Journal of Contemporary Medical Education

available at www.scopemed.org, www.jcmededu.org

Original Research

The effect of rescheduling National Board of Medical Examiners subject examinations on United States Medical Licensing Examination Step 2 Clinical Knowledge performance

Brian Rutledge, Gerald Clark, Jessica Bailey, Lou A Woodward

University of Mississippi Medical Center, School of Medicine, Mississippi, USA

Received: August 31, 2013

Accepted: December 16, 2013

Published Online: March 08, 2014

DOI: 10.5455/jcme.20131216020050

Corresponding Author:
Brian Rutledge,
2500 North State Street, Jackson,
Mississippi 39216
brutledge@umc.edu

Keywords: *NBME, USMLE, Examination scheduling, Medical curriculum*

ABSTRACT

Historically, the University of Mississippi School of Medicine administered all National Board of Medical Examiners subject examinations at the end of the third year of the medical curriculum, just before the comprehensive United States Medical Licensing Examination Step 2 Clinical Knowledge examination. Due in part to grading and accreditation concerns, subject examinations were moved in a phased fashion from the end of the year to the end of the clerkship associated with each subject examination; this research aims to determine the effect of this rescheduling. Multiple hierarchical regression was performed on Step 2 Clinical Knowledge examination scores by whether the subject examination was given at end-of-clerkship or end-of-year, while controlling for covariates known to affect examination performance. This analysis found that performance on the Step 2 Clinical Knowledge examination is not significantly affected by the scheduling of subject examinations. The benefits gained from moving the subject examinations from end-of-year to end-of-clerkship, such as timely feedback and compliance with accreditation standards, do not appear to come at the cost of decreased performance on the Step 2 Clinical Knowledge examination.

INTRODUCTION

The University of Mississippi Medical Center (UMMC) School of Medicine (SOM) educates and trains students through a four-year curriculum. In the third year of this curriculum, students progress asynchronously in groups through seven required clinical clerkships [1]. Also in the third year, students take both the National Board of Medical Examiners (NBME) subject examinations in the subject area of six of the clerkships and the comprehensive United States Medical Licensing Examination (USMLE) Step 2 Clinical Knowledge examination. These examinations are used in several important ways: calculating final clerkship grades (NBME) [2-4], as a requirement for promotion (NBME) [3], and as a requirement for graduation (USMLE) [3,5,6]. They are also both used heavily by residency program directors when evaluating residency applicants [2,6-11].

Because of the significance of these examinations, SOM administration, faculty, and students all seek ways to improve performance on these examinations while simultaneously maximizing learning and maintaining compliance with accreditation standards. One such standard requires timely formative and summative feedback from faculty in courses and clerkships [12].

Until recently, the UMMC SOM administered all NBME subject examinations at the end of the third year, just before the administration of the comprehensive USMLE Step 2 Clinical Knowledge

examination. NBME subject examinations were moved in a phased fashion from the end of the third year to the end of the clerkship associated with each NBME subject examination (see Table 1). There are two primary reasons for this change: to allow more timely summative feedback that includes performance on the NBME subject examination and to raise NBME subject examination scores (based on the assumption that more knowledge will be retained with examinations administered immediately following the related clinical clerkship).

Studying the effects of this change is critical to determining whether it will have the intended results or an unintended result: negatively impacting the high stakes USMLE Step 2 Clinical Knowledge examination scores by eliminating the end-of-year NBME subject examinations which served as a required comprehensive review for the USMLE Step 2 Clinical Knowledge examination.

The purpose of this research is to determine the relationship between the scheduling of NBME subject examinations and performance on the United States Medical Licensing Examination (USMLE) Step 2 Clinical Knowledge examination. This research will also add to the base of knowledge in which there is currently a focused gap regarding the specific effect of scheduling of the NBME subject examinations on the performance on the USMLE Step 2 Clinical Knowledge examination.

There is existing literature that shows the relationship between clerkship scheduling and NBME subject examination performance [13], the predictive relationship between NBME subject examination performance and USMLE Step 1 examination performance [14], the predictive relationship between NBME subject examination performance and USMLE Step 2 Clinical Knowledge examination performance, and how curriculum changes can affect performance on the NBME subject examinations [15] and the USMLE Step 2 Clinical Knowledge examination [16,17]. However, there is a gap in knowledge about the specific effect of scheduling of the NBME subject examinations on USMLE Step 2 Clinical Knowledge examination scores.

