2019

Examining the Relationship between Medical Student Satisfaction and Academic Performance in a Pre-Clinical, Flipped-Classroom Curriculum

Matthew J. Mackey
Wright State University - Main Campus, mackey.33@wright.edu

Examining the Relationship between Medical Student Satisfaction and Academic Performance in a Pre-Clinical, Flipped-Classroom Curriculum.

Summary

Medical schools are tasked with the responsibility of facilitating the acquisition of the specific vernacular, skill set, and reasoning capabilities of a physician for a group of individuals with a mélange of diverse experiences. Determining the best methods to train physicians is thus a challenge for those in medical education. Starting with the class of 2021, flipped classroom teaching and learning activities dominate Wright State University Boonshoft School of Medicine's (WSU-BSOM) curriculum for MS1s and MS2s. This new curriculum is under constant evaluation to make sure it is best serving Boonshoft students. However, evaluating student evaluations on course satisfaction is difficult due to the amount of noise in qualitative data. We thus evaluated the quantitative and qualitative data from the surveys completed by the class of 2021 and correlated them with their NBME and final exam results from each module. It was found that a statistically significant relationship exists between student satisfaction and performance on final exams for most modules in this curriculum. Further studies are needed to create a framework for removing noise from these evaluations so those who design the curriculum may strive to improve student satisfaction in ways that will most improve student performance.

Background

The existing literature on flipped classrooms, as it relates to medical education, reports improvements in student satisfaction[1], Student Performance[2],or both[3]. Starting with the class of 2021, Wright State Boonshoft School of Medicine (BSOM) adopted the "Wright Curriculum," a new approach to our curriculum that emphasizes learning in a flipped classroom. Traditional lectures have been replaced with Peer Instruction (PI), Team Based Learning (TBL), and Problem Based Learning (PBL or WrightQ) sessions. PI and TBL sessions require the learner to complete assigned readings and/or other resources before the class session. This allows the class sessions to be active learning sessions and has been associated with better learning outcomes[1].

In PI, students answer multiple choice questions using an electronic clicker. Questions that do not meet the designated threshold for performance are discussed by the class and the class is repolled. The correct answer is revealed, and the instructor then emphasizes the key teaching points of that question. He or she proceeds to answer questions raised by students before the next multiple-choice question appears. TBL consists of an individual quiz, a group quiz, and an application phase. It is meant to challenge students to answer difficult, occasionally ambiguous, questions as a team, as one often does in medical practice. Wright Q sessions are designed around patient cases and help students develop differential diagnoses, order appropriate tests, prescribe treatment, and otherwise diagnose and manage a particular ailment.

Table 1-Wright Curriculum Main Teaching/Learning Activities

Learning Activity	Description
Peer Instruction (PI)	Students are quizzed on their reading using low-stakes, real-time multiple choice questions. Questions where the class collectively scores below a set threshold go to a peer discussion and are then repolled. The instructor then asks a student to explain their answer, the correct answer is revealed, and questions are fielded from the students by the instructor.
Team Based Learning (TBL)	An individual quiz is given at the start of class and is followed by the same quiz given in a group format. Students work together to reach a consensus on the group portion. A more demanding application session follows that is usually case based.
Problem Bases Learning (Wright Q or PBL)	A patient case is presented, and students create a differential diagnosis as a group. The group discusses what exams, labs, diagnostic imaging, treatments, etc. should be performed as the case progresses. The case is opened one week and closed the following week after students have a chance to dive deeper into the various elements of the disease.

Courses for the first two years were structured into nine blocks named Upstream Medicine, Origins1, Origins2, Human Architecture, Host and Defense, Staying Alive, Beginning to End, and Balance Control and Repair. Clinical medicine and research curriculum are provided as longitudinal courses. A description of the major content of each class used in this study can be seen in Table 2 below.

