9-2020

Which Internal Medicine Clerkship Characteristics Are Associated With Students' Performance on the NBME Medicine Subject Exam? A Multi-Institutional Analysis.

Matthew M. Fitz
Loyola University

William Adams
Loyola University

Steven A. Haist
NBME; University of Kentucky

Karen E. Hauer
University of California, San Francisco

Linette P. Ross
NBME

See next page for additional authors

ACADEMIC MEDICINE
Journal of the Association of American Medical Colleges

Author: Fitz Matthew M. MD, MSc; Adams William PhD; Haist Steven A. MD; Hauer Karen E. MD, PhD; Ross Linette P. MS, PhD; Raff Amanda MD; Agarwal Gauri MD; Vu T. Robert MD; Appelbaum Jonathan MD; Lang Valerie J. MD, MHPE; Miller Chad MD; Grum Cyril MD; and the Clerkship Directors in Internal Medicine - National Board of Medical Examiners EXPRESS Study Group

Title: Which Internal Medicine Clerkship Characteristics Are Associated With Students' Performance on the NBME Medicine Subject Exam? A Multi-Institutional Analysis

DOI: 10.1097/ACM.0000000000003322

Academic Medicine

DOI: 10.1097/ACM.0000000000003322

Which Internal Medicine Clerkship Characteristics Are Associated With Students' Performance on the NBME Medicine Subject Exam? A Multi-Institutional Analysis

Matthew M. Fitz, MD, MSc, William Adams, PhD, Steven A. Haist, MD, Karen E. Hauer, MD, PhD, Linette P. Ross, MS, PhD, Amanda Raff, MD, Gauri Agarwal, MD, T. Robert Vu, MD, Jonathan Appelbaum, MD, Valerie J. Lang, MD, MHPE, Chad Miller, MD, Cyril Grum, MD, and the Clerkship Directors in Internal Medicine - National Board of Medical Examiners EXPRESS Study Group

M.M. Fitz is vice chair for faculty development, Department of Medicine, Loyola University Stritch School of Medicine, Maywood, Illinois.

W. Adams is assistant professor of medical education and public health sciences, Loyola University Chicago, Chicago, Illinois.

S.A. Haist was vice president, Test Development Services, National Board of Medical Examiners, Philadelphia, Pennsylvania, when this work began. He is associate dean, University of Kentucky School of Medicine - Northern Kentucky Campus, Highland Heights, Kentucky, now.

K.E. Hauer is associate dean, Competency Assessment and Professional Standards, University of California, San Francisco, School of Medicine, San Francisco, California.

L.P. Ross is senior psychometrician, National Board of Medical Examiners, Philadelphia, Pennsylvania.

A. Raff is internal medicine clerkship director, Albert Einstein College of Medicine, The Bronx, New York.

G. Agarwal is associate dean for clinical curriculum, University of Miami Miller School of Medicine, Miami, Florida.

T.R. Vu is associate professor of clinical medicine, Indiana University School of Medicine, Indianapolis, Indiana.

J. Appelbaum is chair, Department of Clinical Sciences, and education director and professor of internal medicine, Florida State University College of Medicine, Tallahassee, Florida.

V.J. Lang is internal medicine hospitalist, Director of Meliora in Medicine, and director, Hospital Medicine Faculty Development Program, University of Rochester School of Medicine and Dentistry, Rochester, New York.

C. Miller is associate dean of undergraduate medical education, Saint Louis University School of Medicine, St. Louis, Missouri.

C. Grum is associate chair for undergraduate medical education, University of Michigan Medical School, Ann Arbor, Michigan.

Correspondence should be addressed to Matthew M. Fitz, Loyola University Stritch School of Medicine, Department of Medicine, Building 102, Room 7604, Maywood, IL 60153; telephone: (708) 216 - 4813; email: Mfitz1@lumc.edu; Twitter: @MattFitzMD.

Supplemental digital content for this article is available at http://links.lww.com/ACADMED/A847.

Additional authors: Individuals meeting the criteria for authorship who are not listed individually in the byline but included in the CDIM-NBME Express Study Group include: Mark Fagan, MD, Brown University; Jennifer Foster, MD, Florida Atlantic University; Hillary J. Ryder, MD, Geisel School of Medicine at Dartmouth; Bruce Houghton, MD, Creighton University; Ryan Nall, MD, University of Florida; Amy Shaheen, MD, University of North Carolina; Michael Elnicki, MD, and Anna Donovan, MD, University of Pittsburgh; Stuart Kiken, MD, Rosalind

Franklin College of Medicine; Cynthia Ledford, MD, Ohio State University; Shobhina Chheda, MD, MPH, University of Wisconsin; Doug Paauw, MD, University of Washington; Blake Barker, MD, University of Texas Southwestern Medical School; Maureen Lowery, MD, University of Miami Miller School of Medicine; Nina Mingioni, MD, Thomas Jefferson University Sidney Kimmel Medical College; Deepti Rao, MD, University of New Mexico School of Medicine; and William Kelly, MD, Uniformed Services University.

