IDENTIFYING PREDICTORS OF PERFORMANCE ON USMLE® STEP 1

by

SACHIN SHAH

DISSERTATION

Presented to the Faculty of the Medical School
The University of Texas Southwestern Medical Center
In Partial Fulfillment of the Requirements
For the Degree of

DOCTOR OF MEDICINE WITH DISTINCTION IN MEDICAL EDUCATION

The University of Texas Southwestern Medical Center
Dallas, TX

© Copyright by Sachin Shah 2017
All Rights Reserved

ACKNOWLEDGMENTS

This research was supported by the UT Southwestern Office of Medical Education, Office of Student Affairs, and Student Academic Support Services. I thank my mentors, Dr. Dorothy Sendelbach, Dr. Angela Mihalic, and Dr. Arlene Sachs, for their unwavering support and guidance throughout this research project. I thank Shola Rogers for her efforts compiling the database, and I thank Ms. Carol Wortham and Ms. Anne McLane for their parts in compiling data. I also thank Dr. John Sadler for his guidance in obtaining IRB exemption, and Dr. Joan Reisch for her early input on proposed statistical methods.

TABLE OF CONTENTS

INTRODUCTION 2
BACKGROUND 3
METHODS 7
RESULTS 10
DISCUSSION 19
CONCLUSION 25
LIST OF TABLES 28
LIST OF FIGURES 33
REFERENCES 46
VITAE 48
APPENDIX 49

INTRODUCTION

USMLE® Step 1 is considered by residency program directors to be one of the most important factors in selecting medical students for interviews, so it is important for both students and medical schools to maximize scores[1]. However, despite its importance, very little has been studied to determine indicators of performance on the exam. It was the goal of this project to look at undergraduate GPA and MCAT® scores, demographic information, medical school exam performance, and self-reported survey data from the UT Southwestern Medical School Class of 2018 to discover correlations that may allow medical schools to better predict student performance on the exam and guide student preparation for the exam. Importantly, this information would allow administrators to identify future students who are at risk of failing or doing poorly on the exam and intervene early in order to prevent poor performance on the exam. Furthermore, if students are able to better understand how to best prepare for the exam and to be able to identify personal red flags, they may be more likely to independently adjust studying habits early on in order to optimize their performance on the exam and ultimately be more competitive for residency. It is for these reasons that we believe it is important to review and identify predictors of USMLE Step 1 performance. This project has three main goals: 1) determining if data available prior to admission can predict Step 1 performance, 2) analyzing school performance and survey data to suggest studying tips, and 3) developing a model to flag students at risk of performing poorly on the exam for early intervention.

BACKGROUND

USMLE Step 1 is the first of three physician licensure exams taken by medical students and covers the basic science material taught during the pre-clinical years of medical education. It was designed initially to separate those with adequate versus inadequate knowledge as an indicator for preparedness for clinical training, but it has become a tool used heavily to stratify applicants[2]. A student's score on this exam is considered by residency program directors to be the top factor in selecting students for interviews[3]. On the USMLE® Step 1 exam, it is generally agreed that a score below 220 will significantly decrease a student's chance of getting into a top residency program, as this is often a minimum score required to even apply to many programs, even in less competitive specialties[4]. As such, it is important for both students and for medical schools to do what they can to maximize their scores to optimize their chances of being considered for top residency programs.

Starting with the admissions process, medical schools use MCAT® scores and undergraduate GPAs as their top factors in selecting students to interview, with the assumption that high MCAT® scores and GPAs will predict success in medical school, and specifically success on the USMLE® Step 1 exam. A number of studies indicate that these pre-admissions data may be mild predictors of Step 1 scores but are wary of the weight that is applied to these factors in the admissions process because of high variability in predictive value[5]. Since these studies all include only matriculating medical students, it is possible that this factor exhibits a threshold

effect, where after a certain point the predictive value of these data points becomes less significant. Further large-scale studies are needed to be able to make a more accurate assessment.

The success of a pre-clinical medical school curriculum is also judged largely by the performance of students on Step 1, and medical school programs assume that high performance on internal exams indicate that students are prepared for Step 1. Students at many campuses, including UT Southwestern, are offered dedicated studying time for Step 1, and there are dozens of resources available to prepare for the exam, mostly chosen by word-of-mouth recommendations. All of these factors play into one's performance on Step 1, and yet very little has thus far been studied. Only one study at Albert Einstein looked at specific resources and found that question banks are more predictive than other resources, and that the amount of time spent studying has no correlation with scores[6].

Because of the significant consequences of performance on Step 1, studies show that this exam is highly related to medical student burnout. Much of this stress arises from programs teaching information that doesn't overlap directly with Step 1, with 35% of students reporting that they shift their study focus away from their schools' curricula[7]. It would therefore benefit programs to better understand how to guide students in preparing for this exam to not only improve scores on Step 1, but also to improve focus on the school's unique curriculum. It would also be useful for programs to be able to accurately identify students who are at risk of poor performance so that they might be able to provide early intervention. Few studies have looked

into developing a mathematical model, and even fewer have had sufficient predictive success, with one notable exception at Marshall University in December 2016[8]. By expanding on those studies and looking at a wider variety of data, it may be possible to guide other schools on developing individualized models that effectively flag at-risk students.

Due to the importance of the USMLE Step 1 exam, the project aims include:

1. To identify early predictors of performance on USMLE Step 1. First, analyzing pre-matriculation data of undergraduate GPA and MCAT® scores and demographic factors may indicate good predictors of Step 1 scores. This information would potentially allow admissions offices to better determine what weight should be given to these factors in determining which students are offered admission.
2. To evaluate if success in pre-clinical courses at UT Southwestern correlate to success on Step 1. Evaluating the relation of overall GPA to exam performance and determining if specific course or up-trending or down trending scores throughout the year may serve as predictors for exam scores.
3. To determine correlations between specific resources or studying strategies and Step 1 scores to better inform students and the medical school on optimal preparation for the exam. Students at UT Southwestern have completed two sets of surveys – one during the dedicated study period to determine interval scores on practice tests as well as progress through studying materials, and the second after the exam to retrospectively determine how much time students spent studying and which resources they used. . Analysis of these surveys will be performed to identify

factors that are associated with higher scores and may provide guidance to future at risk students; preventing failures or poor scores on USMLE Step 1.

4. To develop a mathematical model that predicts performance on USMLE Step 1 in order for administrators to be able to identify at-risk students and provide early intervention.

Using prior studies as a guide, we first hypothesize that pre-admissions data will provide only mild predictive value of Step 1 scores and should be used cautiously as admissions criteria and may exhibit a threshold effect. Second, performance in medical school is directly correlated with Step 1 scores, but decreasing grades as students get closer to the exam date will not necessarily predict performance on Step 1. Third, students using more active resources and start studying early will perform better on the exam, but will see diminishing returns after each week of study with scores maximizing by weeks 5 and 6. Finally, that a predictive model can be developed to categorize students as low-risk or high-risk.

METHODS

Data Collection and Subject Protection

Data was collected for the UT Southwestern Medical School Class of 2018 (n=238). An independent data broker was used to compile a database of student undergraduate GPA, MCAT® scores, medical school test scores, Comprehensive Basic Science Self-Assessment (CBSSA) test scores, and demographic information. Demographic information included age, sex, race, year of undergraduate graduation, prior degrees, concurrent degrees, leaves of absence, and socioeconomic category on admission.

The data broker also obtained data from 2 surveys that were already being administered by the UT Southwestern Student Academic Support Services and the Office of Student Affairs via SurveyMonkey (***APPENDIX A*** *and* ***B***). The first was a weekly survey during the dedicated 6-week Step 1 preparation period to determine student's scores on their practice tests and students' progress in their study materials. The second survey was a post-exam survey which asked students when they started studying, how many weeks they studied, how many days per week and hours per day they studied, which resources they used, how many practice tests they took, and how involved they were with extracurricular activities. These surveys had student identifiers, which the data broker was able to de-identify using the same codes from the database described in the prior paragraph in order to link the two databases. No students were excluded from the study, but only students who took Step 1 in Summer 2016 were surveyed. Students were provided no incentives for survey completion.

The collective data was assembled by the research team in an MS Excel spreadsheet along with the student's USMLE Step 1 scores. All students were assigned a participant number and their identities were protected by the 'data broker' personnel who maintained the confidential record of links between the unique participant data and project ID number. The investigators and other listed team members did not have access to student identity or demographic information. This project was deemed exempt by the IRB because student identities were protected, and use of the data did not require consent per FERPA 34 CFR § 99.31[9].

Pre-Admissions Data Analysis

The first step of analysis entailed identifying correlations between Step 1 scores and MCAT®, undergraduate GPA, and demographic data. During this part of the analysis, MS Excel was used to perform simple linear regression to identify if correlations existed, with $r>0.5$ suggesting strong correlation, $0.3<r<0.5$ suggesting moderate correlations, and $0.1<r<0.3$ suggesting weak correlation, per human behavior study standards[10]. This corresponded to our use of $R^2>0.25$ indicating strong correlation and $0.10<R^2<0.25$ indicating moderate correlation. P-values of <0.05 were considered significant.

Medical School Grades and CBSSA Scores Data Analysis

As above, MS Excel was again used to perform simple linear regression to identify if correlations existed between medical school grades or practice CBSSA scores and Step 1 scores. Independent groups t-tests were performed with StatPlus in order to better understand differences between those scoring above and below 220.

Survey Data Analysis

For the weekly progress surveys, simple linear regression was used in MS Excel to determine if there were significant correlations between progress made in First Aid, progress made in the Q-bank, and cumulative percent correct in the Q-bank and practice exam scores for that week. Average practice scores and the average delta of practice scores were also analyzed via simple plotting per week in order to determine how scores changed over the course of the dedicated study period.

Similar analysis was performed on the post-exam survey. To analyze medical school grades and survey information, linear regression in MS Excel was used to identify if correlations exist between Step 1 scores and total hours studied, number of weeks studied, hours of extracurricular involvement, Q-bank % completion, and number of practice tests taken. Independent groups t-tests were also performed to compare users of a resource to non-users of a resource for each of the commonly used test-prep resources in order to identify resources that result in higher Step 1 scores.

Developing a Model

The primary goal of the model is to predict if a student is at risk of scoring under 220 on Step 1. Therefore, independent groups t-tests were performed in StatPlus in order to identify factors that were different between those who scored below 220 and those who scored 220 or higher, with a p-value of <0.05 considered significant. A combination of the findings from the linear regression models in prior analysis and these t-tests were used to test the more highly correlated factors via stepwise multiple regression analysis in StatPlus in order to create a predictive model to identify students who are likely to score below a 220.

RESULTS

Pre-Admissions Data

The distribution of pre-admissions scores for all 228 students in the UT Southwestern Class of 2018 can be reviewed in ***Table 1***. Each section of the MCAT®, as well as the combined biological and physical sciences score and the cumulative score, were correlated with Step 1 scores and medical school grades *(**Figures 1*** and ***2**)*. The R^2 values can be seen in ***Table 2***. Of the individual sections, the biological sciences and physical sciences had R^2 values of 0.067 and 0.095, respectively, indicating only a weak correlation with Step 1 scores; the verbal reasoning section had an R^2 value of 0.004, indicating no correlation. The summation of the biological and physical sciences sections resulted in an R^2 of 0.127, indicating a moderate correlation between the science sections of the MCAT® and Step 1 scores.

When correlated with medical school performance (cumulative second year grade percentage), the correlations with MCAT® scores were even weaker (***Figures 3*** and ***4***). Even the strongest predictor, the summation of the biological and physical sciences sections, had an R^2 value of only 0.072, indicating only very weak correlation between medical school performance and MCAT® scores.

Undergraduate GPA exhibited similar trends, but were even less predictive than MCAT® scores (***Figures 5*** and ***6**)*. An R^2 value of 0.104 indicates moderate correlation with Step 1 scores, and an R^2 of 0.058 shows only minimal correlation with medical school grades.

An analysis of demographic factors indicates that race, undergraduate major (science vs non-science), and gender are not significantly different, as the step 1 score averages of each group are within the margin of error ($p>0.05$ for each race, male versus female, and science vs non-science) (***Figure 7***). Although at first glance it seems there was no significant difference among age groups, after isolating only those under and above 30, there was a significant difference in means of almost 8 points (237.7 vs 229.7, $p<0.03$) between those under 30 and those 30 and older. Step score differences among the socioeconomic groups was also not significant (see ***APPENDIX C*** for full details on how students were sorted into each category).

Medical School Grades and CBSSA Scores

The CBSSA exam was administered to each student towards the end of the pre-clinical curriculum, and prior to the dedicated time for Step 1 study. Two hundred eleven of the students in the class took the exam during this time period. Of those, the average scaled score on the exam was 193 (SD ±28). The average change between the CBSSA scaled score and the actual Step Score was 44 points (SD ±18), with a range from -6 to +85. A graph of Step 1 scores versus CBSSA scores shows an R^2 of 0.599, indicating a very high correlation (***Figure 8***). Of all the variables studied over the course of this study, CBSSA scaled score had the highest predictive power of final Step 1 scores.

The second most significant predictive factor was medical school performance, as indicated by the second-year medical school cumulative grade percentage (***Figure 9***). 217 students had completed all courses and taken the Step 1 exam, with an average grade of 84.5%, SD ±5.0, ranging from 69.3% to 98.0%. When second year grades are broken down by quarter, R^2 values indicate that each quarter is more and more predictive of the final

Step 1 scores, with the exception of the final quarter, which is far less predictive than the other three, but still showing a strong correlation with Step 1 scores (***Table 3***).

