

Validity and Reliability of Institutional Assessments in Predicting USMLE STEP 1 Scores: Lessons from a Traditional 2 x 2 Curricular Model

Nitin Puri (✉ purin@marshall.edu)

Joan C Edwards School of Medicine at Marshall University <https://orcid.org/0000-0002-2388-246X>

Sydney Graham-Smith

Joan C Edwards School of Medicine at Marshall University

Michael McCarthy

Joan C Edwards School of Medicine at Marshall University

Bobby Miller

Joan C Edwards School of Medicine at Marshall University

Research article

Keywords: USMLE, STEP1 scores, institutional assessments, Traditional 2 x 2 Curricular Model

Posted Date: August 20th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-56064/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Purpose: We have observed that students' performance in our PreClerkship curriculum does not align well with their USMLE STEP1 scores. Students at-risk of failing or underperforming on STEP1 have often excelled in our institutional assessments. We sought to test the validity and reliability of our course-assessments in predicting STEP1 scores, and in the process generate and validate a more accurate prediction model for STEP1 performance.

Methods: Student pre-matriculation and course assessment data of the Class of 2020 (n = 76) is used to generate a stepwise STEP1 prediction model, which is tested with the students of the Class of 2021 (n = 71). Predictions are generated for the end of each course in the programming language R. For the Class of 2021, predicted STEP1 score is correlated with their actual STEP1 scores and data-agreement is tested with means-difference plots. A similar model is generated and tested for the Class of 2022.

Results: STEP1 predictions based pre-matriculation data are unreliable and fail to identify at-risk students ($R^2 = 0.02$). STEP1 predictions for most year 1 courses (anatomy, biochemistry, physiology) correlate poorly with students' actual STEP1 scores ($R^2 = 0.30$). STEP1 predictions improve for year 2 courses (microbiology, pathology and pharmacology), but reliable predictions are based on truly integrated courses with customized NBMEs as comprehensive exams (0.66). Predictions based on these NBMEs and integrated courses are reproducible for the Class of 2022.

Conclusion: MCAT and undergraduate GPA are poor predictors of students' STEP1 scores. Partially integrated courses with biweekly assessments do not promote problem-solving skills and leave students' at-risk of failing STEP1. Only truly integrated courses with comprehensive assessments are reliable indicators of students STEP1 preparation.

Introduction: -

USMLE STEP 1 is the first of four licensure examinations a medical student must pass in order to become a practicing allopathic medical doctor in the United States. The examination assesses a student's ability to apply basic science concepts to principles and mechanisms associated with the structure, function, disease, and therapeutics of the organ systems. Other uses of USMLE STEP 1 include: (1) a passing score is often a promotion requirement from the PreClerkship to the Clerkship curriculum; and (2) frequently used as screening tool for competitive residency placements(Dillon et al. 2011), (Sutton et al. 2014).

This manuscript will not provide commentary on the utility, effectiveness, or fairness of the USMLE STEP 1 examination in determining, in large part, the career path of a medical student. However, until the point that the design and/or use of USMLE STEP 1 is changed, it remains one of the key indicators of the academic progression of a medical student. In light of the foregoing, success on this standardized examination is of paramount importance to medical students and the medical institution with significant individual and institutional effort spent on preparing for this examination.

The information required for USMLE STEP 1 is primarily acquired during the first two years of the PreClerkship phase of medical education. Joan C. Edwards School of Medicine (JCESOM) at the Marshall University utilizes the traditional 2 × 2 curricular model to instruct medical students. The MS1 curriculum is centered on pedagogy in anatomy, physiology, and biochemistry. The MS2 curriculum is centered on pedagogy in microbiology, pathology of disease and modes of therapy. Medical students' understanding of this information is evaluated using institutional assessments created by the instructors specific to each discipline. Medical students' performances on institutional assessments are included in prediction modeling, which should be an effective tool used to chart the academic progress of individual students, including USMLE STEP 1 performance. This is especially useful in identifying medical students struggling early in the PreClerkship curriculum in order to intervene in a timely fashion.

However, one challenge JCESOM has encountered is accurately predicting a medical student's performance on USMLE STEP 1, despite utilizing prediction modeling (Gullo et al. 2015). For example, students who excel in the first two years of PreClerkship education and are predicted to pass USMLE STEP 1 fail or underperform, and vice versa. Consequently, medical students are unable to reflect on their respective strengths and weaknesses in order to better prepare for USMLE STEP 1, and the medical institution is unable to focus resources or early interventions on students at risk for failing USMLE STEP 1. To address the discrepancies between performance on institutional performance and USMLE STEP 1, we performed a study assessing the validity of institutional assessments in identifying future performances of the students on the USMLE STEP 1 examination.