The idea that scheduling of NBME subject examinations may affect examination performance is not a new one. Even the NBME acknowledges the effect of scheduling on examination performance [18]. In recent years, however, few studies have addressed this area specifically.

In general, the current state of knowledge shows that taking NBME subject examinations later in the year leads to higher scores [2,5,8,19]. This is particularly true for the internal medicine clerkship that draws upon clinical knowledge acquired over a variety of clerkships [2]. In 2010, a study of over 2,000 medical students confirmed that clerkship order was a significant determinant of performance on NBME subject examinations, but did not find that clerkship order was a significant determinant of performance on the USMLE Step 2 Clinical Knowledge examination. Note that this study did not address whether NBME subject examination *scheduling* was a significant determinant of USMLE Step 2 Clinical Knowledge examination performance.

Since there is no requirement from the NBME about when NBME subject examinations must be administered, schools may administer them at the end of each clerkship [9,20], at the end of the semester, or at the end of the third year [2].

Table 1. Change in NBME subject examination scheduling

	NBME subject examinations given at the end of the:				
	Group 0 (reference group)		**Group 1**		**Group 2**
	2007-2008	**2008-2009**	**2009-2010**	**2010-2011**	**2011-2012**
Family Medicine	Year	Year	Year	Year	Clerkship
Internal Medicine	Year	Year	Year	Year	Clerkship
Ob-Gyn	Year	Year	Clerkship	Clerkship	Clerkship
Pediatrics	Year	Year	Year	Year	Clerkship
Psychiatry	Year	Year	Clerkship	Clerkship	Clerkship
Surgery	Year	Year	Year	Year	Clerkship

Like NBME subject examinations, USMLE Step 2 Clinical Knowledge examination performance is also subject to being affected by scheduling. It has been shown that students who take the USMLE Step 2 Clinical Knowledge examination soon after the end of the third year have higher average scores than students who delayed taking the examination into the fourth year [6]. It should be noted that one study showed the opposite: scores on the USMLE Step 2 Clinical Knowledge were higher when taken later in the year [21]. However, this study was flawed due to its small sample size, high failure rate, and use of sample USMLE examination questions rather than actual examination questions [6].

There are other predictors of performance on the USMLE Step 2 Clinical Knowledge examination besides scheduling. Performance on premedical tests (MCAT), performance on preclinical standardized tests (USMLE Step 1 examination), and performance in clerkships (both NBME subject examination scores and final grades) are all predictive of performance on the USMLE Step 2 Clinical Knowledge examination [22]. In addition, race and gender can be predictive of performance on standardized examinations [23-27].

Upon review of the literature, the significance of this research was further strengthened by the confirmation of the many effects of examination scheduling and other factors on student examination performance.

METHODS

This study employed data obtained from the UMMC SOM including matriculation year, USMLE identification number, race, gender, undergraduate grade point average (in biology, chemistry, physics, and mathematics), MCAT score, USMLE Step 1 score, USMLE Step 2 Clinical Knowledge score, and scores in six NBME subject examinations: family medicine, surgery, obstetrics and gynecology, medicine, pediatrics, and psychiatry. These data were collected for students enrolled in the third year during academic years 2007-2008, 2008-2009, 2009-2010, 2010-2011, and 2011-2012. Because each of these groups is defined by no more than a one-year difference, one-year variations are minimal and differences in the examined relationships are not expected.

Because this new research activity presented no more than minimal risk to human participants, the UMMC Institutional Review Board granted expedited review of the research protocol. The IRB defines minimal risk by as "the probability and magnitude of harm or discomfort anticipated in the research are not greater in and of themselves than those ordinarily encountered in daily life or during the performance of routine physical or psychological examinations or tests" [28].

Sampling Procedures

In part because of this phased nature of the change in NBME subject examination scheduling, the convenience sampling method was chosen to select all medical students in the appropriate curricular year as the sample for this research. Additionally, this sampling method is common when analyzing all of a group of students at a single institution or school and also for detecting relationships among phenomena in a specific group, both of which are characteristics of this research. It is particularly useful to document that a particular phenomenon occurs within the group selected for the sample [29].