There was very little change in scores between each quarter grading term, both for individuals and on average (***Table 4***). However, when plotting just the individual change between the third and fourth quarter grades compared to Step 1 scores, individuals whose scores decrease the most between the third and fourth quarters end up with higher Step 1 scores (***Figure 10***). An R^2 of 0.137 on this plot indicates that there is a moderate negative correlation between those whose grades drop between Q3 and Q4 and their Step 1 scores. This same relationship is not seen in the change between the second and third quarters.

Survey Data

SASS Weekly Survey

Weekly surveys were distributed during the dedicated Step 1 preparation period and administered by the Student Academic Support Services. These surveys had 198 unique responders over a 6 week period, with 174 responders in week 1, tapering to 43 responders in week 7. Because identifiers were used, we had the ability to track the same individual student responses over the 7-week period that the surveys were administered.

The most predictive data point on these surveys was the cumulative % correct in the UWorld™ and/or USMLE Rx™ question bank, the two question bank resources used by every student surveyed. A plot of the % correct versus the practice exam score for that week shows an R^2 of 0.507, indicating a strong correlation (***Figure 11***). Less predictive and moderate predictors were the progress made through the Q-bank(s) and First Aid™, respectively, with R^2 values of 0.241 and 0.200 (***Figures 12*** and ***13***). The percentages

include multiple passes but don't distinguish between different resources, so someone who was halfway through a second pass of the U-World™ Q-bank would have 150%, as would someone who completed 100% of the U-World™ Q-bank and 150% of USMLE Rx.

Each consecutive week of study produced smaller gains in points on the USMLE 1. ***Figure 14*** shows that in general, scores continue to increase throughout the first 6 weeks of study, but less and less each week, with a significant drop by Week 7. It is important to note that students were only given 6 weeks of dedicated preparation time, so those who spent longer than 6 weeks had to request extra time. For the first 6 weeks, a quadratic line of best fit was determined to be:

Step Score = -0.4365 x (# weeks)2 + 8.5953 (# weeks) + 201.16

with an R^2 of 0.972, indicating a very strong fit to the data. This equation's vertex is 243.56 at 9.87 weeks, indicating that between weeks 6 and 10, a student may potentially gain 6 points, or less than 2 points per week of study.

Figure 15 includes only data points for students with test scores in two consecutive weeks, and shows that while students could gain an average of 13.9 points in the first week (SD 15.1), the gains drop progressively week-to-week, with an average increase of just 4.1 points between weeks 5 and 6 (SD 12.0), and many students not increasing or even dropping scores at that point. The difference between weeks 6 and 7 was not plotted because there were only 5 students who submitted both week 6 and week 7 practice exam scores. A linear best-fit line on this graph indicates that after week 6, on average no further score increases should be expected, with an R^2 value of 0.938 indicating the strong predictive strength of this model.

Post-Exam Survey

The post-exam survey administered within a few weeks of the Step exam had 140 responses out of 228 total students in the class. The distribution of scores of those who completed the survey compared to the entire class is shown in ***Table 5***. Those who completed the survey scored, on average, 5 points higher than the class as a whole, with a lower standard deviation among survey respondents indicating that a significant proportion of those who didn't complete the survey performed worse on the exam.

A comparison of users versus non-users of each resource indicates that those who use Kaplan™, NBME® practice exams, UWorld™ Self Assessments, BRS™ Physiology, and USMLE Rx™ scored higher (***Figure 16***). Those who used Doctors in Training™ and Sketchy Pharm™ scored lower. In this study, only Sketchy Pharm™ had $p<0.05$, so the differences seen in the graph are not statistically significant.

The number of questions answered in either the UWorld™, USMLE Rx™, or Kaplan™ question bank had a mild to moderate correlation with an R^2 of 0.094. However, considering that UWorld™ is ~2350 questions, grouping the data into those answering fewer than 50% of the Q-bank questions, those answering 50-150%, and those answering greater than 150%, the significance of difference becomes more apparent, especially at the higher end (***Figure 17***). Students who answered fewer than 1700 questions scored an average of 230 ± 10.6 (95% CI), those who answered 1701-4000 questions scored an average of 241 ± 2.4 (95% CI), and those who answered greater than 4000 questions scored an average of 254 ± 5.3 (95% CI).

Students who started studying for Step 1 earlier performed better, although the data doesn't suggest a clear ideal start time, since there is no significant difference between any

of the adjacent categories beginning 1 year, 6 months, 3 months, or 6 weeks prior to the exam (***Figure 18***). Those who started to study 1 year prior to the exam scored on average 254 ± 8.5 (95% CI), those who started 6 months prior scored 252 ± 6.9 (95% CI), those who started 3 months prior scored 242 ± 3.2 (95% CI), and those who waited until the dedicated 6-week preparation time scored 237 ± 3.6 (95% CI).

When asked how many hours students studied within the 6-week preparation time, there were no significant differences found in terms of number of weeks spent studying or total hours studying over the course of the 6 week period (***Figures 19*** and ***20***). Statistical error accounted for differences in every category.

Somewhat surprisingly, extracurricular involvement appears to have no impact on Step 1 scores (***Figure 21***). In fact, those who started to study 1 year and 6 months prior to the exam had higher levels of extracurricular involvement at 7.7 hours and 8.8 hours, respectively, compared to those who started studying later at 6.0 and 6.3 hours. There was also no significant difference in the extracurricular involvement and total hours studied during the dedicated study period.

Students were asked to name the last practice exam they took and the score they received prior to taking the Step 1 exam. ***Figure 22*** shows that NBME® 18 is an accurate and precise predictor of one's final score on Step 1, with an average difference of 0.06 points ±2.0. UWorld™ Self Assessment 2 (+2 points ±5.0, 95% CI) and NBME® 15 (-2.5 points, ±7.2) are accurate but less precise. NBME® 13 (+5.5 points, ±4.8, 95% CI) and NBME® 17 (-4.5 points, ±3.4, 95% CI) are less accurate predictors, and also not as precise. NBME® 14 (+1 point, ±13, 95% CI) is imprecise. NBME® 16 (+3 points, ±2.9, 95% CI) is less precise and accurate than NBME® 18, but is otherwise a good predictor.

Developing a Model to Predict Step 1 Performance

To develop a model to identify those at risk of scoring below 220 on Step 1, t-tests were performed for a number of categories that suggested correlation in the pre-admissions data the surveys, and the medical school grades and CBSSA scores for those students scoring below 220 compared to those scoring 220 or higher. ***Table 6*** shows the averages, standard deviation, and p-values obtained from the t-tests for each tested variable.

The t-tests indicated that 6 categories had significant differences between the groups: undergraduate GPA, MCAT®®® score, # of questions answered in the Q-Bank, total hours studying, CBSSA score, and cumulative MS2 grades.

Based on the averages and standard deviations shown, a rounded threshold roughly 1-2 standard deviations below the average for the <220 group was chosen and a positive predictive value for selecting a student who would score <220 for each of the tests was calculated (***Table 7***). Most notably, a student with a CBSSA score of less than 160 has a 64% chance of scoring below 220, and a student in the bottom 15% of the class has a 60% chance of scoring below 220. The next most predictive factors were the undergraduate GPA and MCAT® scores, with an undergraduate GPA less than 3.5 predicting a 45% chance of scoring below 220, and an MCAT® less than 30 with a 48% chance. The least predictive were answering fewer than 2000 Q-bank questions or spending less than 250 total hours studying, with a positive predictive value of 27% and 30%, respectively.

Using these findings, a series of multiple linear regressions were tested, using a mixture of continuous variables and categorical variables as seen on ***Table 7***. Since the purpose of the model was to flag students at risk of performing poorly early, the number of

Q-bank questions answered and total hours studying were left out of the model, as those are factors that wouldn't be known until the end of the study period. Since there was a significant difference in MCAT® scores and undergraduate GPAs using a threshold, those factors were left as constants either added or left out of the model to increase the predictive value. Finally, the initial CBSSA and cumulative second year grade percentage had such strong correlations that those were included with linear coefficients. The result of the multiple linear regression was the following model:

USMLE1 Score = 100.84962 + 4.85204 (*if Undergraduate GPA <3.5*) + 3.66261 (*if MCAT® <31*) + 0.31215 * [CBSSA Score] + 0.8223 * [Cumulative MS2 Grade %]

The regression statistics are shown in ***Table 8***. Undergraduate GPA and MCAT® scores did not have significant p-values when run through the model, but they improved the overall predictive value of the model, so they were helpful to include. An R^2 value of 0.64 indicates that this model accounts for 64% of the variability in the population, and a standard error of 8.87 indicates that this model will predict 67% of scores within 8.87 points, and 95% of scores within 17.74 points.

When testing this model to identify those at risk of scoring below 220, the standard error was used as the upper end of the model. That is, any student predicted by the model to score 228.87 or lower was considered a positive test. When applied to the group as a whole (186 students had all data points necessary to run this model), this resulted in a sensitivity of 81% (25/31), specificity of 86% (134/155), positive predictive value of 54% (25/46), and negative predictive value of 96% (134/140). That is, even though the model only predicts 67% of scores within the standard error, using this to flag students could

catch 81% of those who ultimately scored below 220. If a student was found to score higher than a 228.87 by the model, then there is a 96% chance that the student will not score below 220. Of note, 38% of the false positives (8/21) scored 225 or below.

DISCUSSION

Pre-Admissions Data

Although significant weight is traditionally placed on MCAT® and undergraduate GPAs, clearly only a mild-to-moderate correlation exists between Step 1 scores and MCAT® scores, only weak correlations between Step scores and undergraduate GPAs and even weaker correlations between MCAT® or GPAs to medical school performance. The verbal section of the MCAT® is particularly poor at predicting any level of performance in medical school. However, these two factors are still the best measures available to easily sort through applicants. In fact, after putting together a model to predict performance, it is clear that those with MCAT® scores 30 or below, or GPAs lower than 3.5 are at a significant risk of performing poorly on Step 1. This data may therefore be skewed by the fact that UT Southwestern students have generally very high MCAT® scores and undergraduate GPAs. This indicates a threshold effect – if a student scores above a 30 on the MCAT® and has greater than a 3.5 GPA, perhaps no additional consideration should be made with regards to those scores. However, this data may suggest that admissions personnel be wary of students who do not meet those thresholds.

Race, undergraduate major, socioeconomic class, age, and gender have no significant effect on Step 1 scores, with any differences of averages well within the margin of error. Students certainly report benefiting from a diverse class in the annual graduation questionnaire, and ample research suggests the many benefits of diverse exposures and experiences[11]. Admissions committees may consider using these factors to help create a

diverse class to improve the learning atmosphere as a whole, but demographic factors have no bearing on how a student is likely to perform on Step 1.

Medical School Grades and CBSSA Scores

It is clear that medical school grades and CBSSA scores are the strongest predictors of Step 1 performance, both showing very high correlations. Although many students have a tendency to shift their focus away from medical school exams as they begin to prepare for Step 1, it should be made clear to students that, at least at UT Southwestern, doing well in courses will translate directly to performing well on Step 1.

CBSSA scores are the single highest predictor of Step 1 scores. This is easy to understand, as this exam is the same format and style of the actual exam, and is a fair assessment of the baseline knowledge of a medical student before going into a dedicated preparation period. Because students are all given essentially the same amount of time, it makes sense that those who start ahead will tend to stay ahead. That said, there is significant variation in how much students can learn over that 4-6 week period, with some students increasing their CBSSA scores by as much as 85 points. Those scoring below 160 (scaled Step score) have a significant risk of doing poorly on the exam and may warrant more intervention early on.

It was surprising to see that students' grades did not vary significantly between Q2 (Quarter 2) and Q3, and even students whose grades decreased, presumably because they started to shift focus towards step, did not perform better on Step 1. That was not true between Q3 and Q4. Out of all of the second year grades Q4 grades were by far the least predictive of Step scores, and students who dropped between Q3 and Q4 tended to score higher on Step 1. This may be explained by selection bias and confounding factors. Students

who could afford to "slack off" and drop their Q4 grades were mostly students who were not already struggling. Furthermore, by the end of Q3, UT Southwestern's curriculum had covered a bulk of the core Step 1 material, with only dermatology, psychiatry, and integrated systems left to cover in Q4. Because those topics are only a small percentage of what is covered on Step 1, it is possible that students benefited by studying more high-yield Step 1 topics during that quarter, which could result in higher scores. These are the same students who would be expected to have started studying between 3 months and 6 months prior to the exam, which was also correlated with higher Step 1 scores.

Overall, the conclusion remains that doing well on medical school exams will translate to doing well on Step 1, something administrators and educators will appreciate. However, it is possible that as one gets closer to the exam, there may be some benefit to shifting focus away from school and towards the exam, particularly if the school curriculum leaves lower-yield topics for the end of the year.

Survey Data

Although the survey participants scored higher than the class as a whole, the surveys provide useful information to guide preparation strategies.

In the comparison of users of resources versus non-users, none of the data points were statistically significant except that users of Sketchy Pharm® performed significantly worse than non-users. However, this is likely a result of the power of this portion of the study. In general, it can still be concluded that users of question-based resources – Kaplan™ Q Bank, NBME® assessments, UWorld™ Self Assessments, and USMLE Rx™ – performed better on exams than non-users. This may indicate that the best preparation strategies are to answer as many questions as possible. This is further supported by the fact that those

who answered more than 4000 Q-Bank questions outperformed those who answered fewer. Users of the two video resources Doctors in Training™ and Sketchy Pharm™ were the worst performers, which may be due to the fact that video resources take a long time to go through and impede the ability to do more practice questions.

Looking at the weekly surveys, the factors most correlated with high practice exam scores are the cumulative % correct in the Q-bank and progress through the Q-bank, with progress through First Aid™ also correlated with higher scores but less so. This supports the assertion that question bank resources are the highest-yield resources in preparation for this exam.