Our results show that matriculation data is not predictive of a students' USMLE STEP1 performance. Using progressive, stepwise prediction modeling, we find that most MS1 course assessments show only weak correlation with USMLE STEP1 scores. This correlation improves for MS2 courses, but this is not uniformly true. In fact, the most reliable model is based on assessments of only one course i.e. three examinations. The model accuracy is also confirmed in prospective data analysis of students' taking STEP1 in the next academic year.

Material And Methods: -

All student data in the study is deidentified prior to use. USMLE STEP1 predictions for the Class of 2021 are based on data from the Class of 2020 (n = 76). STEP1 predictions for the Class of 2022 are based on data from the class of 2021 (n = 72). We use this "yearly" prediction modeling due to incremental changes in the MD curriculum each year. Especially, in the type and placement of assessments in the PreClerkship Curriculum. Additionally, pre-matriculation data shifted in 2015 with introduction of the new three-digit MCAT score. Class of 2020 is the first cohort with three-digit MCAT for matriculating students. Pre-admission data is extracted from the American

Medical College Application Service (AMCAS) database for students in the Class of 2020 who had subsequently taken United States Medical Licensing Examination (USMLE) Step 1 (n = 76). Only objective, academic data points are considered for generating prediction models.

This same cohort is followed through all MS1 and MS2 courses and their course-exams results are tabulated along with their USMLE STEP1 scores. For generating STEP1 score-predictions for the Class of 2021, aforementioned data for the Class of 2020 (training set) is subjected to linear regression modeling at each milestone with the USMLE STEP1 as the outcome. Models are computed in the programming language R. STEP1 prediction model at each milestone is follows (E = exam; all exams are institutionally developed unless noted otherwise) –

1. Pre-matriculation: Predicted STEP1 score (PSS) = $51.3064 + (0.5342 * \text{First MCAT}) + (1.2277 * \text{BCPM GPA})$
2. End of course 1, MS1 (Elements of Medicine): PSS = $95.90538 + (0.68099 * E1) + (0.38661 * E2) + (0.08675 * E3) + (0.35345 * E4)$
3. End of course 2, MS1 (Structure and Function 1): PSS = $93.5817 + (0.334 * E1) + (0.8754 * E2) + (0.2176 * E3)$
4. End of course 3, MS1 (Structure and Function 2): PSS = $72.5436 + (0.884 * E1) + (0.2817 * E2) + (0.5207 * E3)$
5. End of course 4, MS1 (Structure and Function 3): PSS = $112.490246 + (1.077676 * E1) + (-0.007033 * E2) + (0.268371 * E3)$
6. End of course 5, MS1 (Structure and Function 4): PSS = $123.4593 + (0.6358 * E1) + (0.5348 * E2)$
7. End of course 1, MS2 (Principles of Disease): PSS = $64.7813 + (1.0024 * E1) + (0.2859 * E2) + (0.5547 * E3)$
8. End of course 2, MS2 (Disease and Therapeutics 1): PSS = $56.7937 + (0.8985 * E1) + (1.0028 * E2)$
9. End of course 3, MS2 (Disease and Therapeutics 2): PSS = $42.9588 + (1.2228 * E1) + (0.8839 * E2)$
10. End of course 4, MS2 (Disease and Therapeutics 3): PSS = $90.9247 + (0.5872 * E1) + (-0.009423 * E2) + (1.0778 * \text{NBME score})$
11. End of course 5, MS2 (Disease and Therapeutics 4): PSS = $74.2407 + (0.6616 * E1) + (1.0772 * E2)$

For the Class of 2021 (n = 72), predicted STEP1 score (from the regression model above) generated at the end of each prediction-point (course end) is plotted against the actual STEP1 score of the student. Means-difference plots (Bland Altman plots) show the agreement (Bland & Altman 2016) between the predicted and actual STEP1 scores at each measured point for the students.

For the Class of 2021, curricular modification included inclusion of course-end customized NBMEs for the MS2 courses. Correlation and regression analysis of student performance on these NBMEs and STEP1 is also shown.