Table 2-Wright Curriculum Blocks Evaluated in this Study

Course	Description
Origins 1	Medical Biochemistry and related disorders; molecular basis for cancer
Origins 2	Basics of pathology, histology, muscle physiology, and pharmacology
Human Architecture	Anatomy of the extremities and thorax
Host and Defense	Immune System, Leukemias/Lymphomas, Infectious Diseases
Staying Alive	Physiology and Pathology of the Heart, Lungs, and Kidneys
Beginning to End	Physiology and Pathology of the GI tract and Endocrine System

As stated, studies have shown increases in satisfaction and/or student success, using similar methods of instruction. However, our search failed to discover a study where the flipped classroom was used and evaluated for the entire curriculum for the first two years of medical education in the United States. After collecting survey data and exam grades for students involved in this curriculum, BSOM has the data do conduct several studies concerning student satisfaction and academic performance. Since we seek to analyze our student feedback with the intention of making changes to elements of the curriculum that result in positive changes in student satisfaction and academic achievement, we found it most relevant to the present to begin analyzing various components of student satisfaction as they relate to student grades on major exams. In this study we seek and answer to the most basic question: Is there a meaningful correlation between student satisfaction with the Wright Curriculum and student performance on major exams?

Methods.

Survey data were collected from the Wright State University Boonshoft School of Medicine (BSOM) class of 2021. This data includes Likert scales and comments boxes for class organization, Quizzes, Final Exams, teaching/learning activities, textbooks, non-textbook materials, and other variables. The Likert Scales were five-point scales with a strong negative, a negative, a neutral, a positive, and a strong positive as options. We assigned numerical values, ranging from one to five, to each response. One corresponds to the strong negative, two with the negative, three with the neutral, four with the positive, and five with the strong positive.

These data were collected at the end of each module, resulting in nine total course evaluations. Data from six courses was evaluated for this study since exam scores were readily available to evaluate a correlation with student academic success. While the entire class of 114 was surveyed, the surveys were not mandatory. The number of responses ranged from 85 to 110 for a given course. Descriptive statistics were generated to include the mean, standard deviation, minimum, and maximum. and to identify any outliers.

The National Board of Medical Examiners (NBME) assessments afford Medical schools the opportunity to provide students with standardized exams or for faculty to create custom exams from a standard bank of questions. These NBME exams are taken by BSOM students four times during the first two years of the curriculum. Every student is required to take these exams to pass a given module, so the response rate is nearly 100% except for the Beginning to End course (B2E) where one student had yet to take the exam. Descriptive statistics were generated to include the mean, standard deviation, etc. and to identify any outliers. MCAT scores were used to control for confounding variables.

Some courses, such as Human Architecture and Host and Defense, did not have an NBME final exam. They instead had a final exam prepared by the faculty of BSOM. However, Host and Defense material was important for portions of the NBME exam that followed Staying Alive. Thus, we decided to include analysis of the relationship between satisfaction with Host and Defense and performance on the Staying Alive NBME.

We first decided on the survey data to include and exclude. Likert scales that were related to class organization, Quizzes, Final Exams, teaching/learning activities, textbooks, and non-textbook materials, were kept. These categories were averaged for student in courses where multiple resources, learning/teaching activities, or assessments were present. This provided us with a continuous variable that we could use to compare courses and run multiple linear regression analysis, controlling for age and MCAT percentile. Student overall satisfaction with the course, our independent variable, was given as an average of the scores for the four mentioned categories. The final exam results that corresponded with the material covered by each course was used to measure academic success, our dependent variable. Linear regression analysis between those variables answered our basic question of whether an association existed between student satisfaction and student academic success.

Bivariate analyses (Pearson correlation coefficients) were conducted to determine whether any one component of student satisfaction had a correlation with exam performance greater than 0.4.

Qualitative data from the comments boxes were selected to correlate with the Likert scales. The categories of organization, assessments, teaching modalities, and textbooks were chosen. Additionally,

comment boxes that discussed overall areas of strength or improvement for a class were included. Such data were coded using a custom code developed for our unique needs. A sample of sixty comments and codes were used to examine interrater reliability and a consensus greater than 85% was achieved.

It was decided that we would only evaluate the coded comments that concern areas of improvement (Question 17 in the survey) for each module since these are likely to provide useful, concise information on how to address areas where students are not satisfied.