This data also provides very useful information in guiding students on when to start preparing and for how long. Students at UT Southwestern were all given 6 weeks off for dedicated study, and most students didn't start intense studying until this time. However, it is clear that those who started studying 1 year or 6 months prior to the exam performed better than those who started studying either 3 months prior or 6 weeks prior. That said, once the dedicated study period started, the results were less clear, as there was no significant difference in scores among different weeks studying or total hours studying during the dedicated preparation period. The average scores for those who studied fewer than 250 total hours and those who studied over 600 hours were within the margin of error, along with each interval in between. More important, it seems, was the level at which students came into the preparation period. That said, there was a significant difference in hours studied between those who scored below 220 and those who scored 220 or higher, so it would not be unreasonable to suggest that students should still aim to study 300 hours, which is 1 standard deviation below the average for those who scored 220 or higher.

This correlates to roughly 50-60 hours per week for 5-6 weeks. Of note, this was retrospective data and may not accurately represent the exact time studying, as it may include time spent on breaks, stretching, eating, etc.

This number is further supported by the weekly surveys. It is clear that students have diminishing returns after each week of study. Fitting a quadratic function to the first six weeks (week 7 was an outlier because these were students who requested extra time and were the lowest performers) suggests that after Week 6, students should expect, on average, no more than 1-2 points per week, and even if they study 10 weeks, they are unlikely to get more than 6 points. By looking at just the delta, the line of best fit is even more conservative, stating that students are unlikely to make any increase in scores after week 6. Of course, there are exceptions, but for most students, it appears that scores may start to either stabilize or even decrease after week 6. Given this data, it appears that 5-6 weeks is optimal for dedicated preparation time among students at UT Southwestern.

Developing a Model to Predict Step 1 Scores

There were a number of factors that did not show strong correlations when looking at the entire group of students, but when only comparing those who scored <220 to those who scored 220 or higher, a few more significant relationships emerge. Undergraduate GPAs and MCAT® scores, as expected, were significantly different between the two groups. Also significant were the number of questions answered and the total hours studying, both with p values <0.02. This suggests that although on the whole the correlation may be weak, those who perform better meet at least a certain threshold of studying. The other notable significant factors are the CBSSA scores and cumulative MS2 grades, also as expected based on the prior data.

The model developed can serve as a very useful resource for educators to flag students at risk of poor performance on Step 1, defined as scoring below a 220. It uses a threshold for undergraduate GPA and MCAT® scores, with points deducted if a student has lower than a 3.5 or scores 30 or below on the MCAT®, and it applies a coefficient to the students' CBSSA score and medical school grades. The undergraduate GPA and MCAT® don't individually have low p-values in the model and contribute minimally compared to the CBSSA and medical school grades, but they do still improve the predictive value of the model.

The model is able to predict 67% of scores within 8.87 points, and 95% of scores within 17.39 points. More importantly, though, it can serve as a good way to flag students at risk of scoring poorly on the exam. If we use the 67% confidence interval and mark the upper limit of the model at 228.87, it will catch 81% of those who would score below 220 with 86% specificity. It also has a negative predictive value of 96%, so it can also alleviate concerns about a majority of the remainder of the student body. Of note, although the model can serve as a useful resource for educators and administrators to flag at-risk students, individual students should use it with caution. Some of those who are predicted to score low may very well do very well on the exam, just as those predicted to perform well may do poorly on the exam. The model is a useful guide, but it does not predict destiny.

CONCLUSION

This project demonstrates that objective data analysis can be used to explain studying strategies and to guide students towards optimal preparation for Step 1. The same data can be used by educators to flag students at risk of poor performance to intervene early.

The first aim of the project was to analyze pre-matriculation data. At UT Southwestern, it was shown that most pre-matriculation data serve as only weak predictors of medical school or USMLE® exam performance, and that demographic information has no influence on performance in medical school. However, there is a certain threshold that indicates basic competence, under which students will be at risk of poor medical school performance. This suggests that medical school admissions officers should use GPA and MCAT® scores to make sure students reach a certain threshold, but be cautious about placing too much weight on these characteristics beyond that point and look instead to other aspects of the applications. Future studies in this area could further explore the predictive power of MCAT® and GPAs for a more powerful study, and also explore other aspects that may be useful in sorting applications.

The second aim of the project was to determine if performance in medical school correlated to performance on the Step 1 exam. It is abundantly clear that medical school grades are some of the strongest predictors of Step 1 performance, so students who perform well in medical school should be expected to perform well on board exams. However, it is important to note the small caveat shown in students who shifted their focus in the last few months of the school year during less high-yield topics – those whose scores

went down in the last quarter performed better on Step 1. Although the implications of this data is limited at this point, as this class was the first to switch to pass/fail at UT Southwestern and the curriculum has been since reformed for future classes, this highlights the importance of ensuring that the medical school curriculum remains highly relevant throughout the preclinical years, and perhaps points towards a need for more Step-focused coursework towards the end of the pre-clinical years.

The third goal of this project was to identify if the surveys might provide useful studying tips. The data indicates that the best way to prepare for these board exams is to go through as many practice questions as possible. The data also indicates that studying earlier will lead to higher scores, but that once the dedicated preparation period has begun, 5-6 weeks seems optimal. Two distinct models both show that scores will increase only minimally, if at all, after the 5-6 week mark, so studying further than that amount would be a poor use of time. Further studies should look into new resources, as well as aim for more power for each resource in order to better determine what the best resources truly are.

Finally, this project aimed to develop a model to catch those at risk of scoring below 220 on Step 1. Utilizing a threshold of undergraduate GPA and MCAT® scores, and a coefficient multiplier for medical school grades and initial CBSSA scores, a model was developed with significant predictive potential. This project shows that a mathematical screening tool might be able to result in tangible success by virtue of preventing failure for a significant number of at risk students. Next steps include validating this model across different classes at UT Southwestern and across multiple campuses before it can be confidently used to stratify students, but it at least provides guidance on what medical school educators and administrators should look for as they prepare students for Step 1

exams. Future studies should also determine if early intervention will improve the outcome for at-risk students, because it is possible that these students may not perform significantly higher even with appropriate interventions.

LIST OF TABLES

***Table 1**. Distribution of Pre-Admissions Scores for UT Southwestern Class of 2018.*

Variable	Average	Standard Deviation
MCAT® Score	33.79	2.94
Undergraduate GPA	3.83	0.18

***Table 2.** R^2 values for MCAT® sections compared to Step 1 scores and Medical School Grades*

MCAT® Section	R^2 for Step 1 Scores	R^2 for Medical School Grades
Biological Science (BS)	0.06717	0.06577
Physical Science (PS)	0.09543	0.02945
Verbal Reasoning (VR)	0.00041	0.00081
BS + PS	0.12715	0.07188
Total Score	0.10168	0.05981

***Table 3.** Correlation statistics for second year quarter term grades compared to Step 1 scores.*

Grading Term	R^2 for Step 1 Scores
Quarter 1	0.4425
Quarter 2	0.45742
Quarter 3	0.4886
Quarter 4	0.29051
Cumulative	0.52221

Table 4. *Distribution and correlation statistics for grade change between second year quarter term grades compared to Step 1 scores.*

Terms	**Average Grade Change (%)**	**Standard Deviation**	R^2 for **Step 1 Scores**
Q1-Q2	-2.36	3.32	0.13717
Q2-Q3	1.48	3.26	0.0088
Q3-Q4	2.81	3.52	0.0184

Table 5*. Distribution of scores of survey population versus actual population of UT Southwestern Class of 2018*

	Survey	**Actual**
Number of Students	140	228
Average	242	237
Standard Deviation	14	17.5
Min	202	169
Max	272	272

Table 6. *Distribution and t-tests statistics comparing students who scored <220 versus those who scored 220 or higher on Step 1*

	<220		**220+**		
Category	**AVG**	**SD**	**AVG**	**SD**	**p-value**
Ugrd GPA	3.71	0.29	3.83	0.184	0.01741
MCAT®	31.77	2.94	34.15	2.55	0.00003
# Questions	2173.33	568.48	2527.49	686.76	0.01838
# Weeks Studying	5.73	0.729	5.73	0.928	0.4937
# Days per Week	6.47	0.48	6.54	0.6	0.28827
Total Hours Studying	350	109	420	120	0.01804
Passes Through First Aid	2.09	0.99	2.46	1.02	0.097
Number of Practice Exams Taken	4.2	1.47	4.75	1.47	0.09245
CBSSA Score	159	14	199	25	0
Change between NBME and Step scores	46.4	15.8	43.2	18.5	0.15894
Cumulative MS2 Grades %	78.98	3.84	85.51	4.52	1.42E-11

Table 7. *Positive predictive value for scoring <220 on Step 1 for various risk factors*

Risk Factors of Scoring <220	**Postitive Predictive Value**
Undergraduate GPA <3.5	45%
MCAT® <31	48%
<2000 Q-bank questions answered	27%
<250 total hours studied	30%
Initial CBSSA scaled score <160	64%
Cumulative Grades <80% (in bottom 15% of class)	60%

Table 8. *Regression statistics for multiple linear regression model for predicting USMLE Step 1 scores*

Regression Statistics				
R	0.80236			
R-square	0.64379			
Standard Error	8.86639			
N	142			
USMLE® Step 1 Score = 100.84962 + 4.85204 (*if Undergraduate GPA <3.5*) + 3.66261 (*if MCAT® <31*) + 0.31215 * [CBSSA Score] + 0.8223 * [Cumulative MS2 Grade %]				
	Coefficient	**Standard Error**	**p-level**	**H0 (5%)**
Intercept	100.84962	14.81136	2.83E-10	rejected
Undergraduate GPA	4.85204	4.70029	0.30376	accepted
MCAT® Score	3.66261	2.98453	0.22185	accepted
CBSSA Scaled Step 1 Score	0.31215	0.04042	2.15E-12	rejected
Cumulative MS2 Grade %	0.8223	0.22227	0.00031	rejected

LIST OF FIGURES

Note: Error bars on all graphs indicate a 95% confidence interval.

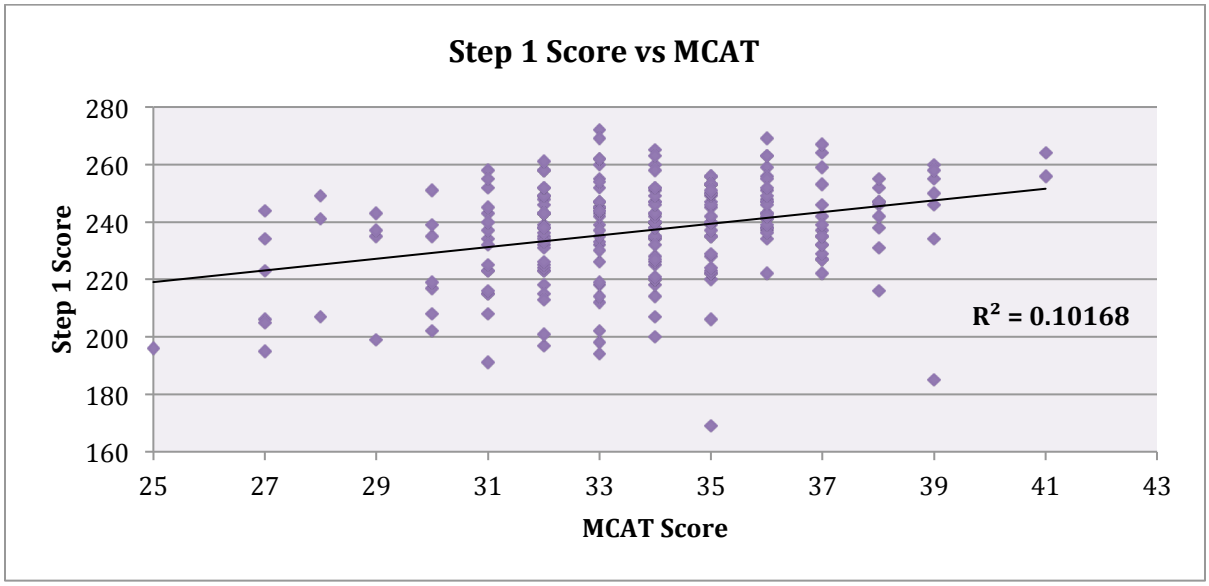

***Figure 1**. Correlating USMLE Step 1 scores with MCAT® Scores. R^2 indicates moderate correlation.*

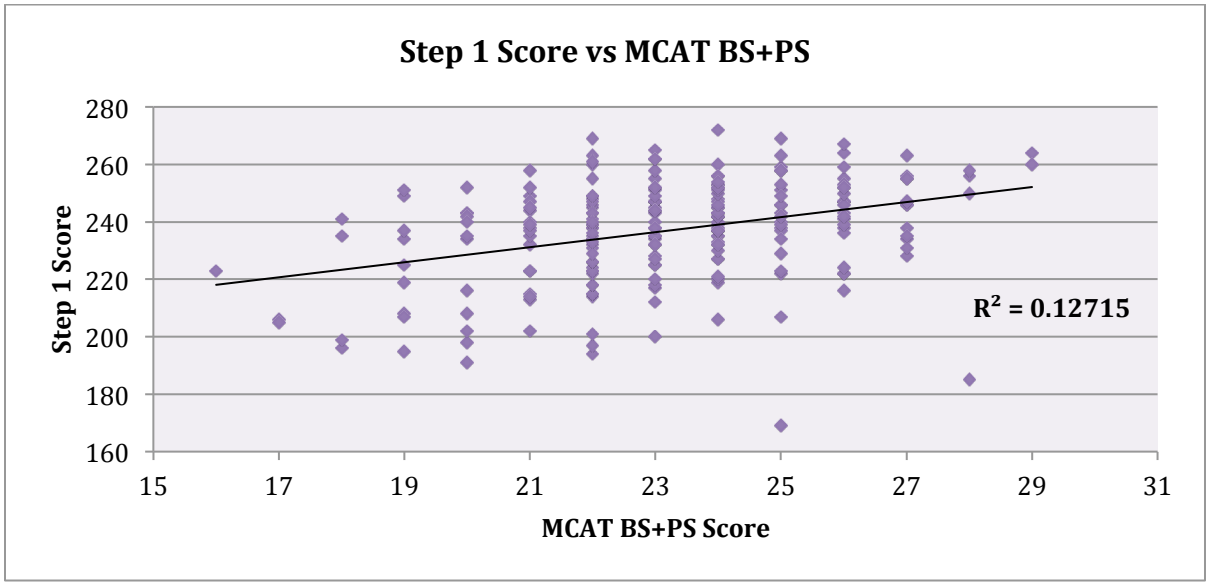

***Figure 2.** Correlating USMLE Step 1 scores with sum of Biological Sciences and Physical Sciences MCAT® Scores. R^2 indicates moderate correlation.*