STEP1 prediction model for the Class of 2022 (n = 53) is based on exam and STEP1 scores of the Class of 2021. The model builds in a stepwise fashion and generates new prediction at the end of each

milestone. Individual course exams are included or rejected from the prediction model based on change in adjusted R^2 and the standard error of prediction. The model is used only when the adjusted R^2 is greater than 0.5 and is presented here—

1. Predicted STEP1 score at the end of PoD—**Prediction 1**— $(((SF4Exam1)*0.8884)+((POD2)*1.2103) + 45.2093)$; SE = 20.2518 ($R^2 = 53.48$)
2. Predicted STEP1 score at the end of DT1—**Prediction 2**— $(((SF4Exam1)*1.0502)+((DT1NBME)*1.2860) + 30.8867)$; SE = 19.9711 ($R^2 = 57.65$)
3. Predicted STEP1 score at the end of DT3—**Prediction 3**— $((DT3Exam1)*0.9499+(DT3NBME)*0.9078 + 78.8965)$; SE = 12.14 ($R^2 = 66.88$)

For the Class of 2022, predicted STEP1 score (from the regression model above) generated at the end of each “prediction-point” is plotted against the actual STEP1 score of the student.

This study (IRB # 1630008-1) has been approved by the Marshall University Institutional Review Board under the exempt approval status.

Results: -

Pre-matriculation data does not accurately predict STEP1 score—

As shown in **Fig. 1**, STEP1 score predictions generated by linear modeling three-digit MCAT scores and the undergraduate BCPM GPA of the candidates do not correlate well with the actual STEP1 scores of the candidates. The mean-difference plot highlights that MCAT/BCPM overpredict STEP1 scores for students scoring below the school mean (220) and underpredicts for students scoring above.

STEP1 predictions based on MS1 course exams have weak correlation with STEP1 (Fig. 2 A-E and 3)—

Similar to MCAT/BCPM GPA, STEP1 predictions are generated at the end of each MS1 course. The predicted STEP1 scores are correlated with actual STEP1 scores for the class of 2021—

1. Elements of Medicine (EoM) course (**Fig. 2A**): the first course of the MD curriculum is primarily focused on the pedagogy of biochemistry, cell biology, and genetics ($R^2 = 0.2457$).
2. Structure and Function 1 (SF1) course (**Fig. 2B**): the second course of the MD curriculum is primarily focused on the pedagogy of the musculoskeletal system and is the first course of the curriculum with anatomy labs ($R^2 = 0.1366$). It is notable that this correlation is weaker than at the end of the first course of the MD curriculum.
3. Structure and Function 2 (SF2) course (**Fig. 2C**): the third course of the MD curriculum is primarily focused on the pedagogy of anatomy and physiology of the nervous system ($R^2 = 0.2216$).

4. Structure and Function 3 (SF3) course (**Fig. 2D**): the fourth course of the MD curriculum is primarily focused on the pedagogy of anatomy and physiology of cardiovascular, renal and respiratory systems ($R^2 = 0.3755$).
5. Structure and Function 4 (SF4) course (**Fig. 2E**): the fifth and final course of the MS1, MD curriculum is primarily focused on the pedagogy of anatomy and physiology of gastrointestinal, genitourinary and endocrine systems ($R^2 = 0.4289$).
6. **Fig.3** shows predicted STEP1 score for all MS1 exams vs actual STEP1 scores for the Class of 2021. Combined mean-difference and correlation plot highlights that at the end of the MS1 curriculum STEP1 prediction models overpredict scores for most students scoring less than school average on STEP1 (220). MS1 exams combined underpredict scores for students performing better than the school average on USMLE STEP1.

Regression and Means-Difference Plots of MS2 course STEP1 predictions with actual STEP1 scores (Fig. 4A-E and S1A-E)–

As before, STEP1 predictions are generated at the end of each MS2 course and correlated with actual STEP1 scores for the Class of 2021. Prediction modeling is based on course exam and STEP1 scores of the Class of 2020. Means-difference (Bland Altman) plots are constructed at each prediction point–

1. Principles of Disease (PoD) course (**Fig. S1A, 4A**): the first course of the MS2 MD curriculum is primarily focused on the pedagogy of microbiology, immunology, and general pathology ($R^2 = 0.4386$). Means-difference plot shows overpredictions for students scoring below the national mean on STEP1 (≈ 231).
2. Disease and Therapeutics 1 (DT1) course (**Fig. S1B, 4B**): the second course of the MS2 MD curriculum is primarily focused on the pedagogy of the diseases of the musculoskeletal and the hematopoietic systems ($R^2 = 0.3820$). As before, means-difference plot shows overpredictions for students scoring below the national mean on STEP1 (≈ 231).
3. Disease and Therapeutics 2 (DT2) course (**Fig. S1C, 4C**): the third course of the MS2 MD curriculum is primarily focused on the pedagogy of the diseases of the nervous system and behavioral health ($R^2 = 0.3741$). Means-difference plot shows mostly overpredicted scores for all range of STEP1 scores.
4. Disease and Therapeutics 3 (DT3) course (**Fig. S1D, 4D**): the penultimate course of the MS2 MD curriculum is primarily focused on the pedagogy of the diseases of the cardiovascular, renal and respiratory systems ($R^2 = 0.6631$). Means-difference plot is more evenly spread out with smaller ranges of under and overpredictions for all STEP1 scores.
5. Disease and Therapeutics 4 (DT4) course (**Fig. S1E, 4E**): the final course of the MS2 MD curriculum is primarily focused on the pedagogy of the diseases of the genitourinary, gastrointestinal and

endocrine systems ($R^2 = 0.5917$). Means-difference plot for most students is more narrowly spread but shows overpredictions for students performing poorly on STEP1.

