting integrated basic science learning in clinical contexts, and enhancing

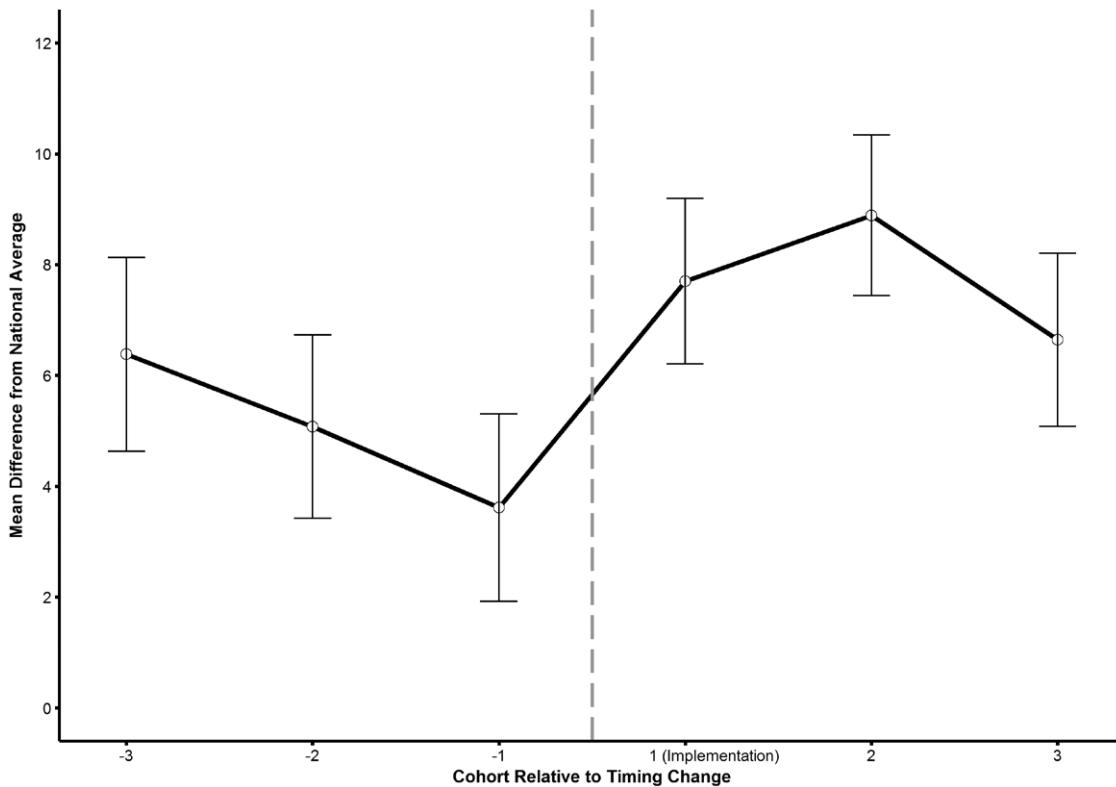


Figure 1 Mean difference from the national average United States Medical Licensing Examination Step 1 score by cohort, along with the 95% confidence interval for each mean. Step 1 scores from four schools that changed the timing of Step 1 from after basic science curricula to after core clerkships were included. For each cohort, scores were examined three years pre change and three years post change. Scores from 2008 to 2016 were used.