Data Coding

In order to use a hierarchical multiple regression method (the primary statistical treatment to be discussed in the following section), categorical variables with more than two levels must be "dummy coded" in order to be meaningfully interpreted [30,31]. In this research, only the *race* and *scheduling group* variables are categorical with more than two levels.

Preliminary Analyses

Standard descriptive statistical analyses were performed for each of the numerical dependent and independent variables. This analysis included minimum, maximum, mean, median, and standard deviation. A frequency distribution was performed for the demographic variables. Descriptive statistics for the scheduling group was not performed because though the values are numerical, they are nominal only.

Hierarchical multiple regression requires that the minimum ratio of cases to independent variables be 5:1 [32]. With an N of over 500, this sample size ratio of at least 125:1 easily accommodated the six independent variables (GPA, MCAT score, USMLE Step 1 score, scheduling group, race, and gender).

With the sample size requirement met, the assumptions for multiple regression must be met. There are many assumption violations for multiple regression, [33] but this study focused on validating three major assumptions: normality, linearity, and homogeneity of variance [32,34].

Once these three assumptions for regression of the preliminary analyses are met or the data are transformed, the primary analysis can then be performed. The primary analysis is described in the following section.

Primary Analysis

The primary statistical treatment for this research was hierarchical multiple regression. While researchers use

a standard multiple regression to study the relationship between multiple independent variables and one dependent variable, researchers can use a hierarchical multiple regression to do the same while controlling for the effect of a separate set of independent variables on the dependent variable [32]. This model also accommodates the type of data used in this research: a continuous dependent variable with a mix of continuous and categorical independent variables. In hierarchical multiple regression, researchers enter the control independent variables or covariates into the model first (block 1), and then they enter the focus independent variables into the model (block 2) [35].

In this research, the focus is to determine the relationship between NBME subject examination scheduling and performance on the USMLE step 2 Clinical Knowledge examination, but the literature suggests that other variables such as GPA, MCAT score, Step 1 score, race, and gender, may also affect performance. To ensure that these variables do not interfere with evaluating just the relationship between the scheduling of NBME subject examinations and the dependent variable (the USMLE Step 2 Clinical Knowledge examination score), the covariates, or control variables, are entered into the regression model first, as block 1. The focus independent variable (scheduling of NBME subject examinations) is then entered into the model in block 2. To examine the effect of scheduling and clerkship on the dependent variable and to account for the fact the hybrid group is non-randomized, we also examined interactions between NBME scores for the six subject areas and the three different scheduling groups. Findings were not significant and are not included in the tables.

The statistical model used for testing is shown in table 2 below:

Table 2. Statistical Model

Dependent variable	Independent variables – Block 1 (control)	Independent variables – Block 2 (focus)
Step 2 CK score	BCPM GPA MCAT score Step 1 score Race (four dummy variables) Gender	Scheduling group (two dummy coded variables)

The output also has values for R, R^2, and adjusted R^2. Unique to hierarchical multiple regression, it is the change in R^2 after the addition of block 2 (focus independent variable) to block 1 (control independent variables) that is analyzed, rather than the R^2 value for the overall model with all variables entered [32]. This measure explains how much variation in the dependent variable(s) can be explained by variation of the independent variable(s).

Table 3. Frequencies for demographics

	Frequency	Percent
Race		
White	431	81.8
Hispanic	4	0.8
Black	50	9.5
Asian	40	7.6
American Indian	2	0.4
Gender		
Female	231	43.8
Male	296	56.2
Scheduling Group		
Clerkship	110	20.9
Hybrid	211	40.0
Year	206	39.1

RESULTS

Descriptive Statistics

Table 4 displays the descriptive statistics for the independent variable, covariates, and the dependent variables.

Data Coding

In order to be appropriately included in the primary analysis (hierarchical multiple regression), the two categorical variables in this research that have more than two levels (*race* and *scheduling group*) must be "dummy coded." This method of coding allows the use of categorical predictor variables, using only ones and zeros to convey all of the necessary information about group membership [36].

Preliminary Analyses

The correlations of each of the 11 variables were analyzed using Spearman correlation.

The three selected assumptions (normality, linearity, and homogeneity of variance) for hierarchical multiple regression were tested and were met so the primary analysis was performed.