Correlation of customized NBMEs in MS2 courses with STEP1 scores (Fig. 5A-D)–

For the Class of 2021, customized NBMEs were introduced as course-end comprehensive assessments for all Disease and Therapeutics courses for the MS2 curriculum. Score performance on each of those exams is correlated against students' score, as seen in **Fig. 5 A-D**. For the class of 2020, only DT3 course included customized NBME at the end of the course. DT3 customized NBME has the strongest linear relationship with students' STEP1 score ($R^2 = 0.6051$).

Prospective analysis for the Class of 2022 (Fig. 6A-C)–

Stepwise, progressive model is used to predict STEP1 scores for the Class of 2022. Regression analysis for each prediction vs students' STEP1 score is shown in Fig. 6.

Discussion: -

In this study, we assessed the validity of institutional assessments in identifying future performances of medical students on the USMLE STEP1 examination. Using linear regression and Bland-Altman analysis, we were able to quantify the ability of our PreClerkship assessments in predicting USMLE STEP 1 scores. Our results indicate that the majority of year one and two institutional assessments do not accurately predict students' performance on the USMLE STEP 1. Specifically, students' scores on PreClerkship assessments lead to an overpredicted USMLE STEP 1 score. This is disadvantageous for faculty and administrators, who are unable to accurately identify high-risk students and provide early individualized interventions. Moreover, medical students engaging in self-reflection are unable to self-identify areas of needs and opportunities.

The first key finding of this study is the unreliability of MCAT and GPA in predicting USMLE STEP1 performance. When used alone, either have minimal predictive value for STEP1 performance of our students. Even combined, their correlation with STEP1 is weak at best and nonexistent at worst. Similar observations were made in our earlier studies(Gullo et al. 2015). However, prospective application of prediction modeling in this study shows that traditionally used pre-matriculation data is all but useless in predicting STEP1 scores for the candidates. Value of standardized pre-matriculation assessments has been questioned in recent years(Gauer et al. 2016) and our results highlight the need for a comprehensive, 360° evaluation of candidates for admission into the MD program. One possible reason for lack of STEP1 predictive value of MCAT and undergraduate GPA in our study could be the evolution of these standardized tests themselves. In recent years, STEP1 has moved away from simple recall of facts to more rounded assessment of problem-solving skills in medical knowledge, biostatistics, evidence-based medicine and medical ethics. We should note that this data does not suggest that MCAT or GPA

have no role in predicting performance on standardized tests like STEP1. These parameters appear to have predictive value for the cohort, just not for the individual. The average MCAT for students matriculating in to JCESOM, Marshall University is around the 50th percentile, and so is the average for our USMLE STEP1 scores.

The traditional 2 × 2 curricular model delivers information in a manner that restricts integration of basic sciences with clinical application throughout the first year. Our organ-system based MS1 curriculum is focused on the pedagogy of normal structure and function of the human body while the MS2 curriculum covers principles of disease and therapeutics. This design narrows the scope of assessments of each course, limiting them to only the covered disciplines. One argument in favor of this design is that students have the opportunity to learn each organ system twice. But, in practice, it creates informational compartments where MS1 courses cannot adequately assess complex problem-solving skills due to lack of context of the disease process. Effectively, the organ-system based MS1 curriculum is just a different organization of the classical three discipline MS1 curriculum, biochemistry, physiology and anatomy. This setup renders the majority of first year assessments to memorization and recall of facts, ranked lower on the Bloom's cognitive scale(Adams 2015). This style of assessment fails to utilize the high levels of Bloom's taxonomy, application and analysis, which are the mainstays of USMLE STEP 1 questions(Weissbart et al. 2015).

Unsurprisingly, the second key finding of our study is that performance on most PreClerkship course-assessments poorly correlate with student's actual USMLE STEP 1 scores. Most notably, Bland-Altman analysis reveals significant over-predictions for students underperforming on STEP1. Additionally, within the first year, significant instructor and discipline variability is readily apparent. Internal assessments with low variance lead to predictions that fail to accurately identify high-risk students; therefore, students fail to receive early interventions to prevent failure on USMLE STEP 1. These shortcomings inflate the students' sense of preparedness who often show resistance to academic support—as they are “doing well” in the curriculum.