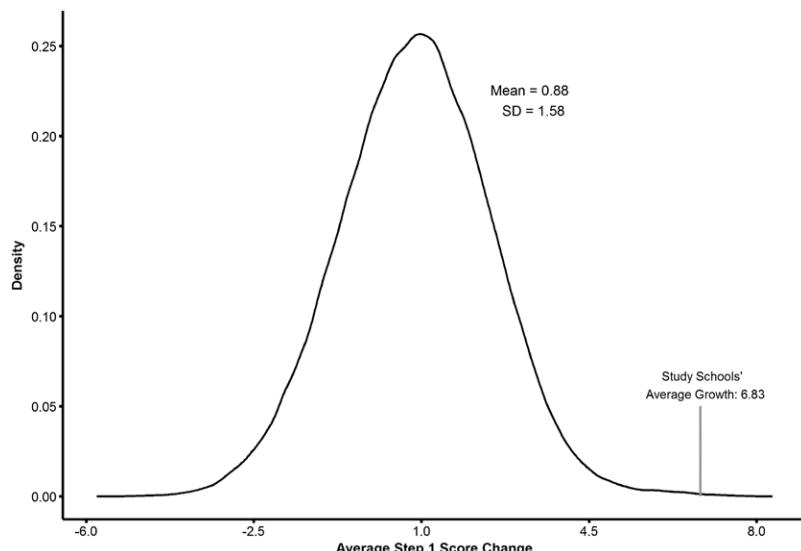


Figure 2 Density distribution of United States Medical Licensing Examination Step 1 score changes from a resampling procedure using 200,000 replications. Cohorts taking Step 1 at four study schools after the timing of Step 1 was changed from after basic science curricula to after core clerkships were randomly matched to a Liaison Committee on Medical Education–accredited school from a subset of similar schools. See Supplemental Digital Appendix 2 at <http://links.lww.com/ACADMED/A600> for additional description of the resampling procedure.

foundational science retention. We were optimistic that, at a minimum, delaying Step 1 would not negatively impact scores, particularly as the USMLE program has moved to incorporate more clinically oriented vignettes on Step 1.¹³ The results of our analyses indicated a small but statistically significant increase in scores for cohorts after the timing change when controlling for rising national averages and differences in cohorts' MCAT scores. The degree of this increase was larger than typically seen among schools with similar MCAT scores in the same time period. We also observed a statistically significant decrease in failure rates after the timing change.

Medical schools want their students to become licensed physicians, so any reduction in Step 1 failure rates would be embraced by learners and institutions. The manner in which a Step 1 failure could be detrimental to a student's residency match success is another reason that reduction in Step 1 failure rates would be desirable. Given the small number of overall failures in our intervention schools, this trend, although promising, is not large enough to draw conclusions.

In terms of score increases, although the aggregate increase of 6.83 was statistically significant, the magnitude of gain was small relative to the Step 1

score scale, and may be negligible at the individual level. For example, the 6.83 value falls slightly above the standard error of measurement of Step 1 scores (5 score points) and within the standard error of the difference (8 score points).¹⁴ Therefore, we caution against interpreting these changes as educationally meaningful, as small changes in scale scores are unlikely to represent important differences in foundational knowledge or ability to deliver clinical care.

Notably, we were able to demonstrate "noninferiority" for the change in Step 1 timing (i.e., the change in Step 1 timing is not worse than the comparator of Step 1 pre clerkships). For institutions undergoing curricular reform, trying to diverge from the traditional 2 + 2 model, this finding could provide guidance for both faculty and administrators. The four institutions in this study, on average, shortened their preclerkship basic science curricula by 21 weeks without observing degradation in Step 1 scores. Given the increased emphasis on early clinical exposure,^{15,16} meaningful learning from patients in clinical contexts,¹⁷ and vertical integration,¹⁸ this noninferiority finding may be considered critically important.

Limitations

Some important limitations of this study must be considered. The change in Step 1 timing may be difficult to

disentangle from other curricular changes implemented during this time frame, including the potential for different emphasis placed on Step 1 preparation at the selected schools. These factors could have potentially influenced learner scores. Notably, these institutions all made curricular revisions that increased emphasis on basic science-related content during the core clerkships, and one school increased the time allotted for Step 1 preparation (Table 1). We would not recommend changing the timing of the exam without making other structural changes to curricula to reinforce basic science learning. Nor would we recommend changing the timing of Step 1 in curricula that maintain the traditional 2 + 2 structure, as this could place the examination too close to the submission of residency applications, and perhaps delay an institution's ability to determine a student's potential to be successfully licensed.

Other factors may have influenced examinee scores. Figure 1 shows a slight decrease relative to the national average in the third cohort after implementation. This may represent typical year-to-year variation, or it may indicate a trend back toward typical performance. As mentioned, the implementation of a curricular change may have consciously or subconsciously increased attention, resources, and effort on Step 1 preparation by faculty and/or students, producing higher scores initially. If the novelty of the change wears off, scores might regress back to typical levels.

Of note, three of the four schools examined in this study tended to score historically above average relative to other schools' Step 1 scores, with lower overall fail rates. They also share other characteristics that may not generalize well to the total population of LCME-accredited medical schools. Thus, we are limited in our ability to make claims regarding the generalizability of a Step 1 timing change, particularly regarding how it would impact schools that tend to have lower Step 1 scores, as also discussed in Daniel et al.⁶

Conclusions and future directions

Among the schools studied, our analyses provide evidence that a change in the timing of Step 1 in the curriculum did not negatively affect performance, supporting the goal of *noninferiority*

put forth by Daniel et al.⁶ This result may encourage other institutions to consider similar curricular innovations that integrate the scientific principles necessary for the practice of medicine with students' relevant clinical training. The schools in this study have demonstrated that preparing students for a required licensure exam is not a hindrance to curricular reform and medical education innovation. The findings also highlight several areas for further study.

