Surgery Open Science 1 (2019) 86–89

Contents lists available at ScienceDirect

Surgery Open Science

journal homepage: https://www.journals.elsevier.com/surgery-open-science

Predicting success: A comparative analysis of student performance on the surgical clerkship and the NBME surgery subject exam

Jamil Jaber *, Natasha Keric, Paul Kang, Ara J Feinstein

University of Arizona College of Medicine Phoenix

ARTICLE INFO

Article history:
Received 17 January 2019
Received in revised form 10 July 2019
Accepted 15 July 2019
Available online 17 August 2019

ABSTRACT

Background: The National Board of Medical Examiners surgery shelf is a well-established terminal measure of student medical knowledge. No study has explored the correlation between intraclerkship quizzes and shelf exam performance.

Methods: Weekly quiz and National Board of Medical Examiners scores were collected from 156 third-year students who participated in a 12-week surgical clerkship from 2015 to 2017. Kruskal-Wallis, Wilcoxon rank sum, and linear regression analysis was completed.

Results: Trauma/Burns, Esophagus/Anorectal, and Wound/Intensive Care Unit quiz content corresponded with increased National Board of Medical Examiners performance with β-coefficients of 1.57 ($P < .001$), 1.42 ($P < .001$), 1.38 ($P < .001$), respectively. Wound/Intensive Care Unit and Cardio/Vascular content corresponded with decreased likelihood of scoring <70 points on the National Board of Medical Examiners (OR: 0.75 ($P = .03$), and 0.68 ($P = .02$)). Aggregate quiz scores stratified by academic block were 67 (IQR 64–69.5), 77 (IQR 74.5–80), 76.5 (IQR of 67–89.5), 83 (IQR of 76–85) corresponding to academic blocks 1, 2, 3, and 4, respectively ($P < .001$).

Conclusion: Modeling National Board of Medical Examiners outcomes as a function of weekly quizzes taken during a 12-week surgery clerkship is a viable concept.

INTRODUCTION

The third-year surgical clerkship curriculum represents a significant change in the learning environment as compared to the first and second years of traditional medical education. During the surgical clinical clerkship, the educational benchmarks, objectives, and evaluations are all superimposed on the complexities of clinical care. Furthermore, the subtleties of patient communication, interdisciplinary patient care, technical skill development, and professional acumen present a challenging learning environment for third-year medical students. In this setting, evaluating student performance is integral to both curricular viability and individual student success. Optimal evaluation of medical student performance remains a controversial topic among program directors and medical educators [1]. Many institutions rely heavily on subjective evaluations of student performance by surgery residents and attending faculty. However, these heavily weighted subjective measures of evaluation have been shown to be conflated and unreliable predictors of student performance on the National Board of Medical Examiners (NBME) surgery subject examinations (Surgery shelf); a well-established objective assessment of student medical knowledge [2–4].

The discordance between subjective and objective evaluation methodologies presents a particularly difficult problem for medical educators. One such problem is that student who are struggling with assimilating medical knowledge in the surgery clerkship setting are not readily identified until the end of the clerkship rotation when the NBME surgery shelf is taken. Since the surgery shelf score represents the endpoint of most surgical clerkships, students and educators are left with little actionable insights into improving student performance. Early identification of students struggling in the surgical clerkship curriculum may provide for an opportunity to connect students with resources for success. In addition, students from the University of Arizona College of Medicine Phoenix (UACOMP) select between one of four clerkship tracts. Ultimately, this allows for some choice of when the surgery clerkship will fall within the academic year. The timing of the surgery clerkship, in relation to other fundamental clerkships, is another potential factor impacting student performance on the NBME surgery shelf.