Two common tests of normality are the Kolmogorov-Smirnov and Shapiro-Wilk tests. Because the Shapiro-Wilk test is more appropriate for small sample sizes (i.e., less than 50 subjects), in this research the Kolmogorov-Smirnov test was used [30,37]. For this test, if the result is significant ($p < 0.05$), the data are not normally distributed. If the result is not significant ($p > 0.05$), the data are normally distributed and meet the assumption of normality.

Table 4. Descriptive statistics – all subjects

	N	Mean	Median	Minimum	Maximum	Std.Dev
GPA	527	3.64	3.69	2.28	4	0.32
MCAT	527	27.99	28.00	19	42	3.46
Step1	527	220.89	223.00	145	267	22.70
Step2	527	233.19	235.00	164	280	21.76
FamMed	523	74.07	73.00	53	98	7.97
Medicine	527	76.34	76.00	54	99	8.33
Ob-gyn	525	69.94	70.00	45	98	7.70
Pediatrics	527	74.57	74.00	55	99	8.62
Psychiatry	527	74.46	74.00	54	98	7.98
Surgery	527	72.70	72.00	47	99	8.53

Table 5 shows that the results of the Kolmogorov-Smirnov tests for each variable was not significant as each was greater than 0.05. Only one variable, the NBME Surgery scores, had to be logarithmically transformed in order to meet the assumption of normality.

Table 5. Kolmogorov-Smirnov test of normality

	Kolmogorov-Smirnov
GPA	0.134
MCAT	0.077
Step1Score	0.059
DHispanic	0.527
DBlack	0.533
DAsian	0.537
DAmerIndian	0.521
Gender	0.375
DSchedHybrid	0.391
DSchedClerkship	0.491
Step2	0.053
FamilyMedicine	0.068
Medicine	0.058
OB	0.056
Pediatrics	0.075
Psychiatry	0.051
Surgery*	0.051

*This variable was corrected through logarithmic transformation.The assumption of linearity was confirmed because the correlation analysis showed that the focus independent variable has a significant linear relationship to the dependent variables (see the shaded cells in Tables 6 and 6a).

The assumption of homogeneity of variance was confirmed because the variance of the dependent variables within the population was found to be equal when using the Levene test. A p value of > 0.05 (unlike testing for significance when a p value of < 0.05 is important) validated this assumption (see Table 7: Levene's test of equality of error variances).

Primary Analysis

With the preliminary analyses complete, the primary analysis (hierarchical multiple regression) was performed. Per the research design, variables were entered in two blocks with USMLE Step 2 Clinical Knowledge examination scores as the dependent variable. The results of the regression analysis are shown in table 8.

As the results indicate, the p value for both of the dummy coded NBME subject examination scheduling variables are greater than α of 0.05 so the null hypotheses is not rejected. Therefore, the scheduling of NBME subject examinations does not significantly affect performance on the scheduling of NBME subject examinations after controlling for GPA, MCAT score, Step 1 score, race, and gender.

In addition to the results of the hypothesis test, the regression analysis also revealed other significant relationships between performance on the USMLE Step 2 examination and the other variables. In this model, GPA and Step 1 score had a positive significant relationship to on USMLE Step 2 examination score, while Gender had a negative significant relationship (since this variable was coded 1=female and 2=male, this means that females performed better than males on the USMLE Step 2 examination).

The results also included R^2 results. As shown in table 9, the R^2 values for the regression for block 1 and block 2 were not significantly different at 0.605 and 0.607, respectively. Since the focus of hierarchical multiple regression is the *change* in R^2 values between block 1 and block 2, this change of 0.002 further indicates that the scheduling of NBME subject examinations cannot significantly explain variation in performance on the USMLE Step 2 Clinical Knowledge examination.