Further research is needed to evaluate other criteria put forth as potential benefits of delaying Step 1, including increased retention of basic science concepts later in students' academic careers. Prior studies by the National Board of Medical Examiners (NBME) have shown that retention of basic science information typically declines over time.^{19,20} A physician workforce with knowledge of basic sciences is vital to advance health care in the 21st century, and our current state of frontloading basic science content without spaced integration and reinforcement is ineffective. Future studies may investigate basic science retention by examining performance on basic science items on the USMLE Step 2 Clinical Knowledge (CK) and Step 3 exams.

Future investigations may also wish to evaluate the impact of curricular change and USMLE timing on student wellness and burnout. Students matriculate in medical school with lower rates of burnout and depression and better quality-of-life indicators than the general population, yet they have worse measures in the end.^{21,22} We are interested in exploring the effects of altering the timing of USMLE Step 1 on student stress levels.

There may be other important consequences associated with a change in Step 1 timing. For example, the later timing of Step 1 may impair early identification of learners who struggle on Step 1. Learners who have not consolidated their basic science knowledge may experience challenges with clerkship knowledge acquisition and performance on NBME clinical subject "shelf" exams. Step 2 CK scores may also be affected, as most students will be farther removed from the core clerkships when they take the exam, although the recency of reviewing for Step 1 may balance this effect.

In sum, USMLE Step 1 scores showed small but statistically significant increases, and the failure rates significantly decreased at our study schools. Given the standard error of Step 1 scores, these findings may not be educationally meaningful; however, we are gaining confidence that moving Step 1 after core clerkships is *noninferior* to taking the examination pre clerkships, a liberating finding for educators at peer institutions looking to implement curricular reform.

Acknowledgments: The authors wish to thank Colleen Ward for her instrumental contributions throughout the planning and development of this manuscript.

Funding/Support: The University of Michigan School of Medicine, Vanderbilt School of Medicine, and New York University School of Medicine have Accelerating Change in Medical Education grants from the American Medical Association. Virginia Commonwealth University School of Medicine receives funding from the American Medical Association for S.A. Santen's consulting on the Accelerating Change in Medical Education grant.

Other disclosures: D. Jurich, M. Paniagua, and M.A. Barone work for the National Board of Medical Examiners, the organization that administers the United States Medical Licensing Examination Step 1 examined in this study.

Ethical approval: This study was reviewed and determined to be exempt by the American Institutes for Research (project number EX00398; June 27, 2016).

Disclaimer: The views expressed are those of the authors and do not reflect the official policy or position of their universities, the National Board of Medical Examiners, the Department of Defense, the United States Air Force, or the United States Government.

D. Jurich is senior psychometrician, National Board of Medical Examiners, Philadelphia, Pennsylvania.

M. Daniel is assistant dean for curriculum and associate professor, Departments of Emergency Medicine and Learning Health Sciences, University of Michigan Medical School, Ann Arbor, Michigan; ORCID: <http://orcid.org/0000-0001-8961-7119>.

M. Paniagua is medical advisor, Test Development Services, National Board of Medical Examiners, and adjunct associate professor, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania.

A. Fleming is associate dean for medical student affairs and associate professor of pediatrics, Vanderbilt University School of Medicine, Nashville, Tennessee.

V. Harnik is associate dean for curriculum and associate professor, Department of Cell Biology, New York University School of Medicine, New York, New York.

A. Pock is associate dean for curriculum and associate professor of medicine, Uniformed Services University of the Health Sciences, Bethesda, Maryland.

A. Swan-Sein is director, Center for Education Research and Evaluation, and assistant professor of educational assessment, Columbia University Vagelos College of Physicians and Surgeons, New York, New York.

M.A. Barone is vice president of licensure, National Board of Medical Examiners, Philadelphia, Pennsylvania.

S.A. Santen is senior associate dean of evaluation, assessment and scholarship of learning, and professor of emergency medicine, Virginia Commonwealth University School of Medicine, Richmond, Virginia, and was previously assistant dean for educational research and quality improvement, University of Michigan Medical School, Ann Arbor, Michigan.