Acknowledgements: The authors wish to thank the following clerkship directors who contributed logistical support, emotional energy, and data to this article: Susan Hingle, MD, William Pieratt, MD, Viju John, MD, Marty Muntz, MD, Temple Ratcliffe, MD, Saurabh Bansal, MD, Asra Khan, MD, Dave Neely, MD, Dennis Chang, MD, Adam Cifu, MD, Asra Khan, MD, Amber Pincavage, MD, Kirk Bronander, MD, Cynthia Burns, MD, Sean Whelton, MD, Doug Walden, MD, Monalisa Taylor, MD, Alisa Peet, MD, Deborah Deeway, MD, Janet Jokela, MD, John Dorsey, MD, Andrew Hoellein, MD, Michael Battistone, MD, Katie Lappe, MD, John Kugler, MD, Renee Dversal, MD, Nicholas Van Wagoner, MD, John Pittman, MD, Gary Stallings, MD, Jeffery Becker, MD, Oliver Cequeria, MD, Briar Duffy, MD, Patrick McMillan, MD, Janet Fitzpatrick, MD, Katherine Chretien, MD, and Krisi Kosub, MD. They also wish to thank Gail Hendler for her help with numerous literature searches during the study.

Funding/Support: None reported.

Other disclosures: None reported.

Ethical approval: During the recruitment period, the participating clerkship directors obtained institutional review board (IRB) approval or exemption from their respective institutions. The co-investigators from the National Board of Medical Examiners received IRB exemption from the American Institutes for Research.

Disclaimers: The views expressed in this article do not necessarily reflect the views of the Uniformed Services University or the Department of Defense.

Previous presentations: Matthew M. Fitz presented the data in this article during grand rounds at Loyola University Stritch School of Medicine (Chicago, Illinois) in the Spring of 2018 and at the University of Central Florida College of Medicine (Orlando, Florida) in the Spring of 2018.

Data: The National Board of Medical Examiners provided the de-identified data used in this analysis.

Abstract

Purpose

To identify which internal medicine clerkship characteristics may relate to National Board of Medical Examiners (NBME) Medicine Subject Examination scores, given the growing trend toward earlier clerkship start dates.

Method

The authors used linear mixed effects models (univariable and multivariable) to determine associations between medicine exam performance and clerkship characteristics (longitudinal status, clerkship length, academic start month, ambulatory clinical experience, presence of a study day, involvement in a combined clerkship, preclinical curriculum type, medicine exam timing). Additional covariates included number of NBME clinical subject exams used, number of didactic hours, use of a criterion score for passing the medicine exam, whether medicine exam performance was used to designate clerkship honors, and United States Medical Licensing Examination Step 1 performance. The sample included 24,542 examinees from 62 medical schools spanning 3 academic years (2011–2014).

Results

The multivariable analysis found no significant association between clerkship length and medicine exam performance (all pairwise $P > .05$). However, a small number of examinees beginning their academic term in January scored marginally lower than those starting in July ($P < .001$). Conversely, examinees scored higher on the medicine exam later in the academic year (all pairwise $P < .001$). Examinees from schools that used a criterion score for passing the medicine exam also scored higher than those at schools that did not ($P < .05$). Step 1 performance remained positively associated with medicine exam performance even after controlling for all other variables in the model ($P < .001$).

Conclusions

In this sample, the authors found no association between many clerkship variables and medicine exam performance. Instead, Step 1 performance was the most powerful predictor of medicine exam performance. These findings suggest that medicine exam performance reflects the overall medical knowledge students accrue during their education rather than any specific internal medicine clerkship characteristics.

In 2016, 169 (94%) U.S. medical schools accredited by the Liaison Committee on Medical Education (LCME) used the National Board of Medical Examiners (NBME) Medicine Subject Examination as an end-of-clerkship assessment for their internal medicine clerkship.[1] Still, clerkship curricular content and examination content often do not align. At most medical schools, exam topics such as neurology, dermatology, ambulatory medicine, and hospital medicine do not align with the curricular content of the internal medicine clerkship. In addition, the structure of the internal medicine clerkship varies across institutions.[1] Multi-institutional studies examining the effects of clerkship characteristics on medicine subject exam performance are limited. One of the most informative multi-institutional studies examined the association between several internal medicine clerkship characteristics related to structure, pedagogy, and patient contact and medicine subject exam scores.[2] The authors of that study found that more small-group hours per week and the use of community preceptors correlated with higher medicine subject exam scores. However, the study was conducted more than 16 years ago, which may limit its generalizability.