During the 12-week surgery clerkship the students are evaluated with both subjective and objective methodologies. Objective measures of student performance include weekly quizzes covering assigned

* Corresponding author at: University of Arizona, College of Medicine Phoenix, 550 E Van Buren St, Phoenix, AZ 85004, USA.
E-mail addresses: JamilJaber@email.arizona.edu (J. Jaber), Natasha.Keric@bannerhealth.com (N. Keric), paulk@email.arizona.edu (P. Kang), ara.feinstein@bannerhealth.com (A.J. Feinstein).

https://doi.org/10.1016/j.sopen.2019.07.002

Table 1a
Quiz grading scale

Point value	Quiz performance
5	85–100%
4	70–84%
3	55–69%
2	40–54%
1	20–39%
0	<19%

Quiz percentiles and corresponding point values.

reading topics, and the NBME Surgery shelf exam score. This retrospective study was designed to examine the relationship between student performance on weekly quizzes taken during the 12-week surgical clerkship at Banner University Medical Center in Phoenix (BUMCP), Arizona and student performance on the NBME surgery shelf exam. We hypothesized that weekly quiz scores would correlate with NBME surgery shelf exam performance.

MATERIALS AND METHODS

Retrospective quiz scores and NBME surgical subject exam scores were collected for third-year students from the UACOMP. Data corresponded with students who completed a 12-week surgical clerkship at BUMCP during academic years 2015–2016 and 2016–2017 (n = 156). For each week of the surgical clerkship students were given specific reading assignments and weekly quizzes were administered to assess student understanding. Quiz questions were written by content experts and approved by the clerkship director. Some questions with poor reliability were abridged during the clerkships. However, questions were pooled and used throughout the academic years. They consisted of multiple choice basic recall and abridged clinical prompts. Quiz scores were assigned a point value from 0 to 5 according to quiz percentiles as shown in Table 1a. Two quizzes were administered per week allowing for a maximum weekly score of 10 points. Weekly quiz topics and scores stratified by academic year are outlined in Table 1b. Quiz scores were also examined based on when students participated in the surgical clerkship during the academic year. Academic blocks 1, 2, 3, and 4 corresponded to April–June, July–September, October–December, January–March, respectively.

Statistical Analysis. Quiz Scores and NBME performance were assessed using medians and interquartile ranges. The Wilcoxon Rank Sum and Kruskal-Wallis Test were implemented to ascertain differences in scores between academic years and academic blocks, respectively. Univariate linear regression was used to ascertain independent associations between each weekly quiz and NBME scores. Weekly quizzes with $P < .20$ were entered into a second model where a backwards variable selection was conducted to ascertain which quizzes were predictive of NMBE scores. Finally, univariate logistic regression was used to ascertain independent associations between each weekly quiz and the likelihood of scoring below 70 points on NBME performance. Once again, a backwards variable selection was implemented to ascertain predictors of the likelihood of poor NBME performance (<70 pts). All *P*-values were 2-sided and $P < .05$ was considered statistically significant. All data analyses were conducted using STATA version 14 (College Station, TX). This study was approved by our institutional review board.

RESULTS

Median scores for each week's quiz topic stratified by year are outlined in Table 1b. Overall aggregate quiz score for both academic years was calculated to be 78 with as interquartile range (IQR) of 73–85. The 2015–2016 academic year (n = 79) had an aggregate score of 83 and an IQR of 75–87. The 2016–2017 academic year (n = 77) had an aggregate score of 76.5 with an IQR of 69.5–81. Table 2 outlines median scores for each topic stratified by academic block. Block 1 (n = 40) had an aggregate quiz score of 67 and an IQR of 64–69.5. Block 2 (n = 40) had an aggregate quiz score of 77 and an IQR of 74.5–80. Block 3 had an aggregate quiz score of 76.5 and an IQR of 67–89.5. Block 4 had an aggregate quiz score of 83 and an IQR of 76–85.