References

- Irby DM, Cooke M, O'Brien BC. Calls for reform of medical education by the Carnegie Foundation for the Advancement of Teaching: 1910 and 2010. *Acad Med*. 2010;85:220–227.
- Christensen CM, Bohmer R, Kenagy J. Will disruptive innovations cure health care? *Harv Bus Rev*. 2000;78:102–112, 199.
- Cooke M, Irby DM, Sullivan W, Ludmerer KM. American medical education 100 years after the Flexner report. *N Engl J Med*. 2006;355:1339–1344.
- Association of American Medical Colleges. Time in the curriculum in which medical schools require students to take United States Medical Licensing Examinations (USMLE): USMLE Step 1. <https://www.aamc.org/initiatives/cir/406430/10c.html>. Published 2017. Accessed August 31, 2018.
- Association of American Medical Colleges. Number of medical schools requiring the United States Licensing Examination (USMLE) for advancement / promotion. <https://www.aamc.org/initiatives/cir/406442/10b.html>. Published 2017. Accessed August 31, 2018.
- Daniel M, Fleming A, Grochowski CO, et al. Why not wait? Eight institutions share their experiences moving United States Medical Licensing Examination Step 1 after core clinical clerkships. *Acad Med*. 2017;92:1515–1524.
- Boshuizen H, Schmidt H, Coughlin L. On the application of medical basic science knowledge in clinical reasoning: Implications for structural knowledge differences between experts and novices. In: Program of the Tenth Annual Conference of the Cognitive Science Society: 17–19 August 1988, Montreal, Quebec, Canada. Wheat Ridge, CO: Cognitive Science Society; 1988:517–523.
- Fisher R. On the interpretation of χ^2 from contingency tables, and the calculation of P. *J Royal Stat Soc*. 1922;85(1):87–94.
- Newcombe RG. Two-sided confidence intervals for the single proportion: Comparison of seven methods. *Stat Med*. 1998;17:857–872.
- Cumming G, Finch S. Inference by eye: Confidence intervals and how to read pictures of data. *Am Psychol*. 2005;60:170–180.
- Olejnik S, Algina J. Generalized eta and omega squared statistics: Measures of effect size. *Psychol Methods*. 2003;7:337–347.

- size for some common research designs. *Psychol Methods.* 2003;8:434–447.
- 12** Cohen J. Statistical power analysis. *Curr Dir Psychol Sci.* 1992;1(3):98–101.
- 13** Committee to Evaluate the USMLE Program (CEUP). Comprehensive review of USMLE. <http://www.usmle.org/pdfs/cru/CEUP-Summary-Report-June2008.pdf>. Published 2008. Accessed August 31, 2018.
- 14** Federation of State Medical Boards and National Board of Medical Examiners. USMLE score interpretation guidelines. http://www.usmle.org/pdfs/transcripts/USMLE_Step_Examination_Score_Interpretation_Guidelines.pdf. Published 2017. Updated May 9, 2018. Accessed August 31, 2018.
- 15** Dornan T, Bundy C. What can experience add to early medical education? Consensus survey. *BMJ.* 2004;329:834.
- 16** Dornan T, Littlewood S, Margolis SA, Scherpbier A, Spencer J, Ypinazar V. How can experience in clinical and community settings contribute to early medical education? A BEME systematic review. *Med Teach.* 2006;28:3–18.
- 17** Lisk K, Agur AM, Woods NN. Exploring cognitive integration of basic science and its effect on diagnostic reasoning in novices. *Perspect Med Educ.* 2016;5:147–153.
- 18** Brauer DG, Ferguson KJ. The integrated curriculum in medical education: AMEE guide no. 96. *Med Teach.* 2015;37:312–322.
- 19** Swanson DB, Case SM, Luecht RM, Dillon GF. Retention of basic science information by fourth-year medical students. *Acad Med.* 1996;71(10 suppl):S80–S82.
- 20** Ling Y, Swanson DB, Holtzman K, Bucak SD. Retention of basic science information by senior medical students. *Acad Med.* 2008;83(10 suppl):S82–S85.
- 21** Dyrbye LN, West CP, Satele D, et al. Burnout among U.S. medical students, residents, and early career physicians relative to the general U.S. population. *Acad Med.* 2014;89:443–451.
- 22** Brazeau CM, Shanafelt T, Satele D, Sloan J, Dyrbye LN. Distress among matriculating medical students relative to the general population. *Acad Med.* 2014;89:1520–1525.