Several clinical specialties have examined the association between clerkship characteristics and subject exam performance.[3-19] For example, research on successive clerkship cohorts from the same specialty showed that students' scores on the NBME subject exams in obstetrics/gynecology,[3,4] surgery,[5,6] and internal medicine[7-10] (but not psychiatry[11,12]) improved toward the end of the academic year. Yet, findings on the effects of clerkship length on exam performance have been mixed. While 2 studies in obstetrics/gynecology found that a greater clerkship length was associated with higher exam scores (especially in the first half of an academic year), the association between psychiatry clerkship length and subject exam performance was more variable.[11-16] In fact, a 2018 study from a single institution showed no association between individual clerkship subject exam scores and clerkship length.[17]

While these mostly single institution studies suggested associations between clerkship characteristics and exam scores, the findings were mixed and the timing of clerkships within the curriculum continues to change. Only a few studies adjusted for United States Medical Licensing Examination (USMLE) Step 1 performance when examining the association between clerkship length and sequence and subject exam performance.[2,6,7,14,18] Yet, at many medical schools, students are starting clerkships earlier in the curriculum.[20] The majority of LCME-accredited schools still have a traditional 2-year preclinical curriculum followed by 2 years of clinical clerkships.[20] Most schools with nontraditional preclinical curricula have students begin their clerkships a few months earlier than the traditional clerkship start date, which is July. Additionally, a few schools have condensed their preclinical years from 24 to 18 or even 12 months. The potential impact of this trend on subject exam scores has not been fully examined. In this study, we examined the following questions: (1) When controlling for USMLE Step 1 scores, what is the association between internal medicine clerkship characteristics and NBME Medicine Subject Examination scores? and (2) What is the association between traditional versus nontraditional clerkship start dates and NBME Medicine Subject Examination scores?

Method

Participants

We recruited internal medicine clerkship directors to participate in our study at the 2014 national Clerkship Directors in Internal Medicine (CDIM) meeting and by phone call over a 10-month period from September 2014 through June 2015. We chose to include data from the most recent academic years at the time of recruitment (2011-2014). Participating clerkship directors obtained institutional review board approval or exemption for our study from their respective institutions. They provided the NBME with their internal medicine clerkship characteristics, and the NBME matched these clerkship characteristics with examinees' medicine subject exam scores.

Subsequently, the NBME provided the first author (M.M.F.) with a completely de-identified dataset for analysis.

Study design

The CDIM-NBME EXPRESS (**Ex**ploration of **Pre**dictors of **S**ubject Examination **S**cores) Study Group, a combination of internal medicine clerkship directors and NBME members, designed this study. (All authors are members of this study group). We analyzed data from 24,542 examinees from 62 LCME-accredited medical schools spanning 3 academic years (2011-2014). We confirmed clerkship characteristics with in-person interviews, phone calls, and follow-up emails to the participating clerkship directors over a 12-month period from 2014 to 2015. Overall, we analyzed medicine subject exam results for students with scores on both the NBME medicine subject exam and USMLE Step 1, which accounted for approximately 46% (R^2 = 0.459) of the variance in medicine subject exam scores.

Since the early 1990s, medicine subject exam scores have been scaled to a mean of 70 with a standard deviation of 8 for the group of first-time examinees from U.S. LCME-accredited medical schools who took the exam as an end-of-clerkship exam that year. During our study period (2011-2014), the mean and standard deviation for first-time examinees were approximately 76 and 8, respectively.[21] The majority of scores were between 45 and 95 on the scaled score. After our data collection, the NBME transitioned to reporting scores on an equated percent score scale. This went into effect in August 2015.[22]

Clerkship characteristics

We defined a *longitudinal student* as a medical student who participated in the care of a cohort of patients over time and had continued learning relationships with these patients' clinicians to achieve clinical competence across multiple specialties in addition to internal medicine. *Academic start month* was the first month of any clinical clerkships at a particular school.

Clerkship length was the duration of the internal medicine clerkship in weeks. Students take the Medicine Subject Examination at the end of this block of time. Having an *ambulatory clinical experience* entailed participating in outpatient clinical care during the internal medicine clerkship; we further refined this variable to be either a structured block format separate from the inpatient experience (ambulatory clinical experience = yes) or integrated into the inpatient experience (ambulatory clinical experience = mixed). A *study day* was the presence of one or more days after clinical responsibilities ended but before the subject exam. A *combined clerkship* included at least one other specialty (e.g., emergency medicine or neurology) in addition to internal medicine. A *pass-cutoff* designated a school's use of any criterion score for passing the medicine subject exam during the internal medicine clerkship. An *honors-cutoff* designated a school's use of any criterion score on the medicine subject exam for receiving an honors grade for the clerkship.

A preclinical curriculum was *traditional* (i.e., discipline-specific basic science subjects or courses taken sequentially), *organ-based* (i.e., centered around body systems such as pulmonary or cardiology with integrated anatomical, physiological, and pathological processes), or *hybrid* (i.e., a mix of the 2 preceding models); a curriculum not clearly described was *other*. *Quarter* indicated the timing of the medicine subject exam during the academic year (i.e., administered in the first, second, third, or fourth quarter of the year). The number of *didactic hours* was the number of hours within the internal medicine clerkship dedicated to the delivery of the formal curriculum, including lectures and case discussions. Finally, the *number of NBME clinical subject exams* was the total number of NBME clinical science subject exams used in the school's clinical years.

Statistical analysis

We listed the number of examinees for each nominal and ordinal clerkship characteristic as valid counts and proportions. This included longitudinal status, clerkship length, academic start month, ambulatory clinical experience, presence of a study day, involvement in a combined clerkship, use of a criterion score for passing the medicine subject exam, whether medicine subject exam performance was used to designate honors in the clerkship, type of preclinical curriculum, and the quarter of the year in which students took the medicine subject exam. We described continuous covariates using median with interquartile range (IQR) values for the number of NBME clinical subject exams used and mean with standard deviation (SD) values for the number of didactic hours and Step 1 score.