The code used is provided in the table below:

Table 3- Code for Q17 "Comment on Areas for Improvement"

Code	Explanation	Example
Organization	The class did not flow appropriately, learning objectives were unclear, material could have been presented better in a different format	"This module did not seem to flow well. I know that it is difficult to present cardiovascular, renal, and pulmonary material in a way that showcases how they are intertwined, but it may have been easier to learn the foundations of each system separately before learning how they act together. "
Address Specific Topic	Student felt that a specific topic was not covered well enough	"Physiology is lacking. I struggled to learn the physiology which made the pathology also even more challenging."
Assessment	Student felt that Final Exam and/or other assessment was unfair or without benefit	"The final should have been more representative of the topics covered in PI. Many students have stated the same feeling about this final saying that they feel like whoever wrote these questions for the final didn't write the MCQ or PI questions"
Teaching Modalities	Student provides suggestion concerning improvements for classroom learning or suggests a change in how often a teaching modality is used	"I think that the application portions of the TBL sessions can be more demanding of us as students. For example, instead of allowing us to make our own adventure when selecting an answer to the "what's on your differential" questions, demand that students find specific pieces of evidence in the patients' history to support or contradict each potential diagnosis on the differential. I think that if you ask us to be more specific and thoughtful in our explanations, we will respond well, and it will create a better learning session for everyone."
Resources	Students major complaint concerns the textbooks used and/or how readings were assigned. Student feels that more resources are needed including podcasts, practice questions, online lectures, reading guides, etc.	"Including a non-mandatory podcast or video lecture would be extremely helpful for big picture ideas, especially if planning to stay in Mark's the next time around. It was difficult to pick out high yield content in the beginning, especially since we were all new to medical school. Thus, more guidance of big picture concepts would have been helpful and may help students in the future."

Comments that contained more than one code were coded multiple times. While several useful analyses could have been conducted, we simply looked to see which codes occurred with the greatest frequency in our data set.

Results

The Boonshoft School of Medicine class of 2021 has a mean age of 24.05 years with a standard deviation of 3.32, a minimum of 22 and a maximum of 48. The class is 54.4% female and 45.6% male. By race, the class is 70.8% White, 8.5% Black, 14.2% Asian/Indian, and 6.6% Bi/Multi/Other. The mean MCAT percentile is 69.50% with a standard deviation of 3.32, a minimum percentile of 16% and a maximum percentile of 97%.

Table 2 - Descriptive Statistics for Final Exams and Satisfaction Surveys

	N	Minimum	Maximum	Mean	Std. Deviation
Final Exam Results					
Origins1 Origins2 NBME	114	60.00%	95.00%	79.72%	7.95%
Human Architecture Lab Practical (Final Exam)	114	68.00%	100.00%	82.11%	8.13%
Host and Defense MCQ (Final Exam)	114	56.00%	98.00%	81.61%	10.98%
Staying Alive NBME	114	55.00%	92.00%	77.63%	8.39%
Beginning to End NBME	113	60.00%	96.00%	83.20%	7.40%
Satisfaction Results					
Origins1 Averaged	106	2.63	4.88	3.87	0.47
Origins2 Averaged	110	2.18	4.73	3.55	0.54
Human Architecture Averaged	106	2	4.82	3.65	0.54
Host and Defense Averaged	85	2.92	5	4.15	0.43
Staying Alive Averaged	96	1.69	4.83	3.33	0.59
Beginning to End Averaged	88	2.22	5	3.66	0.52

Except for Beginning to End, every class has a significant correlation between average student satisfaction and its respective final exam. In some instances, such as for Human Architecture or Host and Defense, these correlations were significant even when MCAT percentiles were not. Student age was not associated with any of the final exams.

Notably, no significant correlation was found between Host and Defense Satisfaction and the result of the Staying Alive NBME exam. This may reflect the small extent of Host and Defense material covered on this NBME and/or be a result of the several months that pass between the Host and Defense final and the Staying Alive NBME.