For the 2015–2016 academic year, the mean NBME score was 78.5 with a standard deviation (SD) of 8.6 ($P = .04$). For the 2016–2017 academic year, the mean NBME score was 75.6 with a SD of 6.9 ($P = .04$). Block 1 (n = 40) had a mean NBME score of 74.1 with a SD of 6.9. Block 2 (n = 40) had a mean NBME score of 76.7 with a SD 7.8. Block 3 (n = 37) had a mean NBME score of 77.6 with a SD of 9.2. Block 4 (n = 39) had a mean NBME score of 80.2 with a SD of 6.9. Fig 1 illustrates NBME scores and quiz scores as a function of academic year and block. Table 3 models NBME performance as a function of quiz scores.

Week 2 Trauma/Burns content corresponded with a β-coefficient of 1.57 ($P < .001$), 1.36 ($P < .001$), and 1.25 ($P = .002$) for models 1, 2, and 3, respectively. Week 3 Esophagus/Anorectal content corresponded with a β-coefficient of 1.42 ($P < .001$), 1.03 ($P = .006$), and .97 ($P = .01$) for models 1, 2, and 3, respectively. Week 7 Wound/Intensive

Table 1b
Weekly quiz performance by academic year

Topics/Weeks	Overall N = 156	2015–2016 N = 79	2016–2017 N = 77	P-value*
	Median (IQR)	Median (IQR)	Median (IQR)	
Acute Abdomen/Radiology	7 (6, 8)	7 (7, 8)	7 (6, 8)	.004
Trauma/Burns	6 (5, 7)	6 (5, 7)	6 (5, 7)	.48
Esophagus/Anorectal	7.5 (6, 9)	8 (7, 9)	7 (6, 8)	<.001
Colon/Gallbladder	7 (6, 8)	7 (6, 8)	6 (5, 7)	<.001
Small Bowel/HPB	7 (6, 8)	7 (6, 8)	7 (6, 8)	.01
Endocrine/Breast	7 (5, 8)	7 (5, 10)	6 (5, 8)	.007
Wound/ICU	8 (7, 9)	9 (8, 10)	7 (6, 8)	<.001
Anesthesia	3 (2, 3)	3 (2, 4)	2 (1, 3)	<.001
Dermatology/Hernia	7 (6, 8)	7 (6, 8)	6 (5, 7)	<.001
Urology/Pediatrics	8 (6, 9)	6 (5, 8)	9 (8, 9)	<.001
ENT/Orthopedics	6 (5, 8)	6 (4, 7)	6 (6, 8)	.15
Cardiac/Vascular	5 (5, 7)	5 (5, 5)	7 (5, 8)	<.001
Total Score	78 (73, 85)	83 (75, 87)	76.5 (69.5, 81)	<.001

Academic year 2016–2017 did not include neurosurgery.
Weekly quiz performance stratified by academic year and overall performance noted with aggregate scoring.
* Wilcoxon rank sum used to compare topics between academic years.

Table 2
Weekly quiz performance by academic block

Topics/Weeks	Block 1 N = 40	Block 2 N = 40	Block 3 N = 37	Block 4 N = 39	P-value*
	Median (IQR)	Median (IQR)	Median (IQR)	Median (IQR)	
Acute Abdomen/Radiology	7.5 (7, 8.5)	7 (6, 8)	7 (6, 8)	7 (6, 8)	.07
Trauma/Burns	6 (4.5, 7)	6 (4, 7)	6 (6, 8)	7 (6, 8)	.01
Esophagus/Anorectal	7 (6, 8)	7 (6, 8)	8 (6, 9)	8 (6, 9)	.03
Colon/ Gallbladder	7 (6, 7)	7 (6.5, 8)	7 (6, 9)	7 (5, 8)	.23
Small Bowel/HPB	7 (6, 8)	8 (6, 8.5)	7 (6, 7)	8 (6, 8)	.14
Endocrine/Breast	6 (5, 7)	6 (5, 7)	7 (5, 9)	8 (7, 10)	<.001
Wound/ICU	8 (7, 9)	8 (7, 9)	8 (7, 9)	9 (7, 10)	.07
Anesthesia	2 (2, 3)	3 (2, 4)	3 (2, 3)	3 (1, 4)	.22
Dermatology/Hernia	7 (6, 8)	7 (6, 8)	7 (5, 9)	7 (5, 8)	.12
Urology/Pediatrics	8 (6, 10)	7 (6, 8.5)	7 (6, 9)	8 (6, 9)	.54
ENT/Orthopedics	6.5 (6, 7)	6 (6, 7)	7 (6, 8)	4 (3, 8)	.27
Cardiac/Vascular	5 (4, 5)	5 (5, 8)	5 (4, 7)	5 (5, 8)	.002
Total Score	67 (64, 69.5)	77 (74.5, 80)	76.5 (67, 89.5)	83 (76, 85)	<.001