Table 6. Spearman Correlation Matrix

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	1	.169**	.321**	-0.009	-.196**	0.028	-0.056	-0.05	0.015	-.031**	.329**	0.303	.341**	0.299	0.32	1	.169**
2	.169**	1	.388**	0.039	-.404**	.096*	-0.063	.257**	-.010**	0.041	.322**	0.239	.250**	.258*	0.247	.169**	1
3	.321**	.388**	1	0.029	-.233**	-0.004	-0.056	.167**	.165**	.050**	0.798	0.675	.688**	0.601	0.68	.321**	.388**
4	-0.009	0.039	0.029	1	-0.028	-0.025	-0.005	-0.055	-0.071	-0.045	0.029	0.036	0.046	0.021	0.026	-0.009	0.039
5	-.196**	-.404**	-.233**	-0.028	1	-.093*	-0.02	-.223**	-.027**	.041**	-.209**	-0.182	-0.244	-.107*	-0.181	-.196**	-.404**
6	0.028	.096*	-0.004	-0.025	-.093*	1	-0.018	-.137**	0.044	-.041*	0.004	-0.047	-.044*	-0.02	0.015	0.028	.096*
7	-0.056	-0.063	-0.056	-0.005	-0.02	-0.018	1	-0.008	-0.05	0.044	-0.054	-0.052	-0.015	-0.058	-0.049	-0.056	-0.063
8	-0.05	.257**	.167**	-0.055	-.223**	-.137**	-0.008	1	-0.012	-.036**	.060**	0.041	.112**	-.101**	0.027	-0.05	.257**
9	0.015	-0.01	.165**	-0.071	-0.027	0.044	-0.05	-0.012	1	-0.42	.079**	0.271	0.053	0.128	0.183	0.015	-0.01
10	-0.031	0.041	0.05	-0.045	0.041	-0.041	0.044	-0.036	-0.42	1	0.049	-0.243	-0.133	-0.005	-0.056	-0.031	0.041
11	.329**	.322**	.798**	0.029	-.209**	0.004	-0.054	0.06	*.079***	*.049***	1.000**	0.721	.729**	0.672	0.724	.329**	.322**
12	.303**	.239**	.675**	0.036	-.182**	-0.047	-0.052	0.041	*.271***	*-.243***	.721**	1	.694**	0.588	0.69	.303**	.239**
13	.341**	.250**	.688**	0.046	-.244**	-0.044	-0.015	.112**	*.053***	*-.133***	.729**	0.694	1.000**	0.617	0.689	.341**	.250**
14	.299**	.258**	.601**	0.021	-.107*	-0.02	-0.058	-.101*	*.128***	*-.005***	.672**	0.588	.617*	1	0.652	.299**	.258**
15	.320**	.247**	.680**	0.026	-.181**	0.015	-0.049	0.027	*.183***	*-.056***	.724**	0.69	.689**	0.652	1	.320**	.247**
16	.292**	.227**	.593**	0.009	-.104*	-0.026	0	-.106*	*.098***	*.068***	.651**	0.607	.586*	0.581	0.635	.292**	.227**
17	.305**	.324**	.690**	0.055	-.277**	-0.015	-0.063	.183**	*.119***	*-.172***	.715**	0.685	.730**	0.575	0.666	.305**	.324**

*p<.05, **p<.01

Table 6a. Spearman Correlation Matrix Key

Code	Variable
1	GPA
2	MCAT
3	Step1Score
4	DHispanic
5	DBlack
6	DAsian
7	DAmerIndian
8	Gender
9	DSchedHybrid
10	DSchedClerkship
11	Step2
12	FamilyMedicine
13	Medicine
14	OB
15	Pediatrics
16	Psychiatry
17	Surgery

Table 7. Levene's test of equality of error variances

	F	Df1	Df2	Sig.
Step2	4.150	2	524	.016
FamilyMedicine	.190	2	520	.827
Medicine	1.671	2	524	.189
OB	2.381	2	522	.093
Pediatrics	.774	2	524	.462
Psychiatry	1.112	2	524	.330
Surgery	3.367	2	524	.053

Knowledge examination.

DISCUSSION

The major findings of this research suggest that there were not significant differences in USMLE Step 2 Clinical Knowledge scores based on when NBME subject examinations were scheduled, while controlling for covariates. This means that the benefits gained from moving the NBME subject examinations from end-of-year to end-of-clerkship, such as timely

feedback and compliance with accreditation standards, do not appear to come at the cost of decreased performance on the USMLE Step 2 Clinical Knowledge examination.

Limitations

For both the USMLE Step 2 Clinical Knowledge examination and NBME subject examinations, the testing format moved from written to 100% computer-based testing during the period for which data were analyzed in this research. Although the NBME purports to have controlled for any risk or benefit to the student caused by this transition, this is another potential source of influence on the findings.