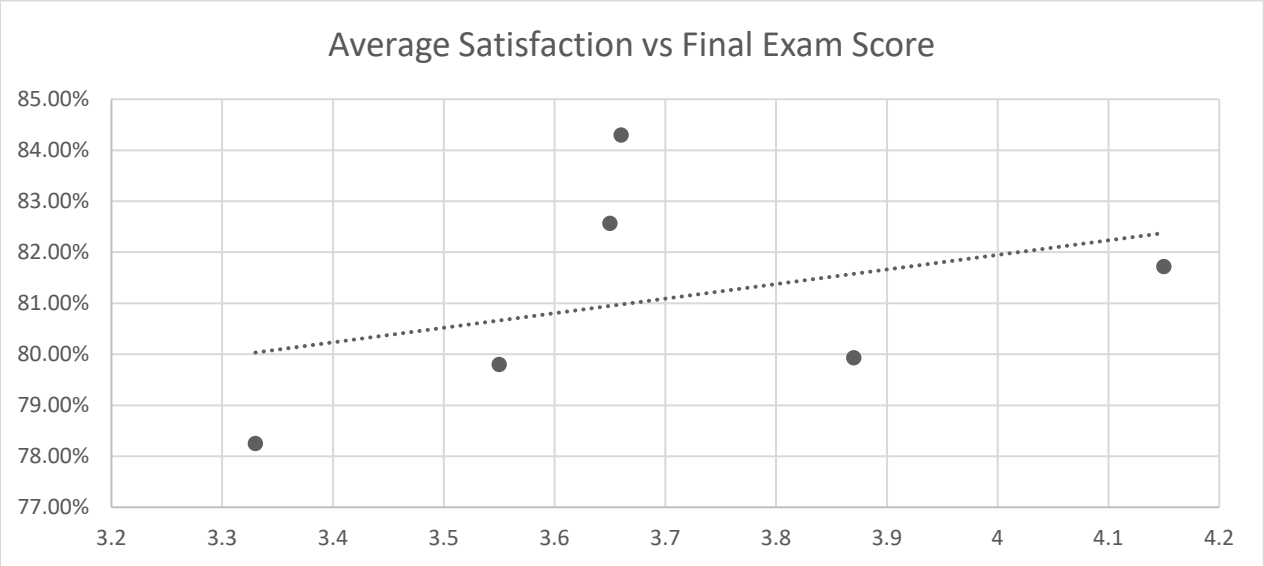
With respect to the beta values, it was found that the change of one point of student average course satisfaction could be associated with anywhere from a 1.91 to an 8.36 difference in the percentage grade of their final exam for that course.

Table 3 - Regression Coefficients for Associations between Course Average Satisfaction and Final Exam Results, Controlling for MCAT Percentiles

Course Average Satisfaction vs Mean Final Exam Grade	N	Independent Variable	Unstandardized Coefficients		Standardized Coefficients		95.0% Confidence Interval for B	
			B	Std. Error	Beta	Sig.	Lower Bound	Upper Bound
Origins 1 vs O1/O2 NBME	106	MCAT Percentiles	0.24	0.04	0.51	*Less than 0.01	0.16	0.31
		Origins 1 Satisfaction Averaged	6.38	1.35	0.37	*Less than 0.01	3.69	9.06
Origins 2 vs O1/O2 NBME	110	MCAT Percentiles	0.23	0.04	0.49	*Less than 0.01	0.15	0.30
		Origins 2 Satisfaction Averaged	2.88	1.26	0.19	0.03	0.37	5.38
Human Architecture (HA) vs Practical Exam	106	MCAT Percentiles	0.10	0.45	0.22	0.02	0.02	0.19
		HA Satisfaction Averaged	4.44	1.39	0.30	*Less than 0.01	1.68	7.20
Host and Defense vs MCQ	85	MCAT Percentiles	0.14	0.08	0.20	0.07	-0.12	0.30
		HD Satisfaction Averaged	8.36	2.83	0.31	*Less than 0.01	2.73	14.00
Host and Defense vs NBME*	85	MCAT Percentiles	0.21	0.06	0.40	*Less than 0.01	0.10	0.33
		HD Satisfaction Averaged	1.78	2.10	0.09	0.40	-2.39	5.95
Staying Alive (SA) vs NBME	96	MCAT Percentiles	0.19	0.05	0.37	*Less than 0.01	0.09	0.28
		SA Satisfaction Averaged	4.13	1.26	0.30	*Less than 0.01	1.62	6.63
Beginning to End (B2E) vs NBME	88	MCAT Percentiles	0.19	0.04	0.45	*Less than 0.01	0.10	0.28
		B2E Satisfaction Averaged	1.91	1.26	0.15	0.13	-0.59	4.41

Overall, student satisfaction is significantly correlated with exam outcomes, sometimes even more so than MCAT percentiles.

Figure 1- Plot of Average Satisfaction vs Final Exam Score



Bivariate Analysis for Staying Alive did not show any correlation coefficients above 0.4. We thus did not deem any particular element of student satisfaction as a significant driver of exam results when considered on its own.