Academic year 2016–2017 did not include neurosurgery.
Weekly quiz performance stratified by academic block and overall performance noted with aggregate scoring.
* Kruskal-Wallis test used to compare topics between blocks.

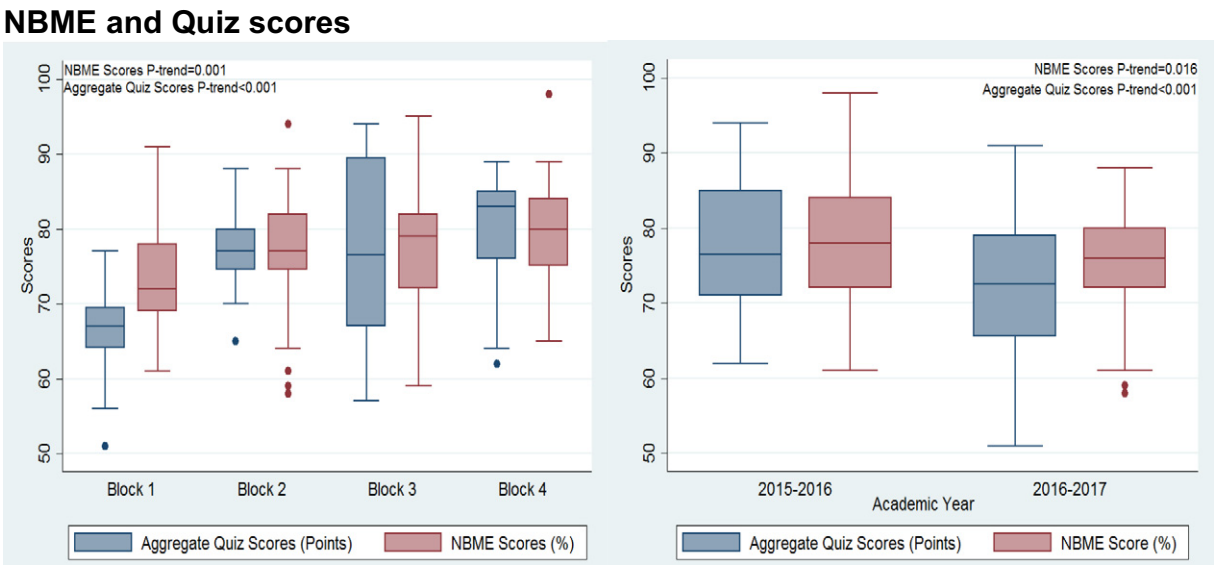

Fig 1. NBME and quiz scores.

Table 3
Model of NBME performance

Predictors	Model 1		Model 2		Model 3	
	β (95% CI)	P-value	β (95% CI)	P-value	β (95% CI)	P-value
Trauma/Burns	1.57 (0.79, 2.35)	<0.001	1.36 (0.61, 2.12)	<0.001	1.25 (0.46, 2.02)	0.002
Esophagus/Anorectal	1.42 (0.73, 2.10)	<0.001	1.03 (0.30, 1.75)	0.006	0.97 (0.22, 1.73)	0.01
Wound/ICU	1.38 (0.68 (2.08)	<0.001	0.88 (0.13, 1.62)	0.02	0.96 (0.12, 1.80)	0.02

Relative NBME performance as a function of quiz performance with β-coefficients.
Model 1 reports univariate analysis using Linear Regression.
Model 2 reports Beta coefficients using multiple linear regressions adjusting for all other variable in the model.
Model 3 contains variables from Model 2 with further adjustment of year and academic block.