We used both univariable and multivariable linear mixed effects models to estimate the average medicine subject exam score as a function of longitudinal status, clerkship length, academic start month, ambulatory clinical experience, presence of a study day, involvement in a combined clerkship, type of preclinical curriculum, the quarter of the year in which students took their medicine subject exam, number of NBME clinical subject exams used, number of didactic hours, and Step 1 performance. Additional covariates included whether a school used a criterion score for passing the medicine subject exam and whether medicine subject exam performance was used to designate honors in the clerkship. Because most of the examinees (> 80%) were in either an 8-week or 12-week clerkship, we treated clerkship length as a nominal (rather than quantitative) variable in the model.

We hypothesized that the included covariates would have a meaningful association with medicine subject exam performance. In our models, we regressed the nominal covariates against a referent and, when necessary, adjusted their confidence intervals (CIs) and significance values for the multiple pairwise comparisons using a Sidak correction to control the Type 1 error rate.

Further, to account for the clustering of examinees within schools, we allowed random intercepts for each medical school contributing to the estimates using a completely general (unstructured) covariance matrix. For all comparisons of the fixed effects, we applied a Kenward-Roger correction to estimate the denominator degrees of freedom.[23]

Regarding model assumptions, we used Akaike's information criterion as a measure of improvement in model fit from univariable to multivariable conclusions, and we assessed the linearity and normality assumptions using residual plots and QQ plots, respectively. Multicollinearity among the covariates was assessed using variance inflation factors and tolerance statistics. Because the adjusted (multivariable) model was a 2-level hierarchical linear model, we estimated the model's effect size or coefficient of determination (R^2) at both the examinee- and school-level as described by Recchia.[24]

Finally, through sensitivity analyses, we assessed whether the academic start month moderated the association between clerkship length and medicine subject exam performance. Because this interaction term was not statistically significant in both our univariable and multivariable analyses, we removed it from the model. Given interest in beginning the academic term earlier in the calendar year,[20] we report stratified summary medicine subject exam performance statistics for each clerkship characteristic by examinees' academic start month in the supplemental digital content (see below).

We used SAS version 9.4 (Cary, NC) for all analyses.

Results

Among the 24,542 examinees included in this study, the majority were in 8-week (10,531; 42.9%) or 12-week (9,544; 38.9%) clerkships. About 5.7% (1,405) were enrolled in a 6-week clerkship, which was the shortest clerkship length in the study. Some examinees were enrolled in 9-week (431; 1.8%), 10-week (1,443; 5.9%), or 11-week (572; 2.3%) clerkships. Only 433

(1.8%) were enrolled in a longitudinal clerkship. Most examinees began their clerkships in July (17,044; 69.4%) with the remainder starting in May (4,682; 19.1%), June (2,644; 10.8%), or January (172; 0.7%).

Roughly half of examinees were from schools with no ambulatory clinical experience (11,969; 48.8%) while the remainder were from schools that used an ambulatory block format (11,583; 47.2%) that was separate from the inpatient experience; 4.0% (990) had an integrated inpatient and ambulatory format. Approximately 12.5% (3,073) of examinees were in a combined clerkship (e.g., a clerkship that combined emergency medicine or neurology with internal medicine), and the majority (13,433; 54.7%) received a study day. Nearly all (21,822; 89.1%) were enrolled in a clerkship with a criterion score for passing the medicine subject exam. For most examinees (20,298; 82.7%), performance on the medicine subject exam was used to designate honors in the clerkship. Fewer than half of examinees (10,073; 41.0%) had a traditional preclinical curriculum, while another 14.2% (3,485) had an organ-based curriculum; approximately 17.3% (4,248) had a hybrid curriculum. The median number of NBME clinical subject exams used at the included schools was 6.00 (IQR: 5 - 7), and examinees received an average of 30.75 (SD = 16.30) didactic hours of education during the internal medicine clerkship. The average Step 1 score was 227.56 (SD = 20.91). See Table 1 for the complete clerkship characteristics.

Univariable analysis showed an association between clerkship length and medicine subject exam performance (overall $P = .001$). Post-hoc pairwise comparisons that adjusted for inflated Type 1 error revealed that examinees in 12- to 20-week clerkships scored 3.17 (95% CI: 0.43 to 5.92) points higher than those in 6-week clerkships ($P = .02$). However, after controlling for all other clerkship characteristic variables and Step 1 performance in the model, we found no statistically

significant association between clerkship length and performance on the medicine subject exam (all adjusted pairwise $P > .05$; see Table 2).

Controlling for all other covariates, there was a significant association between academic start month and medicine subject exam performance (overall $P = .001$). Corrected post-hoc tests revealed that examinees starting their academic term in January scored lower than those beginning in July ($\mu_{diff} = -3.71$, 95% CI: -5.83 to -1.58; $P < .001$). Conversely, examinees at schools that used a criterion score for passing the medicine subject exam scored 1.36 (95% CI: 0.08 to 2.63) points higher than those at schools that did not ($P = .04$). Controlling for all other covariates, later quarter for the medicine subject exam was associated with higher performance (overall $P < .001$), as was a higher Step 1 score ($P < .001$).