Table 4- Correlations between Staying Alive Satisfaction Category Averages and NBME Results

Correlations		SA_NBME	SA Organization	SA MCQs	SA Final Exam	SA Learning Activities	SA Textbooks	SA Non Textbook Materials
SA_NBME	Pearson Correlation	1	.211*	.273**	.246*	.257*	.234*	0.005
	Sig. (2-tailed)		0.039	0.007	0.016	0.012	0.022	0.959
	N	114	96	96	96	96	95	94
SA Organization	Pearson Correlation	.211*	1	.568**	.429**	.618**	0.176	0.096
	Sig. (2-tailed)	0.039		0	0	0	0.088	0.355
	N	96	96	96	96	96	95	94
SA MCQs	Pearson Correlation	.273**	.568**	1	.584**	.698**	.340**	.282**
	Sig. (2-tailed)	0.007	0		0	0	0.001	0.006
	N	96	96	96	96	96	95	94
SA Final Exam	Pearson Correlation	.246*	.429**	.584**	1	.513**	0.149	.213*
	Sig. (2-tailed)	0.016	0	0		0	0.149	0.039
	N	96	96	96	96	96	95	94
SA Learning Activities	Pearson Correlation	.257*	.618**	.698**	.513**	1	.374**	.291**
	Sig. (2-tailed)	0.012	0	0	0		0	0.004
	N	96	96	96	96	96	95	94
SA Textbooks	Pearson Correlation	.234*	0.176	.340**	0.149	.374**	1	.371**
	Sig. (2-tailed)	0.022	0.088	0.001	0.149	0		0
	N	95	95	95	95	95	95	93
SA Non Textbook Materials	Pearson Correlation	0.005	0.096	.282**	.213*	.291**	.371**	1
	Sig. (2-tailed)	0.959	0.355	0.006	0.039	0.004	0	
	N	94	94	94	94	94	93	94

Analysis of the coded comments for Question 17 showed that organization was the most common area for improvement cited by students (37%). Resources (22%), Teaching Modalities (18%), Address Specific

Topic (14%), and Assessment (9%) made up the remainder of the areas where students expressed a desire for improvement.

Figure 2-Occurrence of Codes for Q17 by Type as a Percentage of Total Coded Comments for Q17

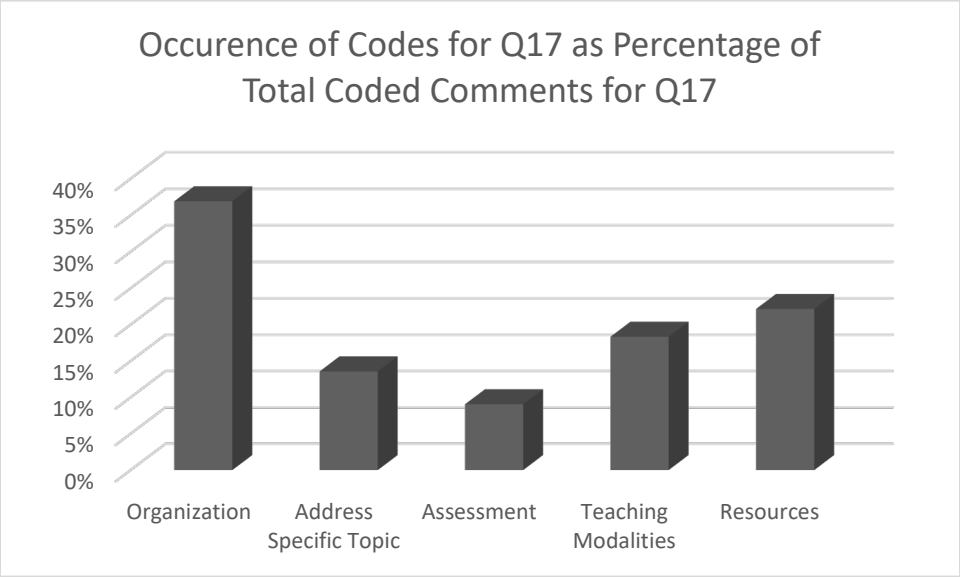


Figure 3-Student Quote, coded for "Resources" and "Teaching Modalities"

"Areas of improvement could be the addition of reading guides to this module so we know where to focus on the details and when to understand the general picture more. Also, I think it would be beneficial to have 3 PIs and 1 TBL a week rather than 2 and 2. This would allow for a little less stress when it comes to reading and studying the material before coming to class, since sometimes it is impossible to do more than read in the time allowed. I feel that PIs serve as a learning environment much more than TBLs do."