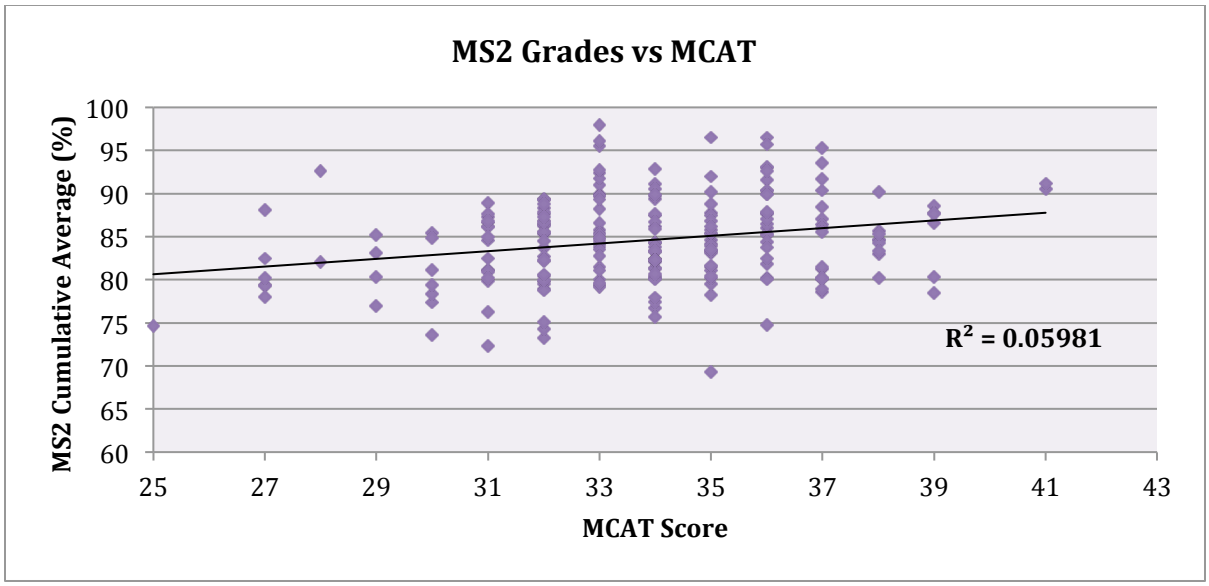

Figure 3. *Correlating second year medical school grades with MCAT® Scores. R^2 indicates weak correlation.*

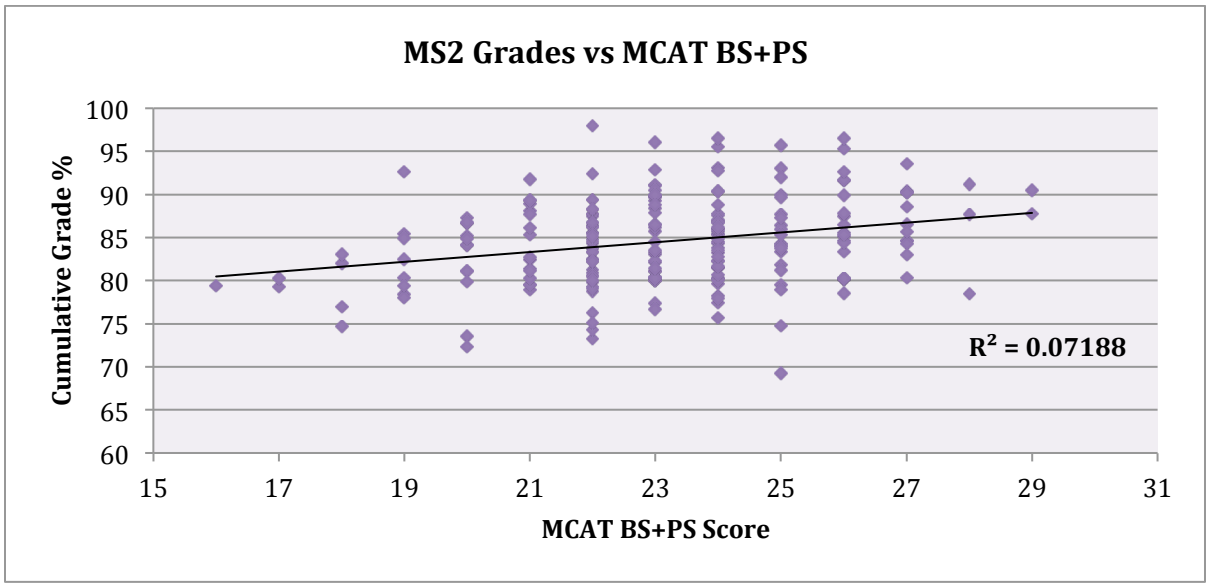

Figure 4. *Correlating second year medical school grades with the sum of Biological Sciences and Physical Sciences MCAT® Score. R^2 indicates weak correlation.*

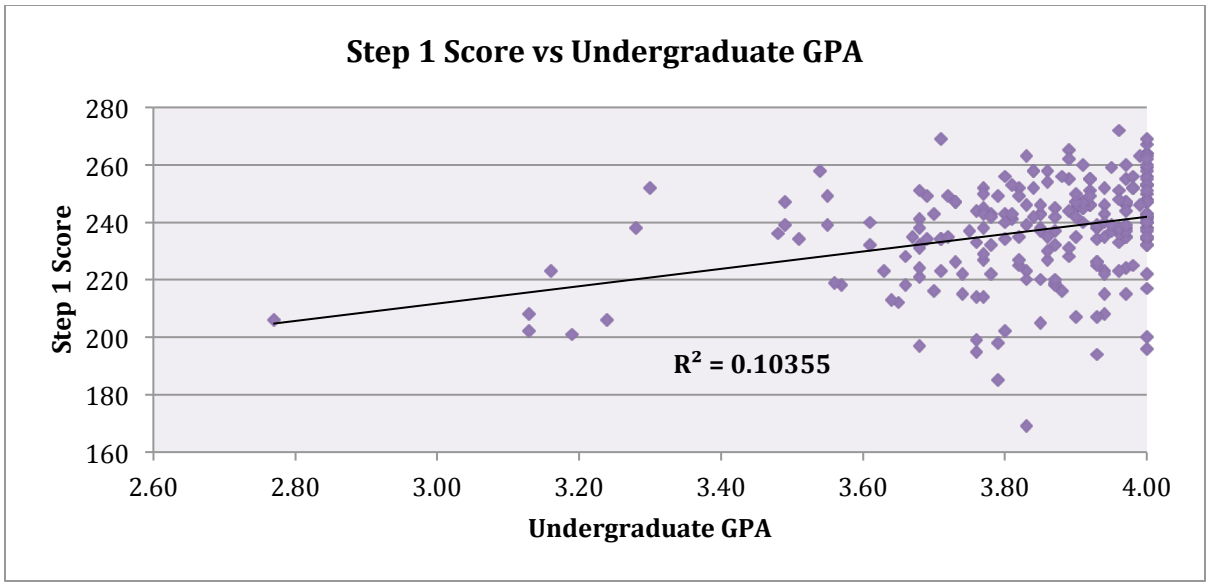

Figure 5. *Correlating Step 1 scores with undergraduate GPA. R^2 indicates moderate correlation.*

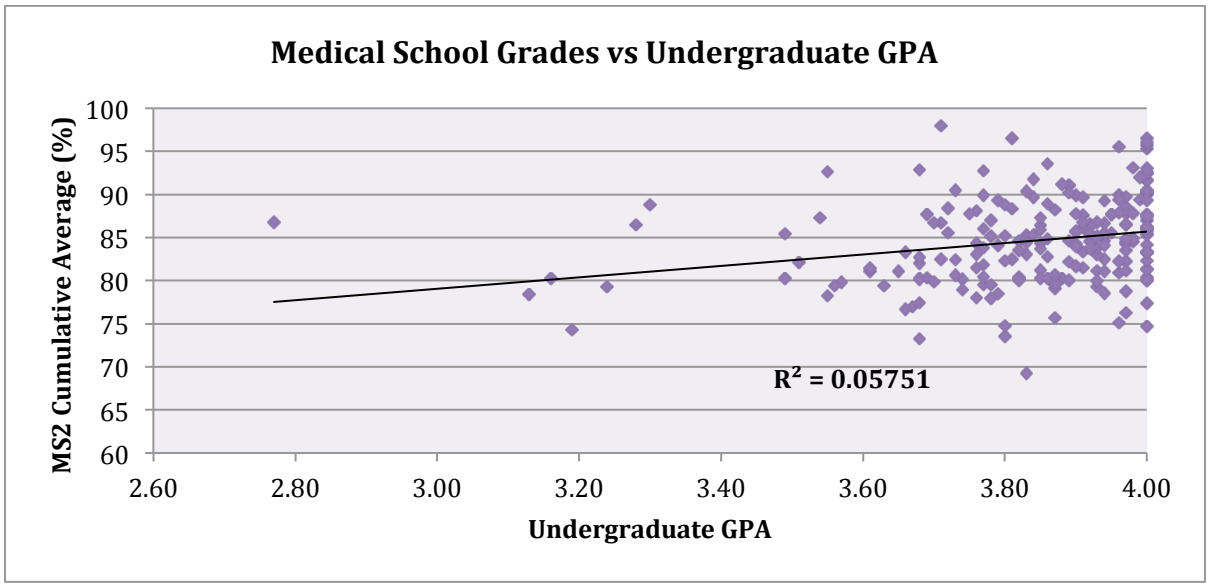

Figure 6. *Correlating second year medical school grades with undergraduate GPA. R^2 indicates weak correlation.*

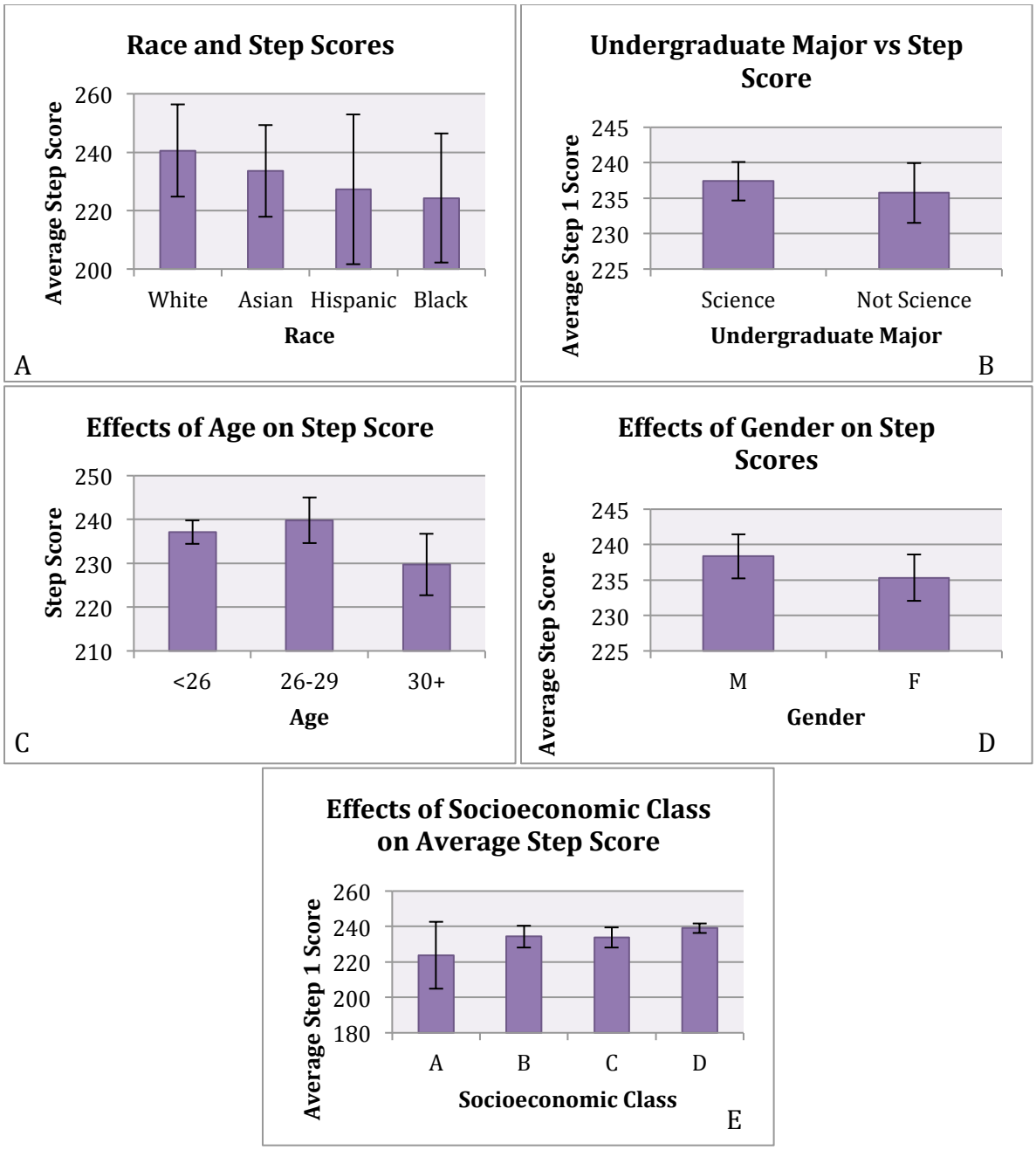

Figure 7. *Average Step 1 scores split by demographic factors. (A) shows averages among different races, (B) shows science versus non-science majors, (C) shows averages among different age groups, (D) shows averages split by gender, and (E) shows split based on socioeconomic categories, with A being low socioeconomic status and D indicating affluent family background.*