Another limitation could potentially exist because of student choice. Because students have some control over the scheduling of the order of their third year clerkships, personal characteristics may preferentially influence these selections, but should be minimized by the initial random assignment to third year student groups.

Generalizability

The results of this research may be of interest to international medical educators whose curricula are similar to the one described here. Recent research has shown that examination content, including NBME and USMLE content, that is used in the United States still applies outside the United States.

In 2010, a collaboration between the National Board of Medical Examiners and four schools in the United Kingdom demonstrated that examination items on the USMLE Step 2 Clinical Knowledge examination are adaptable and can be localized to international medical schools. [38]

Additionally, research has also shown that general testing methodologies and understandings of Western medical education are also applicable in other countries such as Germany, Italy, South Africa, and the Netherlands [39-43], and even in resource-poor countries such as Mozambique [44]. Once cultural and language difference are taken into account, the testing methodologies utilized for the USMLE Step 2 Clinical Knowledge examination and the NBME subject examinations are valid and applicable across the world [45].

Table 8. Regression results

Block		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	56.275	9.295		6.054	0
	GPA	5.98	1.994	0.09	2.998	0.003
	MCAT	0.157	0.207	0.025	0.755	0.45
	Step1Score	0.71	0.03	0.741	23.509	0
	Gender	-3.775	1.297	-0.086	-2.911	0.004
	DHispanic	-4.862	6.946	-0.019	-0.7	0.484
	DBlack	-0.349	2.374	-0.005	-0.147	0.883
	DAsian	-1.41	2.323	-0.017	-0.607	0.544
	DAmerIndian	-7.395	9.805	-0.021	-0.754	0.451
2	(Constant)	56.206	9.297		6.046	0
	GPA	5.868	1.997	0.088	2.939	0.003
	MCAT	0.14	0.208	0.022	0.672	0.502
	Step1Score	0.719	0.031	0.75	23.147	0
	Gender	-3.859	1.301	-0.088	-2.967	0.003
	DHispanic	-5.815	6.997	-0.023	-0.831	0.406
	DBlack	-0.37	2.377	-0.005	-0.156	0.876
	DAsian	-1.307	2.327	-0.016	-0.562	0.575
	DAmerIndian	-7.893	9.821	-0.022	-0.804	0.422
	DSchedHybrid	-1.899	1.395	-0.043	-1.362	0.174
	DSchedClerkship	-0.613	1.66	-0.011	-0.369	0.712

Table 9. R Square

Model	df	R	R Square	Adjusted R Square	Std. Error of Estimate
1	8	.778	0.605	0.601	13.753
2	10	.779	0.607	0.6	13.762

Future research

Because 2011-2012 is only the first academic year during which all NBME subject examinations are administered at end-of-clerkship, the analysis and comparison of scores should continue in future years to further study the effect of scheduling on performance and to determine if the trends described in this chapter continue. Also because 2011-2012 is only the first academic year during which all NBME subject examinations are administered at end-of-clerkship, the UMMC SOM had not yet fully encountered the well-known challenge of clerkship scheduling bias. Now, because NBME subject examinations have been moved to end-of-clerkship, further study on the effects of clerkship timing specifically for UMMC SOM students may be warranted.

Regardless of the results of this research, the focus of medical education researchers is the student. Providing a high-quality, appropriate curriculum with associated accurate, valid, and unbiased assessment tools is the obligation of any school of medicine. Constant, in-depth inquiry into medical education practices, such as this research, is essential to meet this obligation to students.

List of abbreviations

AAMC Association of American Medical Colleges

CK Clinical Knowledge

GPA Grade point average

MCAT Medical College Admissions Test

NBME National Board of Medical Examiners

SOM School of Medicine

UMMC University of Mississippi Medical Center

USMLE United States Medical Licensing Examination

DECLARATION OF INTEREST

The authors declare they have no conflict of interest.

AUTHORS' INFORMATION

Brian L. Rutledge, PhD, MHSA was director of accreditation and assessment for the School of Medicine at the time of this research, and is now chief of staff to the vice chancellor and assistant professor of health sciences at the University of Mississippi Medical Center.

Gerald F. Clark, PhD, MBA is the associate dean for student affairs for the School of Medicine, assistant professor of medicine, and chief student affairs officer at the University of Mississippi Medical Center.

Jessica H. Bailey, PhD, RHIA is interim dean and professor of health sciences for the School of Health Related Professions at the University of Mississippi Medical Center.