Discussion

While it is important that we were able to establish a significant correlation between student satisfaction and academic performance, such a relationship does not begin to imply causation. Nor does our study, at this level, address adequately the specific aspects of student satisfaction that may correlate most closely with improving student exam scores. To achieve a greater degree of insight, a few considerations are provided here for future analysis. First, comment codes can be correlated with student satisfaction and with final exam results. Second, codes can be generated in a tiered structure to provide a more granular view of student satisfaction vs academic achievement. This would be useful in determining which element of a given class seems to be troublesome. For instance, a textbook could be evaluated by this method to determine if satisfaction or dissatisfaction concerning its readership correlated to final exam results. Third, curriculum changes should be accounted for and evaluated by these criteria to look for changes in student satisfaction and student performance. Positive changes in student satisfaction would signify positive results, so long as there was no decrease in academic performance. Specifically, this methodology can be applied to the class of 2022 at the end of this academic year. At that time, STEP scores for both classes can be evaluated as well. It would be useful to compare the two classes even though the modules even though, some course materials, instructors, and

differences in the time allocated to learn a given topic will be different. The effects of those differences on student satisfaction and achievement would be well evaluated in this context.

Lastly, it may be of value to correlate codes and Likert scales with measures of student mental health. The overall prevalence of depression among medical students was found to be 28% in a 2016 meta-analysis of 77 studies [4]. Depression is one of the symptoms of burnout, a frequently described syndrome that results from work-related stress that may contribute to suicidal ideation, substance abuse, and other negative outcomes[5]. The depression and burnout associated with medical students is multifactorial, but academic and time pressures both contribute to the stress this population experiences that leads to burnout[6]. When considering the prevalence of burnout and its association with academic pressures, it is sensible that the LCME requires medical schools to have an "effective system of personal counseling for medical students" and that it is required that medical schools have a formal process to collect and consider medical student evaluations as part of their evaluation of program quality[7].

When a student, such as the student in figure 3, mentions that they feel that there is inadequate time to do more than complete assigned readings and mentions associated stress, it may be worth keeping a closer eye on that student due to the mentioned link between academic and time pressures and burnout. If that student's peers are also reporting similar levels of stress, then student mental health, in addition to academic performance, may be another important reason to reform an aspect of a course. The expectation is that any changes made have a benign or positive effect on academic performance while improving student satisfaction, and possibly, the wellbeing of the student body.

Conclusion

The relationship between student satisfaction with the Wright Curriculum and student exam performance displays a statistically significant positive correlation overall for the class of 2021. Since this class is the first class to have a curriculum designed almost entirely around flipped classroom learning activities and their satisfaction with the curriculum is associated with their success, their evaluations on courses should continue to be considered carefully. Further studies are needed to provide a meaningful framework to properly consider student evaluations to filter out noise and determine the actual strengths and weaknesses of our curriculum.

References

1. Morgan H, Mclean K, Chapman C, Fitzgerald J, Yousuf A, Hammoud M. Practical Teaching The fl ipped classroom for medical students. *Teacher*. 2015;12:155-160. www.theclinicalteacher.com.

2. Tune JD, Sturek M, Basile DP. Flipped classroom model improves graduate student performance in cardiovascular, respiratory, and renal physiology. *Adv Physiol Educ*. 2013;37(4):316-320. doi:10.1152/advan.00091.2013

3. Missildine K, Fountain R, Summers L, Gosselin K. Flipping the Classroom to Improve Student Performance and Satisfaction. *J Nurs Educ*. 2013;52(10):597-599. doi:10.3928/01484834-20130919-03

4. Puthran R, Zhang MWB, Tam WW, Ho RC. Prevalence of depression amongst medical students: A meta-analysis. *Med Educ*. 2016;50(4):456-468. doi:10.1111/medu.12962

5. Dyrbye LN, West CP, Satele D, et al. Burnout among u.s. medical students, residents, and early career physicians relative to the general u.s. population. *Acad Med*. 2014;89(3):443-451. doi:10.1097/ACM.0000000000000134

6. Johnson JC, Degenhardt BF, Smith CK, Wolf TM, Peterson DF. Tool for Predicting Medical Student Burnout From Sustained Stress Levels: Factor Analysis of the Medical Education Hassles Scale-R. *J Am Osteopath Assoc*. 2018;118(3):170. doi:10.7556/jaoa.2018.036

7. Education M, Leading P. Functions and Structure of a Medical School: Standards for Accreditation of Medical Education Programs Leading to the MD Degree. Liaison Committee on Medical Education. http://lcme.org/publications/. Published 2018. Accessed June 27, 2019.