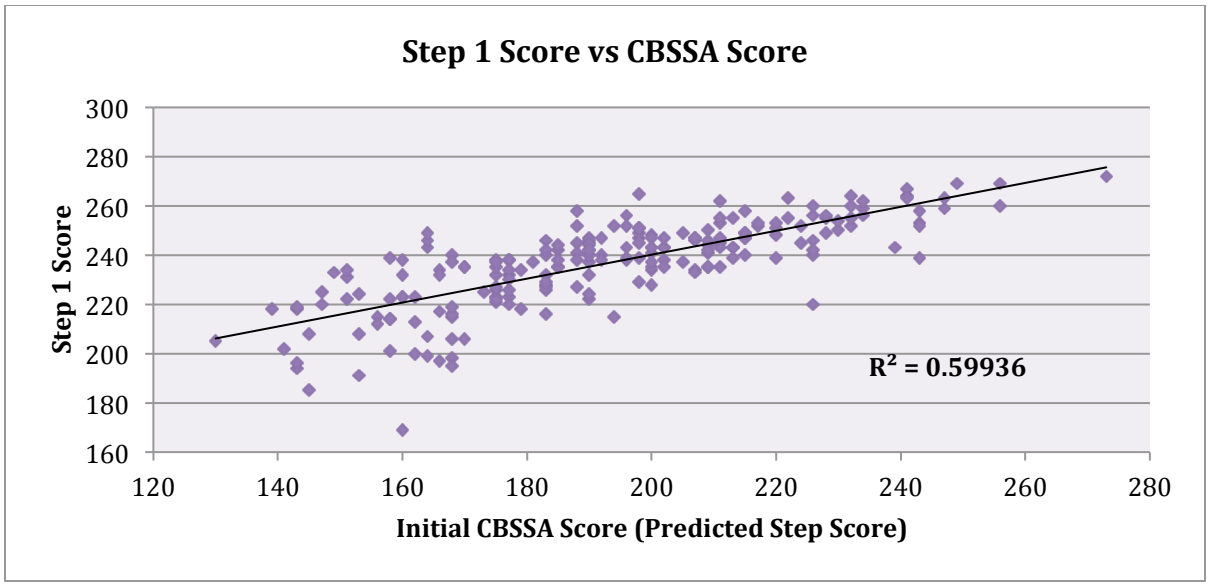

Figure 8. *Correlating Step 1 score with initial CBSSA scaled score. R^2 indicates very strong correlation.*

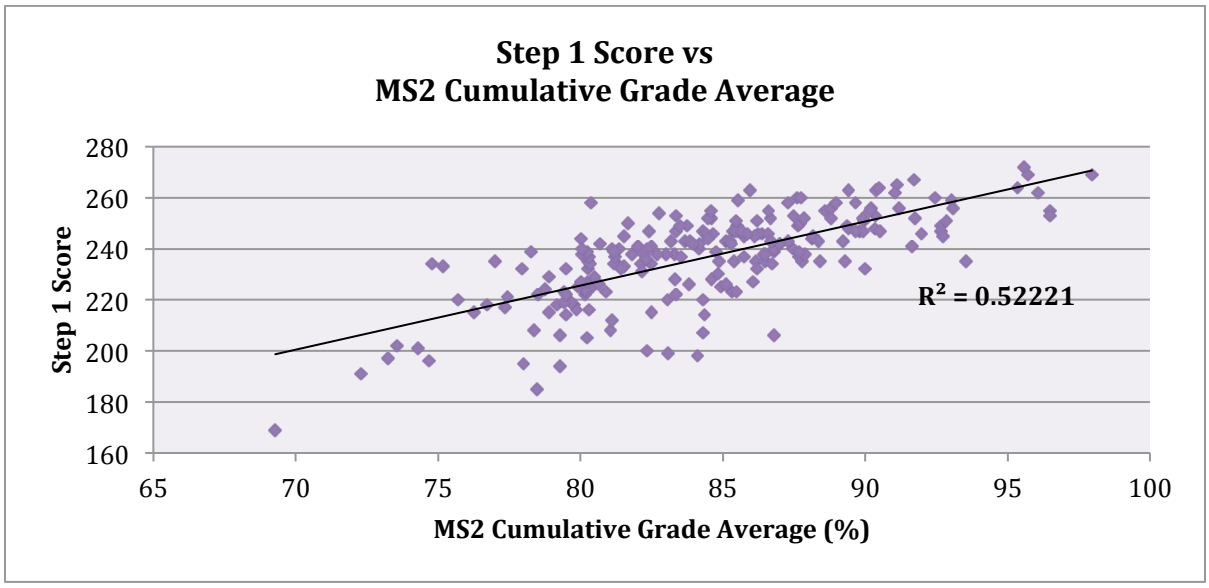

Figure 9. *Correlating Step 1 score with second year medical school cumulative grade average. R^2 indicates very strong correlation.*

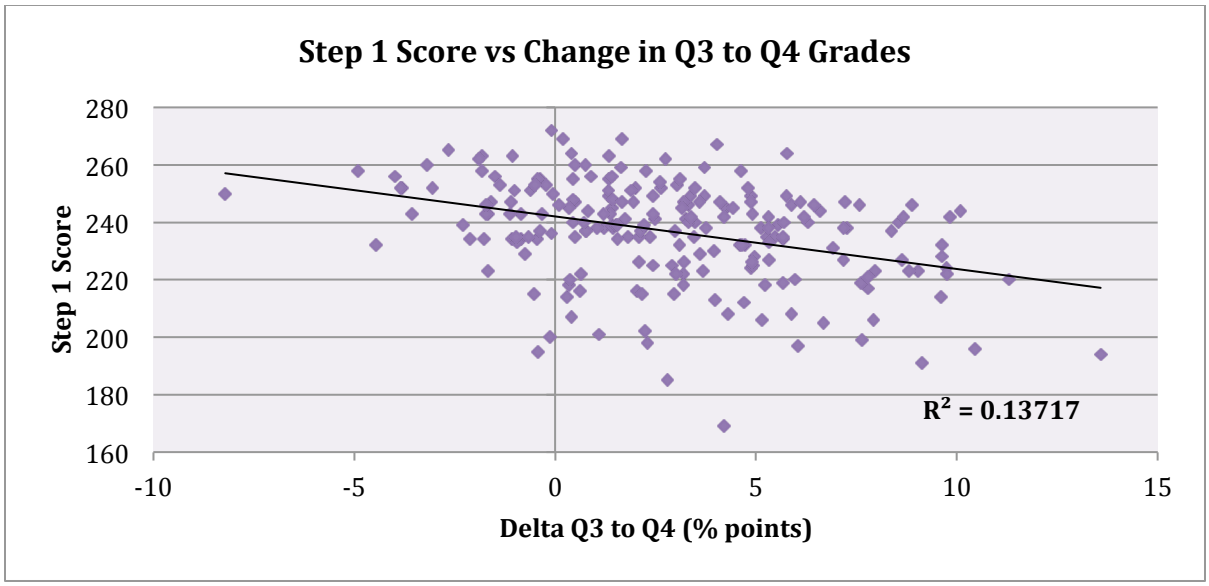

Figure 10. *Correlation between the change in grades between third quarter and fourth quarter grading terms and a student's Step 1 score. This suggests that students whose grades drop between Q3 and Q4 perform better on the exam, with R^2 indicating moderate correlation.*

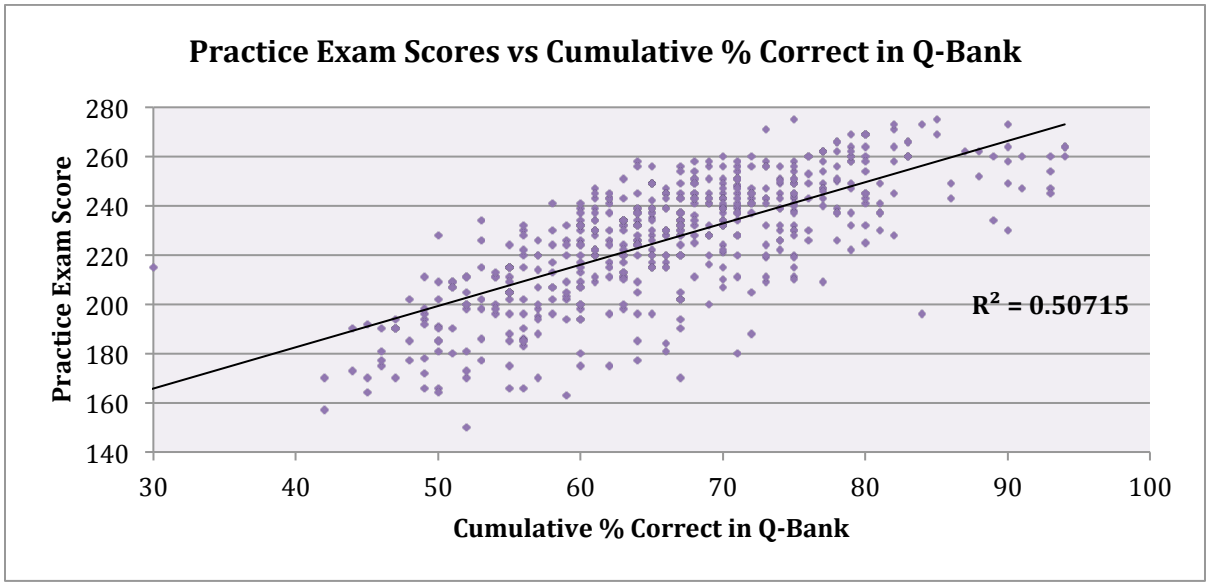

Figure 11. *Correlating survey data of cumulative % correct in the UWorld, Kaplan, or USMLE Rx Q-banks with practice exam scores in the same week. R^2 indicates strong correlation.*

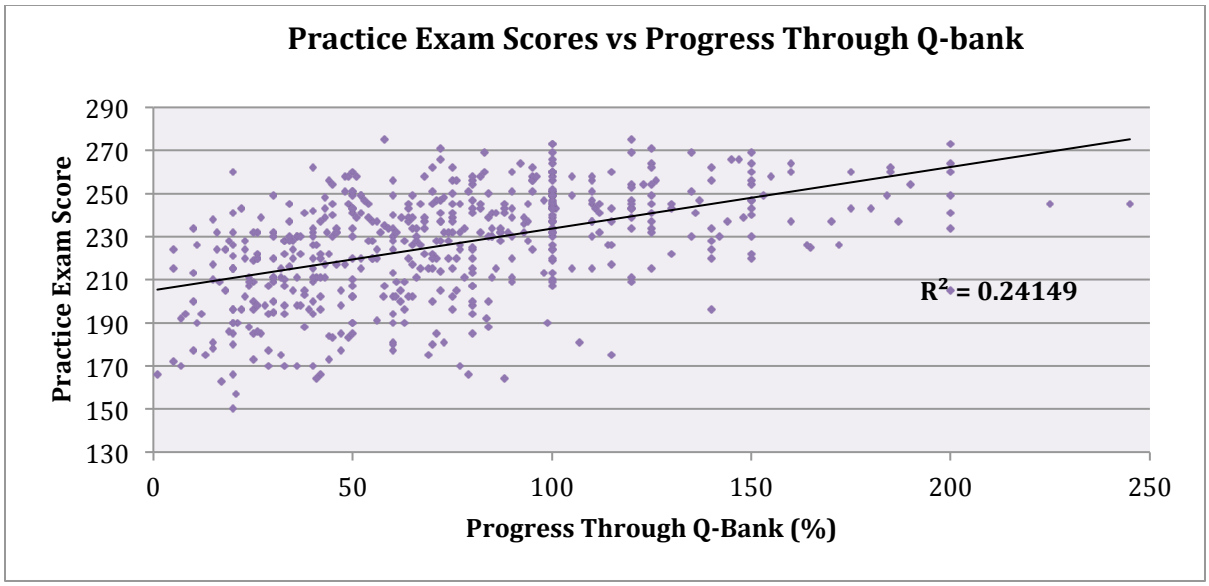

Figure 12. *Correlating survey data of progress through UWorld, Kaplan, and/or USMLE Rx Q-bank with practice exam scores in the same week. 150% can indicate either finishing the Q-bank once and repeating 50%, or finishing 1 full Q-bank and 50% of another. R^2 indicates moderate/strong correlation.*