Theoretically, predictions regarding students' performance on USMLE STEP 1 should be more accurate in the second year of PreClerkship education, because assessments integrate basic science and clinical correlations. However, performances on internal assessments continue to over predict performance on USMLE STEP 1. This is again disadvantageous to students' ability to perform self-directed learning.

The best correlated assessments are from the course DT3, which accounts for 66% variability in students' scores. This course is different in its approach to pedagogy and assessments in our curriculum. It employs more active learning formats, like independent-learnings, team-based leanings and problem-solving exercises. The course is also more integrated than the other MS2 courses. Faculty and instructors spend significant time in the course revisiting foundational concepts of the MS1 curriculum, followed by pathophysiology and pharmacology of the covered systems. DT3 assessments also rank higher on the Blooms scale with greater number of critical-thinking and problem-solving questions. Notably, for the Class of 2020 DT3 was the only course in the PreClerkship MD curriculum to utilize a cumulative NBME

assessment at the end of the course. This assessment utilizes multiple-choice questions that resemble the format of USMLE STEP 1 questions and provides performance feedback to students and faculty. Cumulative assessments may help students retain information for longer, as opposed to “binge and purge” of biweekly assessments of other courses. Additionally, cumulative assessments may aid in improved synthesis and integration of information, as these exams are likely to focus on key concepts rather than minute details. While the latter correlation allows faculty to identify high-risk students more accurately, the timing of intervention is not advantageous. Specifically, students complete this course late into the second year after USMLE STEP 1 preparation has commenced and the time to remediate areas of weaknesses is limited.

Identifying the need for institutional action and need for improved prediction of USMLE STEP1 scores, starting with the Class of 2021, and cumulative, customized NBME assessments are included in all organ-system based courses of the MS2 curriculum. These comprehensive exams better correlate with STEP1 performance than traditional course exams(Brenner et al. 2017). As before, we believe that cumulative assessments aid students’ ability to synthesize and retain information for longer. Additionally, faculty are able to provide early interventions and focus resources on high-risk students and medical students are able to use performance feedback to recognize areas of strengths and weaknesses to guide their preparation for USMLE STEP 1. It is important to note however that not all NBMEs are similarly correlative. Our analysis indicates that this is due to differences in course design and assessments. Course(s) not dedicating enough time to integration of first- and second-year disciplines are not able to choose higher order questions from the NBME question bank. This situation is not dissimilar to institutional assessments i.e. customized NBME assessments may not be superior in of themselves, it’s the type of pedagogy and level of integration in the courses that appears to drive the outcome of reliable predictions of student performance.

The final key finding of this study is the validity and reliability of our STEP1 prediction model generated using targeted assessments, including the course-end NBMEs. As opposed to the retrospective model published by us(Gullo et al. 2015), and others(Coumarbatch et al. 2010) (Giordano et al. 2016), this model is effective prospectively. The model is deployed to identify at-risk students early and provide intervention if needed. Struggling students are identified early in the fall semester of MS2 of the MD curriculum and offered academic remediations, as needed. Reliability and reproducibility of our model is demonstrated with STEP1 scores for the Class of 2022, where multiple students were offered early assistance based on our prediction modeling.

These results go beyond performance on USMLE STEP1. They show us the reproducibility of knowledge acquired and tested during a traditional 2 × 2 MD curriculum. Lack of clinical context in the MS1 curriculum and failure to reconnect to foundations in the MS2 curriculum, render our assessments detail-focused and less reliable. Student often memorize facts in two-week bursts and might fail to understand the interplay and crosstalk of systems network. Critical thinking and problem-solving skills are key to successful practice of medicine. These skills require application of foundational concepts to

clinical problem-solving, which in turn are fostered by integration of concepts across disciplines and systems.

Declaration—

- Ethics approval and consent to participate— This study (IRB # 1630008-1) has been approved by the Marshall University Institutional Review Board under the exempt approval status
- Consent for publication—all authors consent to the publication of the manuscript
- Competing interests—none
- Funding—none
- Authors' contributions—
 - Nitin Puri—design, data collection, data analysis, manuscript writing
 - Sydney Graham-Smith— data analysis and manuscript writing
 - Michael McCarthy—data analysis and manuscript writing
 - Bobby Miller—design, manuscript writing
- Acknowledgements
 - Dr Joseph Shapiro, Dean, Joan C. Edwards School of Medicine, Marshall University
- All data files are available upon request