Care Unit (ICU) content corresponded with a β-coefficient of 1.38 ($P < .001$), 0.88 ($P = .02$), and 0.96 ($P = .02$) for models 1, 2, and 3, respectively. Table 4 shows the likelihood of scoring <70 points as a function of quiz performance. Week 7 Wound/ICU content corresponded with an odds ratio of 0.75 ($P = .03$), 0.65 ($P = .008$), and 0.71 ($P = .06$) for models 1, 2, and 3, respectively. Week 12 Cardio/Vascular content corresponded with an odds ratio of 0.68 ($P = .02$), 0.60 ($P = .006$), and 0.60 ($P = .03$) for models 1, 2, and 3, respectively.

DISCUSSION

Current literature surrounding the use of weekly quizzes as a leading indicator of NBME surgical shelf performance is lacking. Recent studies have examined the accuracy of subjective student evaluations in forecasting performance on the NBME surgery shelf [1–4]. Other studies have demonstrated the importance of clinical experience and surgical case volume on student performance [5]. However, to the best of our knowledge the use of intra-clerkship weekly quizzes as a predictive measure of NBME surgery shelf performance has not been examined to date.

Review of the quiz content and corresponding NBME outcomes illustrates that content related to Trauma/Burns, Esophagus/Anorectal, Wound/ICU is correlated with increased performance on the NBME surgery shelf exam. This content may be integral to the NBME and represent "high yield" teaching points. Further, content related to Wound/ICU was observed to be both associated with increased NBME scores and an increase in the likelihood of scoring above 70 pts. Cardio/Vascular content was also associated with an increased likelihood of scoring above 70 pts. This again may represent content that is highly tested. Analysis of quiz performance based on academic year showed that the 2015–2016 cohort (n = 79) outperformed the 2016–2017 cohort (n = 77) with aggregate quiz scores of 83 and 76.5 respectively ($P < .001$). Interestingly, NBME surgery shelf performance showed a corresponding relative change for each cohort as shown in Fig 1. Considering no significant curricular changes were made between these academic years, the observed inter-class variability may be within

Table 4
Model of NBME outcome

Predictors	Model 1		Model 2		Model 3	
	OR (95% CI)	P-value	OR (95% CI)	P-value	OR (95% CI)	P-value
Wound/ICU	0.75 (0.59, 0.96)	0.03	0.65 (0.48, 0.89)	0.008	0.71 (0.50, 1.01)	.06
Cardiac/Vascular	0.68 (0.48, 0.95)	0.02	0.60 (0.41, 0.86)	0.006	0.60 (0.38, 0.93)	.03

Likelihood of scoring <70 points on the NBME as a function of quiz performance.
Model 1 reports univariate analysis using Linear Regression.
Model 2 reports odds ratios (95% CI) using multiple logistic regression adjusting for all other variable in the model.
Model 3 contains variables from model 2 with further adjustment of year and block.

standard variation. However, more data is needed in order to determine the significance of this observation. Notwithstanding the inter-class variability, the relative change in NBME and quiz scores suggests that there is a link between quiz performance and NBME scores. Further analysis of aggregate scores stratified by academic block suggests that student performance on both the intra-clerkship quizzes and the NBME are correlated, and have a temporal component. That is, students who take the surgery clerkship later in their academic year have better quiz scores and NBME performance. This makes intuitive sense given that students would be expected to have a more robust fund of knowledge, clinical acumen, and more mature test taking strategies. In this context, it is important to note that Students were able to view their quizzes but did not take them home, or document during the review process. It is possible that questions were disseminated by memory, but this is a violation of the institutional honor code. Further, no evidence of such behavior was noted.