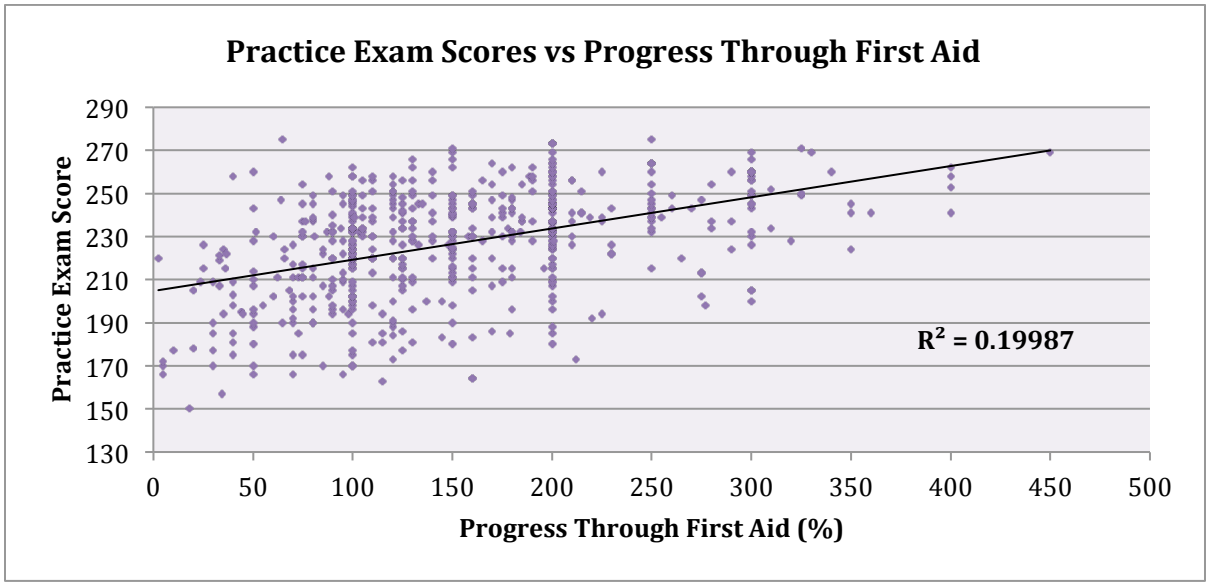

Figure 13. *Correlating survey data of progress through First Aid with practice exam scores in the same week. 150% can indicate either finishing the Q-bank once and repeating 50%, or finishing 1 full Q-bank and 50% of another. R^2 indicates moderate correlation.*

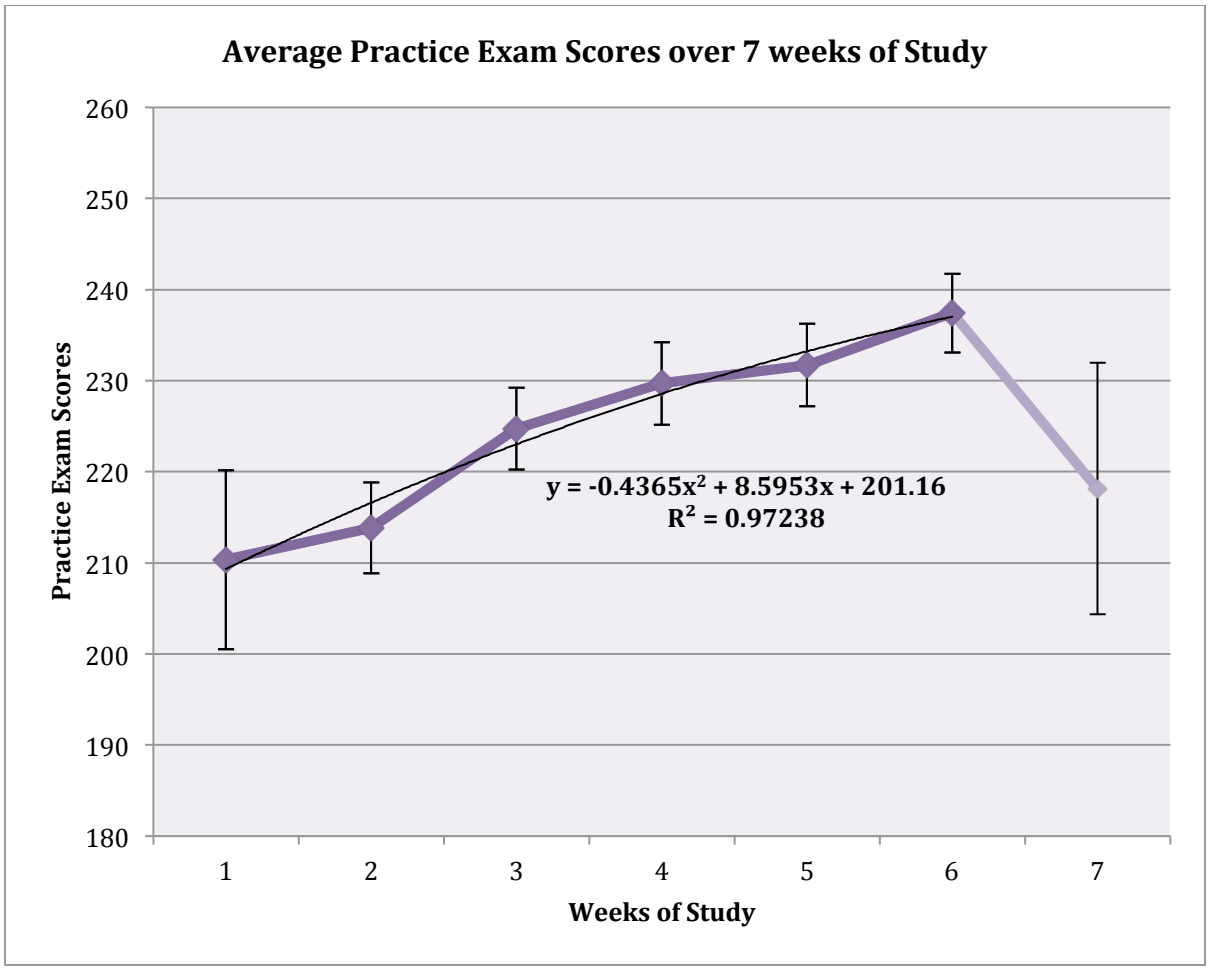

Figure 14. *Survey data showing average practice exam scores for each week of the survey. A quadratic best-fit function for just weeks 1 through 6 is superimposed, with the R^2 indicating very high correlation between the model equation and the data. Week 7 was an outlier because there were so few students, all of which specifically requested extra time to study.*

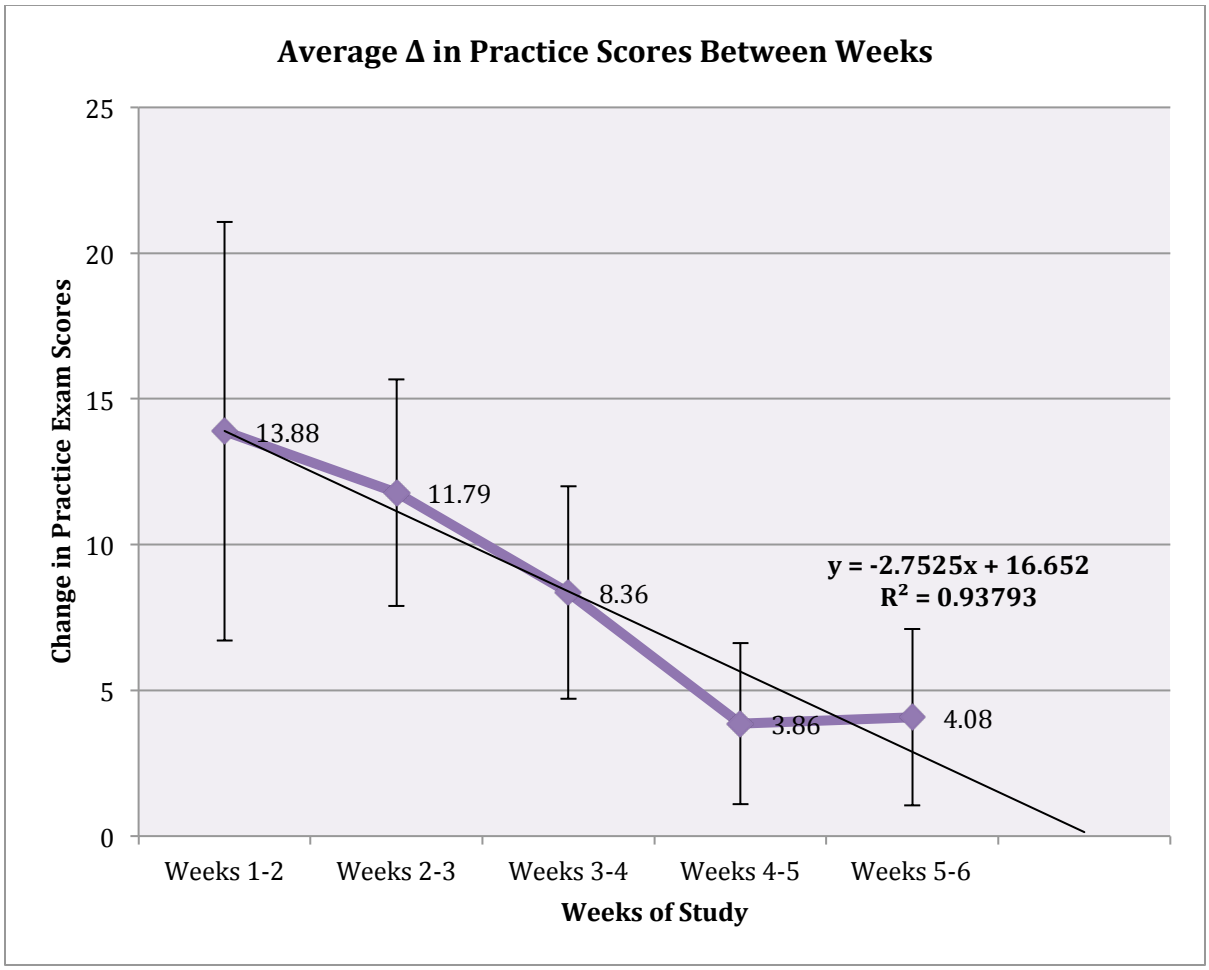

Figure 15. *Survey data showing the change in scores from one week to the next. Only consecutive data points were included. A linear best-fit function for weeks 1 through 6 is superimposed, with the R^2 indicating very high correlation between the model equation and the data. Week 7 was an outlier because there were only 5 students with consecutive data points, all of which specifically requested extra time to study.*

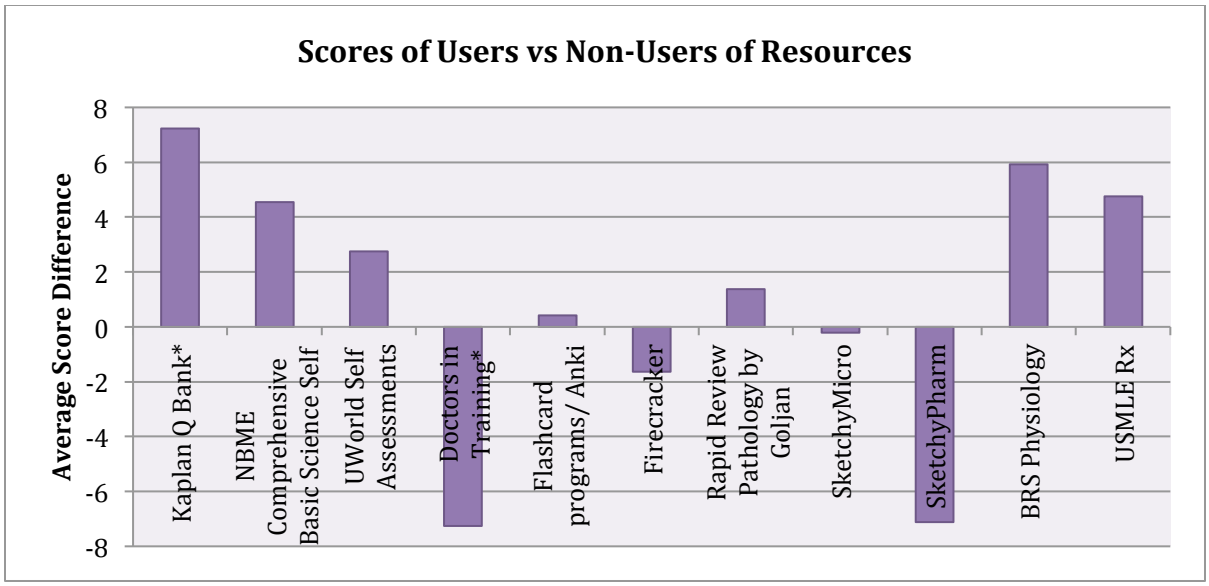

Figure 16. *Average scores of users versus non-users of each resource. Only Sketchy Pharm had p<0.05. Asterisks indicate fewer than 10 users of a resource. Kaplan Q-bank, Firecracker, and USMLE Rx are question banks. NBME and UWorld Self Assessments are practice exams. DIT, Sketchy Micro, and Sketchy Pharm are video-based resources. Rapid Review and BRS are texts.*