Bibliography—

1. Adams NE. 2015. Bloom's taxonomy of cognitive learning objectives. *J Med Libr Assoc.* 103(3):152–153.
2. Bland JM, Altman DG. 2016. Measuring agreement in method comparison studies: Statistical Methods in Medical Research [Internet]. [accessed 2020 Jul 4]. <https://journals.sagepub.com/doi/10.1177/096228029900800204>
3. Brenner JM, Bird JB, Willey JM. 2017. Formative Assessment in an Integrated Curriculum: Identifying At-Risk Students for Poor Performance on USMLE Step 1 Using NBME Custom Exam Questions. *Acad Med.* 92(11S Association of American Medical Colleges Learn Serve Lead: Proceedings of the 56th Annual Research in Medical Education Sessions):S21–S25.
4. Coumarbatch J, Robinson L, Thomas R, Bridge PD. 2010. Strategies for identifying students at risk for USMLE step 1 failure. *Fam Med.* 42(2):105–110.
5. Dillon GF, Clauser BE, Melnick DE. 2011. The role of USMLE scores in selecting residents. *Acad Med.* 86(7):793; author reply 794.
6. Gauer JL, Wolff JM, Jackson JB. 2016. Do MCAT scores predict USMLE scores? An analysis on 5 years of medical student data. *Med Educ Online [Internet].* [accessed 2020 Jul 4] 21.

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5045966/>

7. Giordano C, Hutchinson D, Pepler R. 2016. A Predictive Model for USMLE Step 1 Scores. *Cureus*. 8(9):e769.
8. Gullo CA, McCarthy MJ, Shapiro JI, Miller BL. 2015. Predicting Medical Student Success on Licensure Exams. *MedSciEduc*. 25(4):447–453.
9. Sutton E, Richardson JD, Ziegler C, Bond J, Burke-Poole M, McMasters KM. 2014. Is USMLE Step 1 score a valid predictor of success in surgical residency? *Am J Surg*. 208(6):1029–1034; discussion 1034.
10. Weissbart SJ, Stock JA, Wein AJ. 2015. Program directors' criteria for selection into urology residency. *Urology*. 85(4):731–736.

Figures

Figure 1

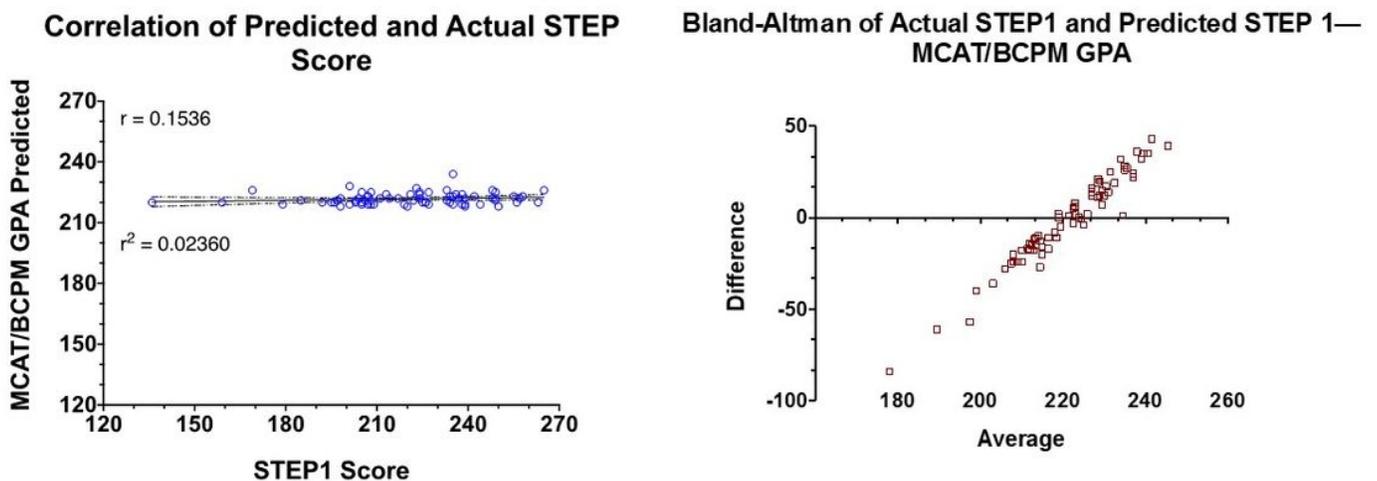


Figure 1

Prediction model generated based on retrospective data of the Class of 2020 is used to predict STEP1 scores for the Class of 2021. A—correlation and regression analysis of actual vs. predict STEP1 scores for the class of 2021 (n = 72). B—Means-difference plot for the same cohort with predicted and actual STEP1 scores.