Clerkship length and medicine subject exam performance did not depend on the month in which examinees began their clerkship; the same findings emerged from our univariable (overall interaction $P = .32$) and multivariable (overall interaction $P = .47$) analyses. Stratified summary statistics for each clerkship characteristic by academic start month (i.e., May, June, and July) are available as Supplemental Digital Appendices 1 - 3 available at http://links.lww.com/ACADMED/A847. Generally, for each academic start month, medicine subject exam performance was higher for longer clerkships. This was particularly true in May, where students in the longest clerkships (12-20 weeks) achieved an average score of 81.03 (SD = 7.99) points.

Discussion

Our study did not reveal differences across the broad array of clerkship characteristics that we hypothesized may be related to NBME Medicine Subject Examination scores. In our large sample, performance on the medicine subject exam was comparable for students in short (6 weeks), medium (8-11 weeks), and long (12-20 weeks) clerkships, after controlling for other

clerkship characteristics including USMLE Step 1 scores. This finding seems reassuring considering the national trend toward a reduction in internal medicine clerkship length. Our findings also corroborate recent results showing no statistical difference in clinical subject exam scores across disciplines at one institution despite a reduction in clerkship length by as much as 25%.[17] Past studies finding higher medicine subject exam scores with increasing clerkship length captured small differences and perhaps are less relevant to current students and curricula.[2,13-16]

One clerkship characteristic of interest in our study was an earlier start date for clerkships. Students who started their clinical year after only 18 months of preclinical study (i.e., January) scored lower on the medicine subject exam than students with a traditional clinical start date (i.e., July). This result may be spurious, however, because the January cohort comprised examinees (0.7%) who were in the first year of a new clerkship structure at one institution. Several schools have subsequently transitioned to a preclinical curriculum that spans only 18 or even 12 months.[25] Our study did not capture data from these curricular changes. It is possible that the effects of clerkship length are masked in our study because of the very large sample of examinees from schools with the traditional 2-year preclerkship period. If more schools transition to earlier clerkship start dates, it will be important to monitor and study the effects of these changes.

Our findings are consistent with those from previous studies regarding the predictive value of USMLE Step 1 scores for performance on the medicine subject exam.[2-7] As in past studies,[2,7-10] we found that students' medicine subject exam scores improved throughout the academic year, suggesting an incremental accrual of general medical knowledge as the year advanced, regardless of the timing of individual clerkships. While we did not capture the timing and order of all clerkships, we infer that our study involving more than 20,000 students across more than 60 schools included a wide distribution of clerkship order and arrangement. Recognizing that

many internal medicine clerkships use the medicine subject exam score in determining clerkship grades and for national comparative data, clerkship directors should incorporate this information when comparing students throughout the academic year.

We found a small but statistically significant difference in medicine subject exam scores among students at schools that used a criterion score for passing the medicine subject exam and those at schools that did not. Students at schools with this pass cutoff had higher subject exam scores. However, the confidence interval barely excludes no difference at all and, in the context of more than 24,000 examinees, does not appear to be a meaningful difference in practice. Nevertheless, it is somewhat surprising to us that a criterion pass cutoff produced this statistical difference. This finding may be relevant to the few students who were near the cutoff between passing and failing the exam at their schools, which can affect grading and remediation decisions.

Despite the seeming benefits of increased integration of clinical frameworks into the preclinical curriculum,[26-29] students at schools with an organ-based or hybrid preclinical curriculum did not perform better on the medicine subject exam compared to their peers at schools with a traditional preclinical curriculum format. We acknowledge that classifying the preclinical curriculum at every school was challenging, and more than a quarter of the schools we studied did not fit into our category scheme. Despite including novel formats for the preclinical curriculum,[26] our findings did not identify a better method for the acquisition and retention of the knowledge typically included in this stage of medical school, as measured by the medicine subject exam. It is possible that the lack of difference may reflect a mismatch between curricula and assessment; the examination also may not capture the incremental benefits of an integrated preclinical curricula or that differences in curricula may have relatively little impact on medicine subject exam performance. In addition, in contrast to an earlier study,[2] we did not find an association between the number of didactic hours during clerkship and medicine subject exam performance.

Further characterization of the quality or nature of the didactic curriculum might yield different results.[2]

It is encouraging that longitudinal students' scores on the medicine subject exam were not significantly different from the scores of students in traditional block clerkships. Past studies investigating outcomes for longitudinal students also showed either no difference or improved assessment scores on standardized exams for these students.[30-35]

Despite the medicine subject exam outline identifying the ambulatory (office or clinic) setting as the site of care for 55% - 65% of the exam,[36] we did not find higher scores at schools with an ambulatory experience or curricula in the internal medicine clerkship. One possible explanation for this finding is that the exam content represents clinical problems that present in both outpatient and inpatient settings. Students without ambulatory experience as part of their internal medicine clerkship may have benefitted from prior ambulatory experience in other rotations.