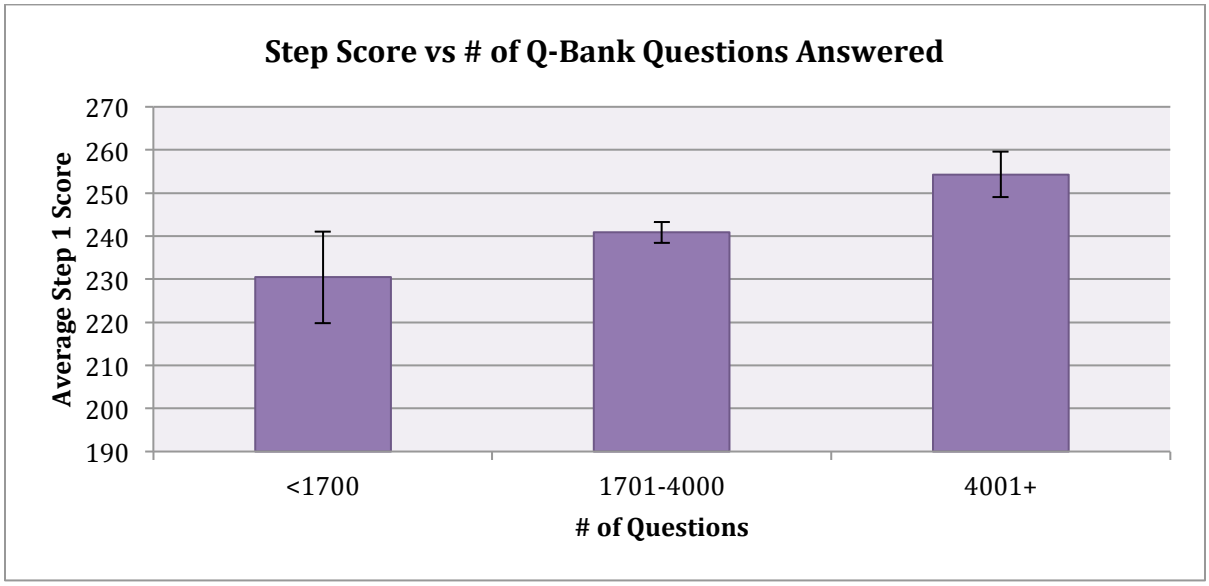

Figure 17. *Average Step 1 score broken down by number of Q-bank questions answered. The full UWorld Q bank, the most commonly used resource, is roughly 2350 questions. 1700 then indicates less than 50% of the Q-bank and 4000 indicates 150%.*

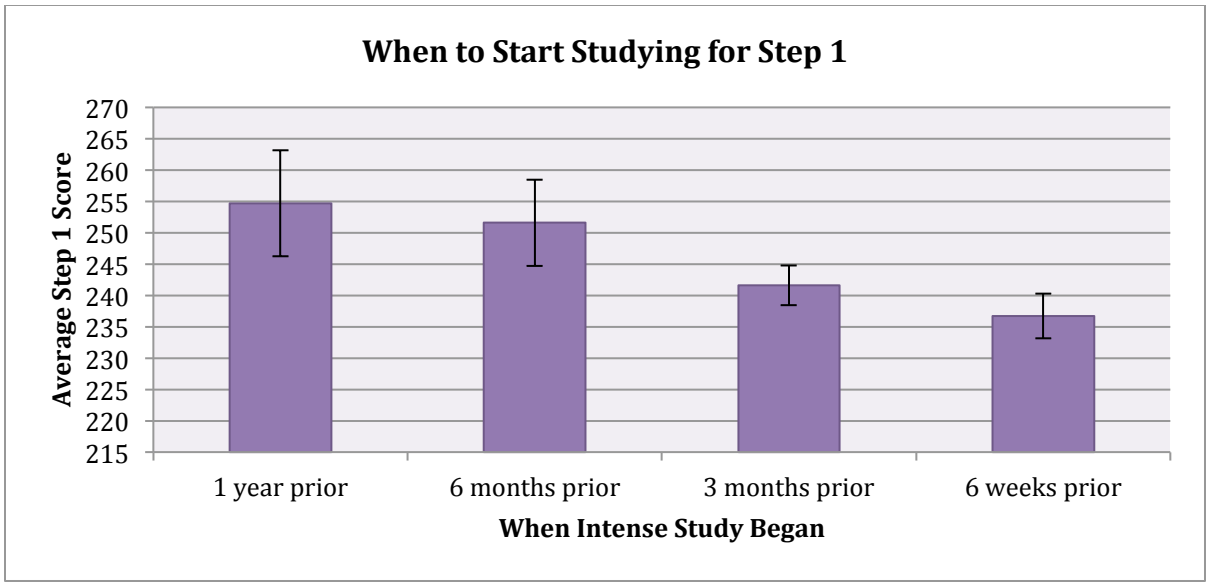

Figure 18. *Average Step 1 scores based on how much prior to taking the exam the student began "intense" study for Step 1. 6 months corresponded to winter break, and 3 months corresponded to spring break. 6 weeks prior was the default, as that was how much time each student was guaranteed for dedicated study time.*

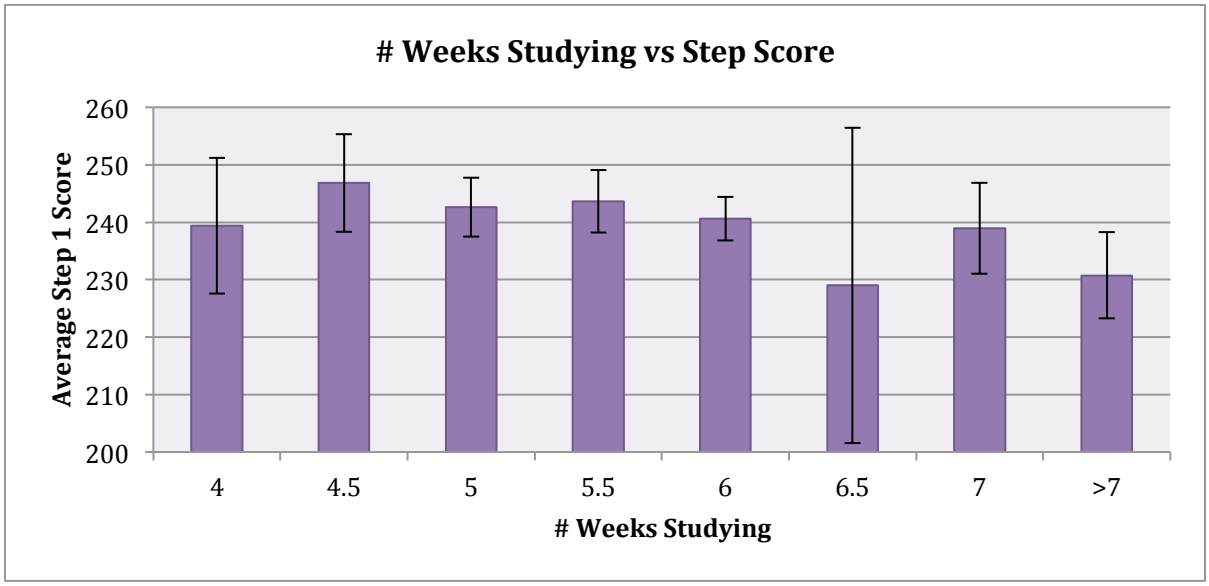

Figure 19. *Average step scores broken down by number of weeks of dedicated study.*

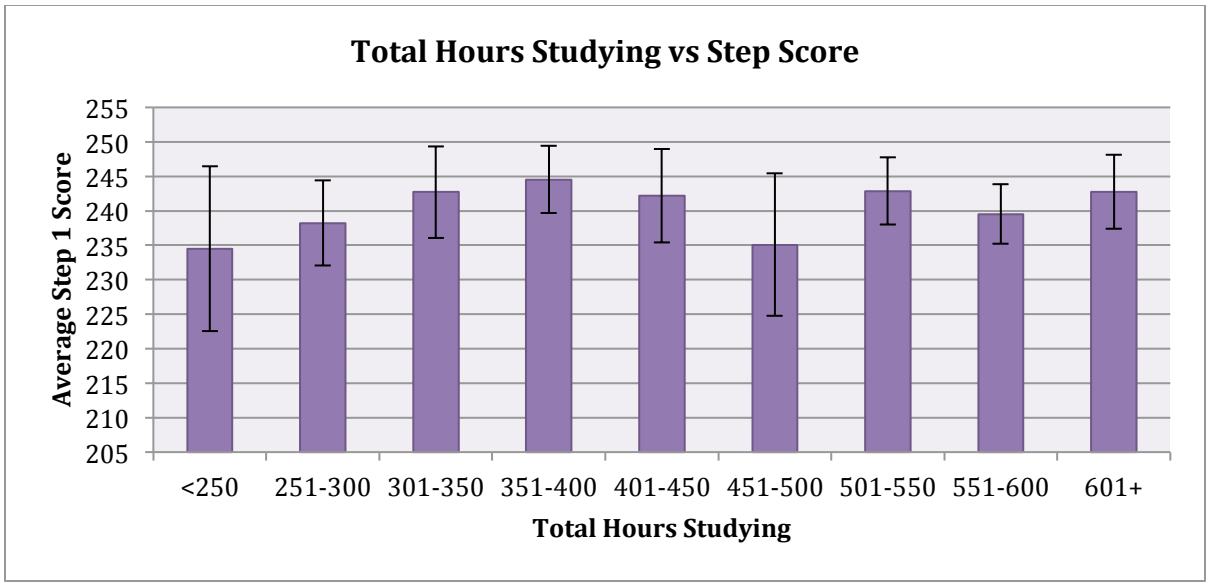

Figure 20. *Average step scores broken down by total hours studying during the dedicated study period.*

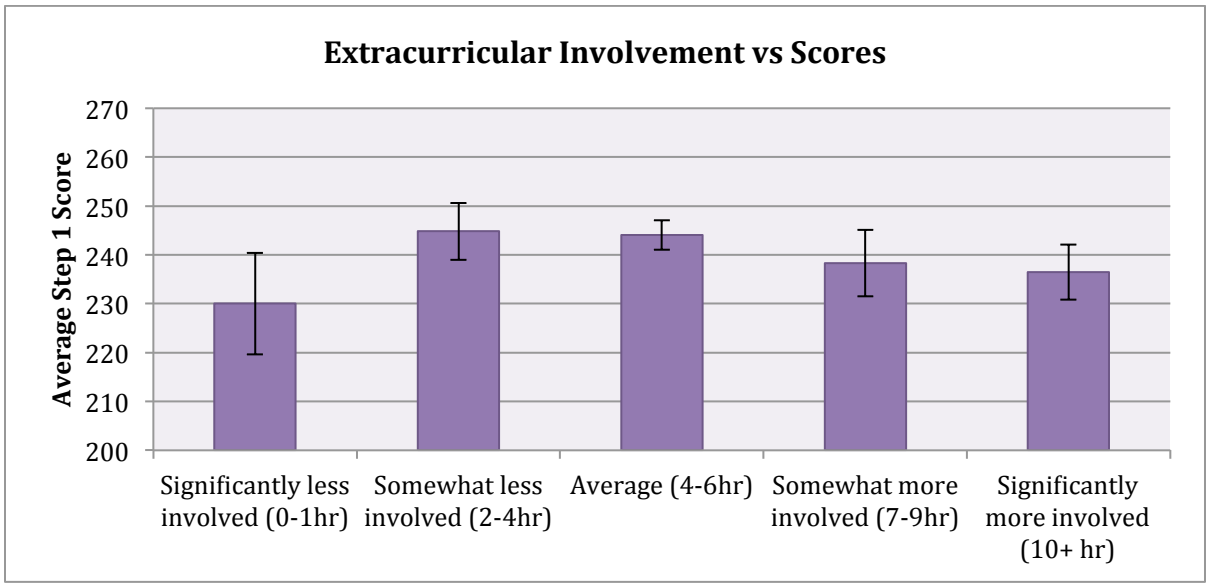

Figure 21. *Average Step 1 scores broken down by level of extracurricular involvement.*

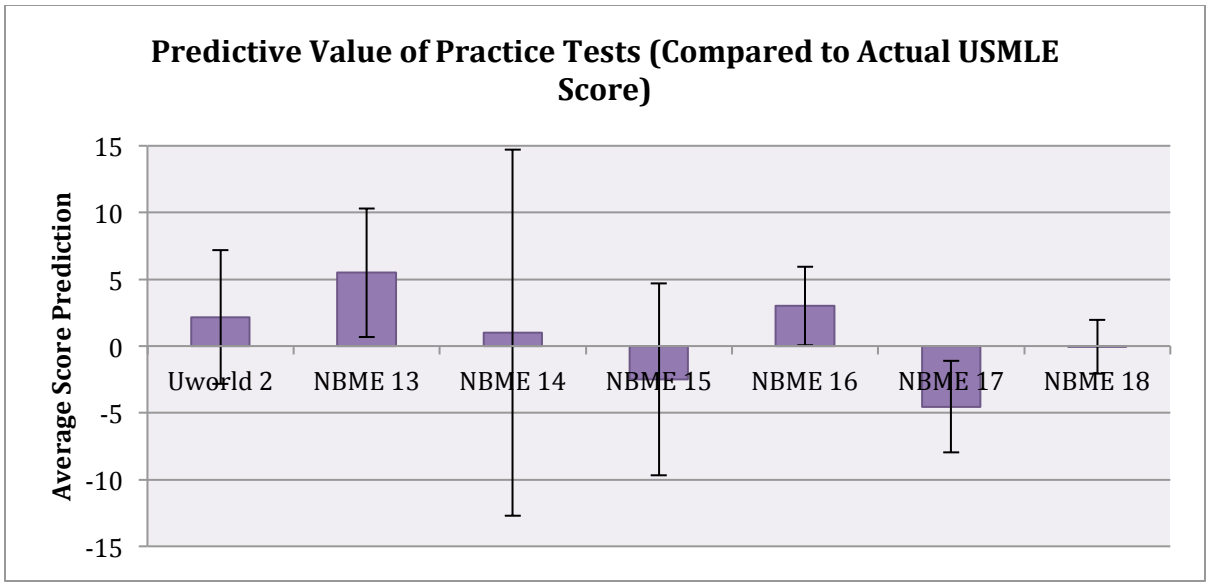

Figure 22. *Average predictive value of practice tests compared to the actual exam score. These exams were mostly taken within a week of the actual exam. Error bars indicate 95% confidence interval.*