LouAnn H. Woodward, MD, is associate vice chancellor for health affairs, vice dean of the School of Medicine, and professor of emergency medicine at the University of Mississippi Medical Center.

REFERENCES

1. M3 Block Schedule. Available via http://studentservices.umc.edu/documents/m3_blocks_schedule.pdf (Accessed June 2012).
2. Widmann WD, Aranoff T, Fleischer BR, Leddy D, El-Tamer M: Why should the first be last? "seasonal" variations in the National Board of Medical Examiners (NBME) Subject Examination Program for medical students in surgery. Curr Surg 2003;60(1):69-72.
3. UMMC: 2011-2012 Bulletin. In. Edited by Center UoMM: University of Mississippi Medical Center; 2011.
4. Lind DS, Deladisma AM, Cue JI, Thomas AM, MacFadyen BV, Nesbit RR: Survey of student education in surgery. J Am Coll Surg 2007;204(5):969-74; discussion 975-966.
5. Ripkey DR, Case SM, Swanson DB: Identifying students at risk for poor performance on the USMLE Step 2. Acad Med 1999;74(10 Suppl):S45-48.
6. Pohl CA, Robeson MR, Veloski J: USMLE Step 2 performance and test administration date in the fourth year of medical school. Acad Med 2004;79(10 Suppl):S49-51.
7. McGaghie WC, Cohen ER, Wayne DB: Are United States Medical Licensing Exam Step 1 and 2 Scores Valid Measures for Postgraduate Medical Residency Selection Decisions? Academic Medicine 2011;86(1):48-52.
8. Cho JE, Belmont JM, Cho CT: Correcting the bias of clerkship timing on academic performance. Arch Pediatr Adolesc Med 1998;152(10):1015-18.
9. Hemmer PA, Markert RJ, Wood V: Using in-clerkship tests to identify students with insufficient knowledge and assessing the effect of counseling on final examination performance. Acad Med 1999;74(1):73-75.
10. Pangaro LN, Gibson K, Russell W, Lucas C, Marple R: A prospective, randomized trial of a six-week ambulatory medicine rotation. Academic Medicine 1995;1995(70).