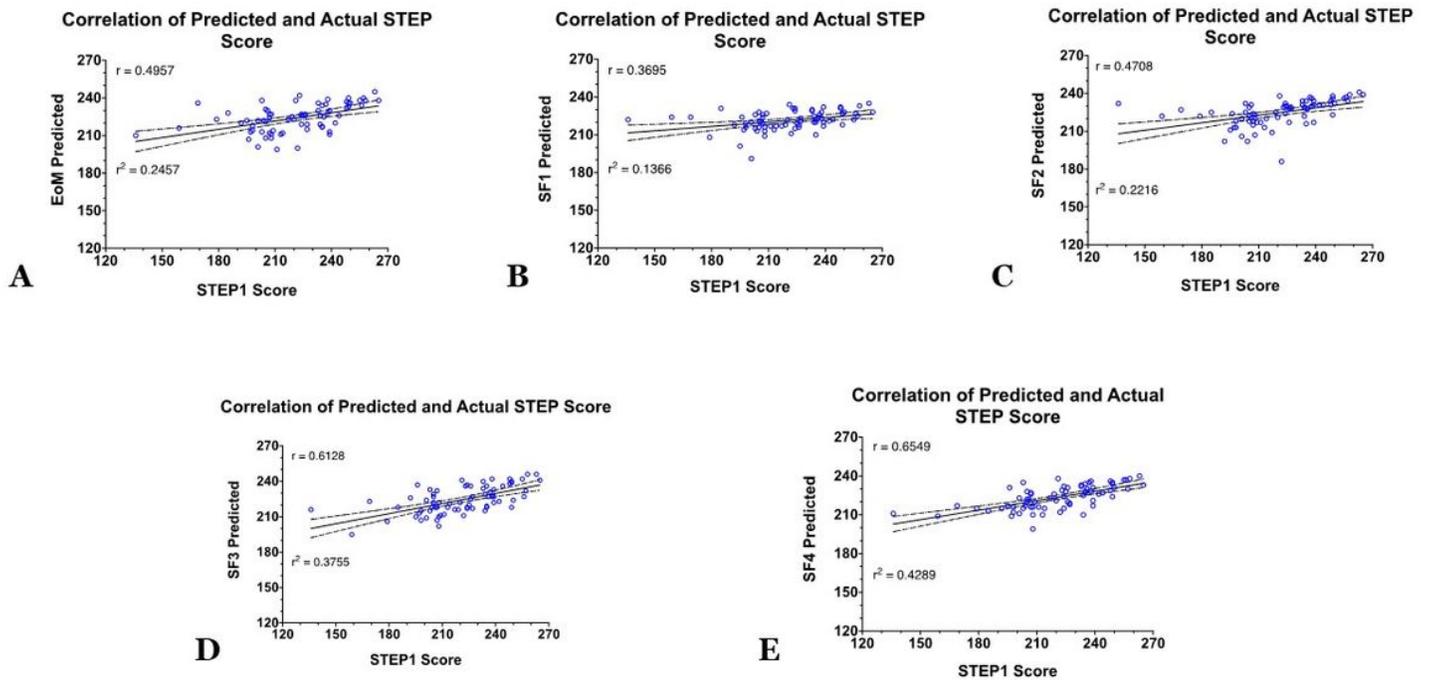


Figure 2

Figure 2

Prediction model generated based on retrospective data of the Class of 2020 is used to predict STEP1 scores for the Class of 2021. Predictions are generated at the end of every MS1 course. A-E— correlation and regression analysis of actual vs. predict STEP1 scores for the class of 2021 (n = 72) at each prediction point.