REFERENCES

1. National Resident Matching Program. (2016, June). *Results of the 2016 NRMP Program Director Survey*. http://www.nrmp.org/. Retrieved February 1, 2016.
2. London, D. A., Kwon, R., Atluru, A., Maurer, K., Ben-Ari, R., & Schaff, P. B. (2016). *More on How USMLE Step 1 Scores Are Challenging Academic Medicine*. Academic Medicine, 91(5), 609-610. doi:10.1097/acm.0000000000001159
3. National Resident Matching Program. (2016, June). *Results of the 2016 NRMP Program Director Survey*. http://www.nrmp.org/. Retrieved February 1, 2016.
4. Search the AMA Residency & Fellowship Database™. (n.d.). Retrieved March 25, 2017, from https://www.ama-assn.org/life-career/search-ama-residency-fellowship-database
5. Kleshinski, J., Khuder, S.A., Shapiro, J.I. et al. Adv in Health Sci Educ (2009) 14: 69. doi:10.1007/s10459-007-9087-x
6. Johnson, J., Jordan, E., Burton, W., & Silhiger, S. (n.d.). *Are Questions the Answer? The Effect of Popular Study Resources on USMLE Step 1 Performance* . Lecture. Retrieved March 10, 2017, from https://members.aamc.org/eweb/
7. Prober, Charles G. MD; Kolars, Joseph C. MD; First, Lewis R. MD; Melnick, Donald E. MD. (2016). *A Plea to Reassess the Role of United States Medical Licensing*

Examination Step 1 Scores in Residency Selection. Academic Medicine, 91(1), 12-15. doi:10.1097/ACM.0000000000000855

8. Gullo, C. A., Mccarthy, M. J., Shapiro, J. I., & Miller, B. L. (2015). *Predicting Medical Student Success on Licensure Exams.* Medical Science Educator, 25(4), 447-453. doi:10.1007/s40670-015-0179-6
9. 34 CFR 99.31 - *Under what conditions is prior consent not required to disclose information?* Cornell LII. (n.d.). Retrieved March 23, 2017, from https://www.law.cornell.edu/cfr/text/34/99.31
10. *Pearson Product-Moment Correlation*. (n.d.). Retrieved March 23, 2017, from https://statistics.laerd.com/statistical-guides/pearson-correlation-coefficient-statistical-guide.php
11. *Medical School Graduation Questionnaire 2016 All Schools Summary Report*. (Rep.). (2016, June). Retrieved March 23, 2017, from AAMC website: www.aamc.org/

VITAE

Sachin Shah (July 30 1991-present) is a fourth year medical student at UT Southwestern Medical School and will be starting his pediatric residency at UT Southwestern in June. He grew up in Plano, Texas, and was a Eugene McDermott Scholar as an undergraduate student at the University of Texas at Dallas. He graduated with a B.S. in Biology in 2013 and immediately began medical school. In medical school, Sachin served as Co-President of his medical school class for all four years and was drawn to medical education. Sachin will graduate with his M.D. and a distinction in Medical Education in June 2017, and he hopes to pursue a career as an attending physician at an academic teaching hospital.

Permanent Address:

8024 Grand Canyon Dr.

Plano, TX 75025

APPENDIX

APPENDIX A – WEEKLY STUDENT ACADEMIC SUPPORT SERVICES SURVEY

Step-1 Support Survey: Week One

Our weekly survey is to gather preparation data for the Step-1 success of current and future students. Thanks so much for your participation!

*** 1. Please select your name from the drop-down list.**

*** 2. What is your goal Step-1 score?**

3. Did you take a practice test this week? If *not scheduled*, skip to question 4...

Yes, and my predicted Step score was

No, because

*** 4. Please indicate progress in your q-bank (by percentage complete)**

*** 5. Please indicate your cumulative % correct in the q-bank:**

*** 6. What % of First Aid have you completed? (eg: halfway through 2nd pass = 150%)**

7. Do you have any questions for us?

We are here to support your success!

If you would like to speak with Dr. Arlene Sachs or Ms. Carol Wortham in Student Academic Support Services about your progress, please call 214/648-9590 to set up a phone or on-campus consultation.

If you would like to speak with someone in Student Wellness & Counseling about maintaining wellness, please call 214/645-8680 to set up an appointment.

Done

Powered by

See how easy it is to create a survey.

APPENDIX B – POST-EXAM 1 STUDENT AFFAIRS SURVEY

Exit this survey

Step 1 Preparation 2016

1. Step 1 Preparation

Thank you so much for participating in this survey. Your participation will really help future students. We are asking for your name in order to link these responses to other data including the weekly surveys you filled out during the dedicated study period. Per IRB protocol, only Carol Wortham will be able to link a person to a survey response and will remove all identifiers before providing the research team with a completely de-identified data set. If you prefer not to have your survey responses linked to other de-identified information, you can skip question one. You also don't have to answer any question with which you feel uncomfortable.

Thanks again for your time!

1. Please select your name from the drop-down list.

2. Which USMLE Step 1 materials/courses did you use?

- [ ] First Aid for Step 1
- [ ] USMLE World
- [ ] Kaplan Q Bank
- [ ] NBME Comprehensive Basic Science Self Assessments (CBSSA)
- [ ] UWorld Self Assessments
- [ ] Doctors in Training Course
- [ ] Pathoma
- [ ] Pismania

PREVIEW & TEST

- [ ] Firecracker
- [ ] Rapid Review Pathology by Goljan
- [ ] SketchyMicro
- [ ] SketchyPharm
- [ ] Med Bullets
- [ ] Online MedEd
- [ ] BRS Physiology
- [ ] Osmosis
- [ ] Other (please specify)

3. How did you decide what resources to use?

4. How many Kaplan Qbank or USMLE World questions did you complete? (UWorld has 2350 questions. You can estimate if unsure.)

5. When in the MS2 year did you start serious (weekly) review/study for Step 1?

- ◯ 1st Semester
- ◯ January
- ◯ Spring Break
- ◯ After Second Year

6. How many weeks did you study full-time for this exam, after MS2 classes ended?

7. When studying full-time, on average how many days per week did you study?

8. When studying full-time, on average how many hours per day did you study?

9. How many passes through First Aid did you make? (Enter 0 if you did not use.)

10. Did you find the NBME Comprehensive Basic Science Self Assessment (CBSSA) given after spring break helpful?

- Yes
- No

Why, or why not?

11. How many practice exams did you take in addition to the one the school provided?

12. What was the FINAL practice exam that you took? (NBME #, UWorld Self Assessment #)

13. Did your practice test over- or underestimate your score?

- ◯ Underestimated
- ◯ Accurate estimate
- ◯ Overestimated

By how many points?

14. How many days before your actual test date did you take your last practice exam?

15. Did you end up moving your USMLE Step 1 exam date?

- ◯ Yes, I moved it earlier.
- ◯ Yes, I moved it later by a few days.
- ◯ Yes, I moved it later by a week or so.
- ◯ Yes, I moved it later by a month.
- ◯ No

If yes, do you regret this decision?

16. How much money did you spend on Step 1 preparation materials/courses? (not including exam administration cost)

17. If finances influenced your choice of materials, please comment.

18. Please indicate your level of agreement with the following:

I felt very prepared for the USMLE Step 1 based on the curriculum at UT Southwestern.

- ◯ Strongly Agree
- ◯ Agree
- ◯ Undecided
- ◯ Disagree
- ◯ Strongly Disagree

Please add comments if your agreement or disagreement has a caveat.

19. Please indicate your level of agreement with the following:

UT Southwestern provided me with enough dedicated study time to prepare for USMLE Step 1.

- ◯ Strongly Agree
- ◯ Agree
- ◯ Undecided
- ◯ Disagree
- ◯ Strongly Disagree

Comments

20. Compared to your peers, how involved are you in extracurriculars, volunteering, and/or research?

- ◯ Significantly less involved
- ◯ Somewhat less involved
- ◯ Average
- ◯ Somewhat more involved
- ◯ Significantly more involved

21. On an average school week, how many hours did you spend on extracurriculars, volunteering, and/or research?

22. Performance on USMLE Step 1 would be improved by placing a larger emphasis on the following topics in the MS1 MS2 curriculum. (Please list specific things)

23. My best piece of GENERAL advice for Step 1 preparation is:

24. Please rate your level of agreement with the following statement as it applies to each service you used.

I found this service beneficial to my preparation for Step 1.
(SASS= Student Academic Support Services)

	Strongly Agree	Agree	Undecided	Disagree	Strongly Disagree	Not Utilized
Results of this Survey from the Prior Class	◯	◯	◯	◯	◯	◯

AOA Step-up to Step 1 review sessions	○	○	○	○	○	○
AOA/SASS Panel Discussion on Step 1 Preparation	○	○	○	○	○	○
Library Simulated Practice Exams	○	○	○	○	○	○
SASS Step 1 Resource Fair and Packet	○	○	○	○	○	○
SASS Calendar Clinic	○	○	○	○	○	○
Personal Consults in SASS	○	○	○	○	○	○
SASS Step 1 Preparation Course	○	○	○	○	○	○
SASS reaching out to me during the preparation period	○	○	○	○	○	○

Comments

25. Please rate your level of agreement with the following statement as it applies to each resource you used.

I found this resource beneficial to my preparation for Step 1.

	Strongly Agree	Agree	Undecided	Disagree	Strongly Disagree	Not Applicable
First Aid for Step 1	○	○	○	○	○	○
USMLE World	○	○	○	○	○	○

Kaplan Qbank	○	○	○	○	○	○
NBME CBSSA	○	○	○	○	○	○
U World Self Assessments	○	○	○	○	○	○
Doctors in Training course	○	○	○	○	○	○
Pathoma	○	○	○	○	○	○
Picmonic	○	○	○	○	○	○
Flashcard programs/ Anki	○	○	○	○	○	○
Firecracker	○	○	○	○	○	○
Rapid Review Pathology by Goljan	○	○	○	○	○	○
Sketchy Micro	○	○	○	○	○	○
Sketchy Pharm	○	○	○	○	○	○
Med Bullets	○	○	○	○	○	○
Online MedEd	○	○	○	○	○	○
BRS Physiology	○	○	○	○	○	○
Osmosis	○	○	○	○	○	○

26. If applicable, please provide your feedback/comments on First Aid for Step 1.

27. If applicable, please provide your feedback/comments on USMLE World.

28. If applicable, please provide your feedback/comments on Kaplan Qbank.

29. If applicable, please provide your feedback/comments on NBME Comprehensive Basic Science Self Assessments.

30. If applicable, please provide your feedback/comments on U World Self Assessments.

31. If applicable, please provide your feedback/comments on the Doctors in Training course.

32. If applicable, please provide your feedback/comments on Pathoma.

33. If applicable, please provide your feedback/comments on Picmonic.

34. If applicable, please provide your feedback/comments on Flashcard programs/ Anki

35. If applicable, please provide your feedback/comments on Firecracker.

36. If applicable, please provide your feedback/comments on Rapid Review Pathology by Goljan.

37. If applicable, please provide your feedback/comments on SketchyMicro.

38. If applicable, please provide your feedback/comments on SketchyPharm.

39. If applicable, please provide your feedback/comments on Med Bullets.

40. If applicable, please provide your feedback/comments on Online MedEd.

41. If applicable, please provide your feedback/comments on Osmosis.

42. If applicable, please provide your comments on other resources you used to prepare for Step 1.

43. Before starting the dedicated study period, what was your initial goal for Step 1.

44. In order to evaluate effectiveness of preparation strategies to benefit future students, please indicate your actual score on Step 1.

Thank you very much for your time. The MS2s will be quite grateful as are we.

APPENDIX C – RULES FOR CALCULATING SOCIOECONOMIC CATEGORY

SOCIOECONOMIC STATUS QUESTIONS & POINT VALUES

GROUP DESIGNATIONS WITH RANGE OF SCORES:

Maximum points available to be assigned = 85

Group A = 45-85 total points assigned
Group B = 30-40 total points assigned
Group C = 20-25 total points assigned
Group D = 0-15 total points assigned

PARENT'S EDUCATIONAL LEVEL:

Only one set of points, 15 maximum, will be included in the SES calculation for parent's educational level. The points will be assigned to the parent or significant parental role person with the lowest level of educational attainment.

The significant parental role person is considered only if the applicant lived with this person at least 3 years while attending high school. If applicant lived with the significant parental person for at least 3 years while in high school, that person is eligible to be considered in assigning the Parents' Educational Level points in the SES calculation.

Applicant must answer in the following way:

Significant Male:

- Did someone other than your biological father play a significant male parental role in your life? YES
- Did you live with this person for at least three years while attending High School? YES
- If yes, please provide education level.

Significant Female:

- Did someone other than your biological mother play a significant female parental role in your life? YES
- Did you live with this person for at least three years while attending High School? YES
- If yes, please provide education level.

The program will check educational level of father, mother and each eligible significant parental person and assign only one set of SES points based on the lowest level of educational attainment.

Parents Education Level:

Points Assigned	Choices
15	Non-H.S. Graduate
10	High School Graduate/GED
5	Some College
5	Associate's Degree
0	Bachelor Degree
0	College Beyond Bachelor Degree
0	Master's Degree
0	Doctoral Degree
0	Post Doctoral Studies

APPLICANT'S HOMETOWN AREA:

How would you describe the area where you grew up?

Points Assigned	Choices
10	Rural
0	Urban
10	Inner City
0	Suburban
0	Military or Government Installation
0	Other

HOUSEHOLD IN WHICH APPLICANT WAS RAISED OR SPENT MAJORITY OF LIFE FROM BIRTH TO AGE 18:

Size of Household:

Points Assigned	Choices
0	4 or Less
10	5
15	6
20	7 or More

Estimated Value of Residential Property (Owned or Rented):

Points Assigned	Choices
15	0-$50,000
10	$50,001 –$75,000
5	$75,001-$100,000
0	$100,001 - $150,000
0	$150,001 - $200,000
0	Over $200,000

ADDITIONAL SES QUESTIONS:

1. Are you a member of the first generation in your family to apply to, attend or graduate from an undergraduate program?
   a. Yes = 10 points
   b. No = 0 points

2. Were you required to contribute to the overall family income (as opposed to working primarily for your own discretionary spending money) while attending elementary and/or high school?
   a. Yes = 15 points
   b. No = 0 points