11. Berber E, Brooks C, Erdmann J: Use of USMLE to select residents. Academic Medicine 1993;68:753-59.
12. LCME: Functions and Structure of a Medical School. In., June 2010 edn: Liaison Committee on Medical Education;2010.
13. Reteguiz JA, Crosson J: Clerkship order and performance on family medicine and internal medicine National Board of Medical Examiners Exams. Fam Med 2002;34(8):604-08.
14. Glew RH, Ripkey DR, Swanson DB: Relationship between students' performances on the NBME Comprehensive Basic Science Examination and the USMLE Step 1: a longitudinal investigation at one school. Acad Med 1997;72(12):1097-1102.
15. 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(4):397-404.
16. Blake R, Hosokawa M, Riley S: Student performances on Step 1 and Step 2 of the United States Medical Licensing Examination following implementation of a problem-based learning curriculum. Academic Medicine 2000;75:66-70.
17. Enarson C, Cariaga-Lo L: Influence of curriculum type on student performance in the United States Medical Licensing Examination Step 1 and Step 2 exams: Problem-based learning vs. lecture-based curriculum. Medical Education 2001;2001(35).
18. Butler A: Associate Vice President, National Board of Medical Examiners. In., Personal Communication edn;2002.
19. Hampton HL, Collins BJ, Perry KG, Jr., Meydrech EF, Wiser WL, Morrison JC: Order of rotation in third-year clerkships. Influence on academic performance. J Reprod Med 1996;41(5):337-40.
20. Manley M, Heiss G: Timing bias in the psychiatry subject examination of the National Board of Medical Examiners. Acad Psychiatry 2006;30(2):116-19.
21. Petrusa E, Reilly C, Lee L: Later is better: projected USMLE performance during medical school. Teach Learn Med 1995;7:163-67.
22. Ogunyemi D, Taylor-Harris D: Factors that correlate with the U.S. Medical Licensure Examination Step-2 scores in a diverse medical student population. J Natl Med Assoc 2005;97(9):1258-62.
23. Case SM, Becker DF, Swanson DB: Performances of men and women on NBME Part I and Part II: the more things change. Acad Med 1993;68(10 Suppl):S25-27.
24. Cuddy MM, Swanson DB, Clauser BE: A multilevel analysis of examinee gender and USMLE step 1 performance. Acad Med 2008;83(10 Suppl):S58-62.
25. Weinberg E, Rooney JF: The academic performance of women students in medical school. J Med Educ 1973;48(3):240-47.
26. Cuddy MM, Swanson DB, Clauser BE: A Multilevel analysis of the relationships between examinee gender and United States Medical Licensing Exam (USMLE) step 2 CK content area performance. Acad Med 2007;82(10 Suppl):S89-93.
27. Dawson B, Iwamoto CK, Ross LP, Nungester RJ, Swanson DB, Volle RL: Performance on the National Board of Medical Examiners. Part I Examination by men and women of different race and ethnicity. JAMA 1994;272(9):674-679.
28. Policy and Procedure Manual. Available via http://irb.umc.edu/newdoc/P&P%2005%20v3.1.pdf (Accessed June 2012).
29. Convenience sampling. Available via http://www.experiment-resources.com/convenience-sampling.html (Accessed June 2012).
30. Kleinbaum DG, Kupper LL, Muller KE, Nizam A: Applied Regression Analysis and Other Multivariable Methods, 3rd edn, Duxbury Press, Washington, 1998.
31. Multiple Regression with Categorical Variables. Available via http://www.psychstat.missouristate.edu/multibook/mlt08m.html (Accessed June 2012).
32. Schwab AJ: Hierarchical Multiple Regression. In. University of Texas at Austin, Austin, TX, 2006.
33. Do your data violate multiple linear regression assumptions? Available via http://www.basic.northwestern.edu/statguidefiles/mulreg_ass_viol.html (Accessed date June 2012).
34. Testing the assumptions of linear regression, Available via http://www.duke.edu/~rnau/testing.htm (Accessed June 2012).
35. Schwab AJ: Multiple Regression - Basic Relationships. In. University of Texas at Austin, Austin, TX, 2006.
36. What is dummy coding? Available via http://www.ats.ucla.edu/stat/mult_pkg/faq/general/dummy.htm (Accessed June 2012).
37. Testing for Normality using SPSS. Available via https://statistics.laerd.com/spss-tutorials/testing-for-normality-using-spss-statistics.php (Accessed June 2012).
38. Swanson DB, Holtzman KZ, Butler A, Langer MM, Nelson MV, Chow JW, Fuller R, Patterson JA, Boohan M, The Multi-School Progress Testing C: Collaboration across the pond: the multi-school progress testing project. Medical teacher 2010;32(6):480-85.
39. Schuwirth L, Bosman G, Henning RH, Rinkel R, Wenink AC: Collaboration on progress testing in medical schools in the Netherlands. Med Teach 2010;32(6):476-79.
40. Norman G, Neville A, Blake JM, Mueller B: Assessment steers learning down the right road: impact of progress testing on licensing examination performance. Med Teach 2010;32(6):496-99.
41. Albano MG, Cavallo F, Hoogenboom R, Magni F, Majoor G, Manenti F, Schuwirth L, Stiegler I, van der Vleuten C: An international comparison of knowledge

levels of medical students: the Maastricht Progress Test. Med Educ 1996;30(4):239-45.

42. Muijtjens AM, Schuwirth LW, Cohen-Schotanus J, Thoben AJ, van der Vleuten CP: Benchmarking by cross-institutional comparison of student achievement in a progress test. Med Educ 2008;42(1):82-88.
43. Cronbach LJ, Linn RL, Brennan RL, Haertel EH: Generalizability Analysis for Performance Assessments of Student Achievement or School Effectiveness. Educational and Psychological Measurement 1997;57(3):373-99.
44. Aarts R, Steidel K, Manuel BA, Driessen EW: Progress testing in resource-poor countries: a case from Mozambique. Med Teach 2010;32(6):461-63.
45. Verhoeven BH, Snellen-Balendong HA, Hay IT, Boon JM, van der Linde MJ, Blitz-Lindeque JJ, Hoogenboom RJ, Verwijnen GM, Wijnen WH, Scherpbier AJ et al: The versatility of progress testing assessed in an international context: a start for benchmarking global standardization? Med Teach 2005;27(6):514-20.