Our study has multiple limitations. All the included schools were U.S. LCME-accredited medical schools, so our findings may not be applicable to DO-granting or international medical schools. Additionally, our data were from the 2011-2014 academic years. While the medicine subject exam framework and content have been consistent, internal medicine clerkships may have changed, and other clerkship characteristics may affect the interactions among the curricula, training environment, and students' medicine subject exam performance. A few schools have now adopted even earlier starts to their clerkships with NBME clinical subject exams predating USMLE Step 1. We were not able to include these changes in our study. However, our large study had a similar representative distribution of internal medicine clerkship variables across the included schools as did other studies of survey data from the Association of American Medical Colleges, CDIM, and NBME during our study period.

In conclusion, we found that many of the clerkship variables we hypothesized would be associated with NBME Medicine Subject Examination performance had no statistically significant association. Some internal medicine clerkships were twice as long as others, and some had significant ambulatory curricula or clinical experiences. Yet we did not find any difference in medicine subject exam scores based on these clerkship characteristics after controlling for USMLE Step 1 scores. Medicine subject exam performance reflects students' overall medical knowledge and may reflect the knowledge they obtained during the internal medicine clerkship, in the preclinical years, from previous clerkships, or from independent study, rather than the unique characteristics of the internal medicine clerkship. Thus, clerkship directors and medical schools that use the medicine subject exam should consider that the exam measures overall medical knowledge and may not reflect or measure the characteristics or experiences that affect students' learning during the internal medicine clerkship itself.

References

1. National Board of Medical Examiners. 2016 NBME Clinical Clerkship Subject Examination Survey: Summary of Results. 2016. https://www.nbme.org/PDF/SubjectExams/Clerkship_Survey_Summary.pdf. Accessed February 18, 2020.
2. Griffith CH, Wilson J, Haist SA, et al. Internal medicine clerkship characteristics associated with enhanced student examination performance. Acad Med. 2009;84:895-890.
3. Clark KH, Jelovsek FR. Effect of clerkship timing on third-year medical students' grades and NBME scores in an obstetrics-gynecology clerkship. Acad Med. 1992;67:865.
4. Hampton HL, Collins BJ, Perry KG, Meydrech EF, Wiser WL, Morrison JC. Order of rotation in third-year clerkships: Influence on academic performance. J Reprod Med. 1996;41;337-340.
5. Baciewicz FA, Arent L, Weaver M, Yeastings R, Thomford NR. Influence of clerkship structure and timing on individual student performance. Amer J Surg. 1990;159:265-268.
6. Ripkey DR, Case SM, Swanson DB. Predicting performances on the NBME Surgery Subject Test and USMLE Step 2: The effects of surgery clerkship timing and length. Acad Med. 1997;72:S31-S33.
7. Ouyang W, Cuddy MM, Swanson DB. US medical student performance on the NBME subject examination in internal medicine: Do clerkship sequence and clerkship length matter? J Gen Int Med. 2015;30:1307-1312.
8. Whalen JP, Moses VK. The effects on grades of the timing and site of third-year internal medicine clerkships. Acad Med. 1990;65:708-709.

9. Kies SM, Roth V, Rowland M. Association of third-year medical students' first clerkship with overall clerkship performance and examination scores. J Am Med Assoc. 2010;304:120-126.
10. Grum C, Woolliscroft JO, Case SM, Swanson DB, Ripkey DR. Impact of block assignments on development of diagnostic skills in a medicine clerkship. In: Rothman AI, Cohen R, eds. Proceedings of the Sixth Ottawa Conference on Medical Education. Toronto, Ontario: University of Toronto Bookstore Custom Publishing; 1995.
11. Case SM, Ripley DR, Swanson DB. The effects of psychiatry clerkship timing and length on measures of performance. Acad Med. 1997;72:S34-S36.
12. Manley M, Heiss G. Timing bias in the psychiatry subject examination of the National Board of Medical Examiners. Acad Psych. 2006;30:116-119.
13. Myles TD. Effect of a shorter clerkship on third-year obstetrics and gynecology final examination scores. J Reprod Med. 2004;49:99-104.
14. Edwards RK, Davis JD, Kellner KR. Effect of obstetrics-gynecology clerkship duration on medical student examination performance. Obstet Gynecol. 2000;95:160-162.
15. Bostwick JM, Alexander C. Shorter psychiatry clerkship length is associated with lower NBME psychiatry shelf exam performance scores. Acad Psych. 2012;36:174-176.
16. Niedermier J, Way D, Kasick D, Kuperschmidt R. Effect of curriculum change on exam performance in a 4-week psychiatry clerkship. Acad Psych. 2010;34:216-219.
17. Monrad SU, Zaidi NLB, Gruppen LD, et al. Does reducing clerkship lengths by 25% affect medical student performance and perceptions? Acad Med. 2018;93:1833-1840.
18. Ryan MS, Bishop S, Browning J. Are scores from NBME subject examinations valid measures of knowledge acquired during clinical clerkships? Acad Med. 2017;92:847-852.