With potential student career planning and curricular development implications, these findings certainly warrant further investigation. Ultimately, this study correlates NBME outcomes with weekly quizzes and substantiates the concept that intra-clerkship quizzes could be used to identify students at risk of poor NBME outcomes. If this subset of students can be identified early, perhaps they could be mentored through adjunct educational models. Unconventional approaches such as case-based learning, mobile application use, and resident/student-led teaching models have shown promise in improving student performance on the NBME surgery shelf and could be of utility in future investigation [6–8].

This study was done in the context of a 12-week clerkship, though it may be possible to apply these findings to shorter length clerkships as well. By reducing the number of quizzes and focusing on content more highly-correlated with increased NBME performance, struggling students could still benefit from early identification in shorter surgical clerkships. This study is intended to explore and set the stage for future explorations into possible interventions and their respective efficacy. Although no Interventions were designed or implemented here, it is possible that future studies can further objectify the effectiveness, utility, and efficacy of intra-clerkship quizzes in the context of clerkships of varying length.

In conclusion, this study demonstrates the viability of modeling NBME surgical shelf exam outcomes as a function of weekly quiz performance during a 12-week surgery clerkship. Quizzes with content corresponding with Wound/ICU, Esophagus/Anorectal, and Trauma/Burns, demonstrated a relative increase in NBME surgical shelf exam performance. Wound/ICU and Cardio/Vascular quizzes were predictive of students scoring <70 points on the NBME surgical shelf exam. Students participating in the surgery clerkship later in the third year have increased performance on weekly quizzes and the NBME surgical shelf exam. Further examination of "high-yield" subject matter, the timing of the surgical clerkship, and the use of intra-clerkship quizzes, has the potential to not only predict student performance on the NBME surgical shelf exam, but provide them with additional information to succeed in their surgical clerkship.

Author contributions

Dr. Jamil Jaber: Primary author
Dr. Natasha Keric: Author and editor
Mr. Paul Kang: Statistical analysis
Dr. Ara Feinstein: Author and editor

Conflict of interest

None of the authors have any conflicts of interest to report.

Funding sources

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

[1] Butler KL, Hirsch DA, Petrusa ER, et al. Surgery clerkship evaluations are insufficient for clinical skills appraisal: the value of a medical student surgical objective structured clinical examination. J Surg Educ 2017;74(2):286–94.
[2] Bowen RE, Grant WJ, Schenarts KD. The sum is greater than its parts: clinical evaluations and grade inflation in the surgery clerkship. Am J Surg 2015;204(4):760–4.
[3] Farrell TM, Kohn GP, Owen SM, et al. Low correlation between subjective and objective measures of knowledge on surgery clerkships. J Am Coll Surg 2010;210(5):680–3.
[4] Goldstein SD, Lindeman B, Colbert-Getz J, et al. Faculty and resident evaluations of medical students on a surgery clerkship correlate poorly with standardized exam scores. Am J Surg 2014;207(2):231–5.
[5] Myers JA, Vigneswaran Y, Gabryszak B, et al. NBME subject examination in surgery scores correlate with surgery clerkship clinical experience. J Surg Educ 2014;71(2):205–10.
[6] Cendan JC, Silver M, Ben-David K. Changing the student clerkship from traditional lectures to small group case-based sessions benefits the student and the faculty. J Surg Educ 2011;68(2):117–20.
[7] Schwartz RW, Donnelly MB, Nash PP, et al. Problem-based learning: an effective educational method for a surgery clerkship. J Surg Res 1992;53(4):326–30.
[8] Wirth K, Malone B, Turner C, et al. A structured teaching curriculum for medical students improves their performance on the National Board of medical examiners shelf examination in surgery. Am J Surg 2015;209(4):765–70.