19. Colbert C, McNeal T, Lezama M, et al. Factors associated with performance in an internal medicine clerkship. Proc (Bayl Univ Med Cent). 2017;30:38-40.
20. Association of American Medical Colleges. Academic Level Length Distribution in US and Canadian Medical Schools: 2016-2017. Curriculum Inventory. https://www.aamc.org/initiatives/cir/451832/ci05.html. Accessed February 18, 2020.
21. National Board of Medical Examiners. Subject Examination Program. 2013. https://www.nbme.org/pdf/samplescorereports/clinical_sci_score_report.pdf. Accessed February 18, 2020.
22. National Board of Medical Examiners. Subject Examination Program. 2015. https://www.nbme.org/pdf/SampleScoreReports/Clinical/Clinical%20Score%20Report.pdf. Accessed February 18, 2020.
23. Kenward MG, Roger JH. An improved approximation to the precision of fixed effects from restricted maximum likelihood. Computational Statistics and Data Analysis. 2009:53:2583-2595.
24. Recchia A. R-squared measures for two-level hierarchical linear models using SAS. J Stat Software. 2010;32:1-9.
25. Jurich D, Daniel M, Paniagua M, et al. Moving the United States Medical Licensing Examination Step 1 after core clerkships: An outcomes analysis. Acad Med. 2019;94:371-377.
26. Haist SA, Butler AP, Paniagua MA. Testing and evaluation: The present and future of the assessment of medical professionals. Adv Physiol Educ. 2017;41:149-153.
27. Brandi K, Schneid SD, Smith S, et al. Small group activities within academic communities improve the connectedness of students and faculty. Med Teach. 2017;39:813-819.

28. Pfeifer CM. A progressive three-phase innovation to medical education in the United States. Med Educ Online. 2018;23:1427988.
29. Association of American Medical Colleges. Curriculum Change in US Medical Schools: Types of Change, 2016-2017. AAMC Curriculum Inventory. https://www.aamc.org/initiatives/cir/427196/27.html. Accessed February 18, 2020.
30. Poncelet AN, Mazotti LA, Blumberg B, Wamsley MA, Grennan T, Shore WB. Creating a longitudinal integrated clerkship with mutual benefits for an academic medical center and a community health system. Perm J. 2014;18:50-56.
31. Schauer RW, Schieve D. Performance of medical students in a non-traditional rural clinical program, 1998-99 through 2003-04. Acad Med. 2006;81:603-607.
32. Ogur B, Hirsh D, Krupat E, Bor D. The Harvard Medical School-Cambridge Integrated Clerkship: An innovative model of clinical education. Acad Med. 2007;82:397-404.
33. Zink T, Power D, Finstad D, Brooks K. Is there equivalency between students in a longitudinal, rural clerkship and a traditional urban-based program? Fam Med. 2010;42:702-706.
34. Hirsh D, Gaufberg E, Ogur B, et al. Educational outcomes of the Harvard Medical School-Cambridge Integrated Clerkship: A way forward for medical education. Acad Med. 2012;87:643-650.
35. Latessa R, Beaty N, Royal K, Colvin G, Pathman DE, Heck J. Academic outcomes of a community-based longitudinal integrated clerkships program. Med Teach. 2015; 37:862-867.
36. National Board of Medical Examiners. Subject Examinations: Content Outlines and Sample Items. 2019. https://www.nbme.org/PDF/SubjectExams/SE_ContentOutlineandSampleItems.pdf. Accessed February 18, 2020.

Table 1

Characteristics of Internal Medicine Clerkships at 62 Medical Schools Included in a Study of the Association Between Clerkship Characteristics and NBME Medicine Subject Examination Performance, 2011-2014[a]

Clerkship characteristic	No. of examinees	% of examinees	Medicine subject exam score	
			Mean	SD
Longitudinal student				
No	23,673	98.2	78.39	7.96
Yes	433	1.8	77.70	7.53
Total	24,106	100.0	78.38	7.95
Clerkship length				
6 weeks	1,405	5.7	76.18	7.61
8-11 weeks	12,977	52.9	77.94	7.79
12-20 weeks	10,160	41.4	79.27	8.11
Total	24,542	100.0	78.39	7.96
Academic start month				
January	172	0.7	74.55	6.91
July	17,044	69.4	78.53	7.94
June	2,644	10.8	77.66	7.98
May	4,682	19.1	78.43	8.00
Total	24,542	100.0	78.39	7.96
Ambulatory clinical experience				
No	11,969	48.8	77.85	7.73
Yes	11,583	47.2	79.07	8.15
Mixed	990	4.0	76.96	7.73
Total	24,542	100.0	78.39	7.96
Study day				
No	11,109	45.3	78.56	7.91
Yes	13,433	54.7	78.24	8.00
Total	24,542	100.0	78.39	7.96
Combined clerkship				
No	21,469	87.5	78.23	7.93
Yes	3,073	12.5	79.46	8.09
Total	24,542	100.0	78.39	7.96

Pass cutoff				
No	2,669	10.9	77.35	7.70
Yes	21,822	89.1	78.52	7.98
Total	24,491	100.0	78.39	7.96
Honors cutoff				
No	4,244	17.3	77.99	7.91
Yes	20,298	82.7	78.47	7.97
Total	24,542	100.0	78.39	7.96
Preclinical curriculum				
Hybrid	4,248	17.3	78.82	8.18
Organ-based	3,485	14.2	78.94	8.04
Traditional	10,073	41.0	78.57	7.93
Other	6,736	27.4	77.56	7.76
Total	24,542	100.0	78.39	7.96
Quarter				
First	6,911	28.2	77.35	7.86
Second	5,229	21.3	77.93	8.12
Third	6,685	27.2	78.82	7.83
Fourth	5,717	23.3	79.56	7.89
Total	24,542	100.0	78.39	7.96

Abbreviations: NBME, National Board of Medical Examiners; SD, standard deviation; IQR, interquartile range.

[a]Included in this study were 24,542 examinees from 62 medical schools (2011 - 2014). The medicine subject exam score is scaled (μ = 70, SD = 8). The median number of NBME clinical subject exams used was 6.00 (IQR = 5 - 7). The mean number of didactic hours was 30.75 (SD = 16.30). The mean Step 1 score was 227.56 (SD = 20.91).

Table 2

Association Between NBME Medicine Subject Examination Performance and Internal Medicine Clerkship Characteristics at 62 Medical Schools, 2011-2014[a]

Clerkship characteristic	No. of examinees[b]	Unadjusted β (95% CI)	*P*	Adjusted[c] β (95% CI)	*P*
Longitudinal student	24,106	0.51 (-0.69 to 1.71)	.40	0.59 (-0.29 to 1.47)	.19
Clerkship length (vs. 6 weeks)	24,542		.001[d]		.04[d]
8-11 weeks		1.84 (-0.88 to 4.56)	.24[e]	0.65 (-1.36 to 2.66)	.70[e]
12-20 weeks		3.17 (0.43 to 5.92)	.02[e]	1.67 (-0.49 to 3.82)	.15[e]
Academic start month (vs. July)	24,542		.13[d]		.001[d]
January		-0.60 (-2.32 to 1.12)	.79[e]	-3.71 (-5.83 to -1.58)	< .001[e]
May		0.20 (-1.59 to 1.99)	.99[e]	0.27 (-1.06 to 1.60)	.94[e]
June		-1.71 (-3.66 to 0.25)	.10[e]	-0.57 (-2.64 to 1.51)	.87[e]
Ambulatory clinical experience (vs. no)	24,542		.004[d]		.08[d]
Yes		1.18 (-0.08 to 2.44)	.07[e]	0.55 (-0.60 to 1.71)	.48[e]
Mixed		-1.87 (-4.03 to 0.29)	.10[e]	-1.83 (-4.27 to 0.61)	.17[e]
Study day	24,542	0.07 (-0.65 to 0.80)	.84	-0.16 (-0.78 to 0.45)	.60
Combined clerkship	24,542	0.67 (-0.43 to 1.76)	.23	0.45 (-0.84 to 1.74)	.49
Pass cutoff	24,491	1.21 (-0.67 to 3.09)	.20	1.36 (0.08 to 2.63)	.04
Honors cutoff	24,542	0.52 (-0.99 to 2.04)	.49	0.54 (-0.57 to 1.64)	.33
Preclinical curriculum (vs. traditional)	24,542		.45[d]		.13[d]
Hybrid		-0.35 (-1.48 to 0.78)	.84[e]	0.04 (-0.91 to 0.98)	.99[e]
Organ-based		-0.69 (-1.81 to 0.44)	.37[e]	-0.85 (-1.87 to 0.17)	.13[e]
Other		-0.47 (-1.58 to 0.64)	.67[e]	-0.26 (-1.13 to 0.62)	.86[e]
Quarter (vs. first)	24,542		< .001[d]		< .001[d]
Second		0.49 (0.16 to 0.83)	.002[e]	0.65 (0.39 to 0.91)	< .001[e]
Third		1.53 (1.22 to 1.85)	< .001[e]	1.71 (1.46 to 1.95	< .001[e]
Fourth		2.25 (1.91 to 2.58)	< .001[e]	3.04 (2.78 to 3.30)	< .001[e]
No. NBME clinical subject exams used (per 1-exam increase)	24,542	0.02 (-0.36 to 0.40)	.92	-0.03 (-0.36 to 0.29)	.84
No. didactic hours (per 10-hour increase)	24,353	0.24 (0.06 to 0.42)	.01	0.03 (-0.13 to 0.18)	.73
Step 1 score (per 10-point increase)	22,302	2.55 (2.52 to 2.59)	< .001	2.58 (2.54 to 2.61)	< .001

Abbreviations: NBME, National Board of Medical Examiners; CI, confidence interval.

[a]The NBME score is a scaled score ($\mu = 70$, $SD = 8$).

[b]No. of examinees is the number of examinees from 2011 to 2014 whose data the authors used to compute the univariable (unadjusted) estimates. The number of examinees used to compute the multivariable (adjusted) estimates was 21,680.

[c]For the adjusted model, $R^2 = 0.479$ (examinee-level) and $R^2 = 0.639$ (school-level).

[d]The authors used a Type 3 test for the fixed effects.

[e]For clerkship length, academic start month, ambulatory clinical experience, and preclinical curriculum, the authors adjusted the CIs and significance values for the multiple pairwise comparisons using a Sidak correction to control the Type 1 error rate.