Bland-Altman of Actual STEP1 and Predicted STEP 1—
ALL MS1 Exams

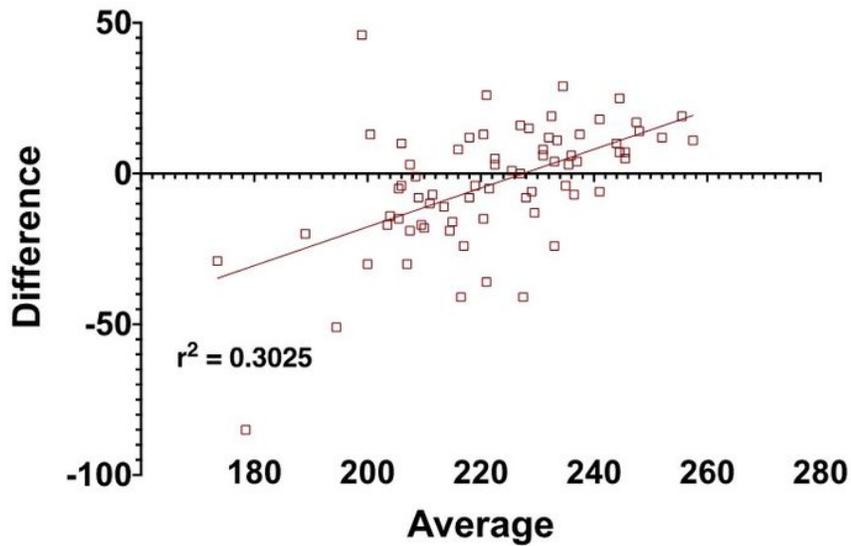


Figure 3

Figure 3

Means-difference analysis of predicted and actual STEP1 scores for the Class of 2021 (n = 72) at the end of MS1, PreClerkship curriculum. Adjusted R2 for the regression analysis is also shown on the graph.

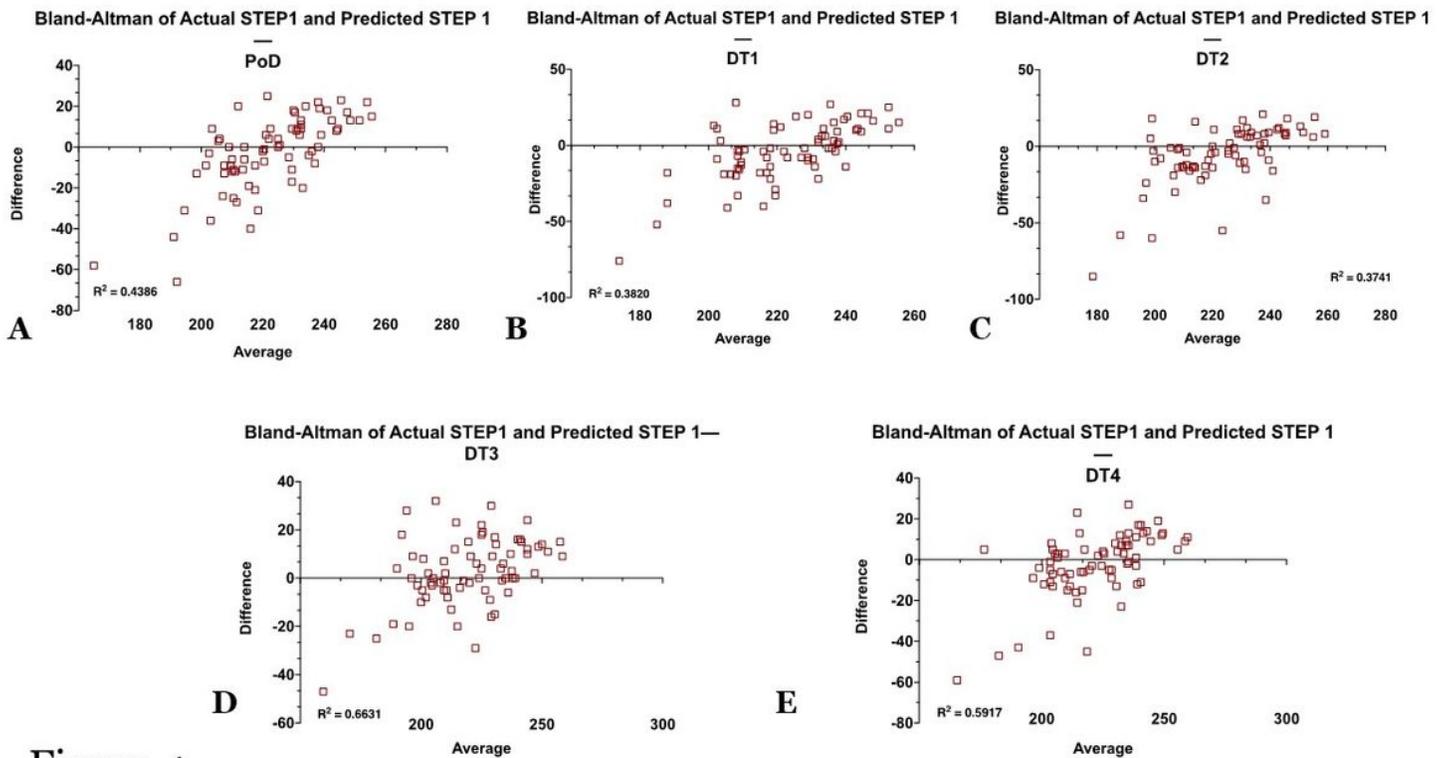


Figure 4

Figure 4

Means-difference analysis of predicted and actual STEP1 scores for the Class of 2021 (n = 72) at the end each MS2 course of the PreClerkship curriculum (A-E). Adjusted R2 for the regression analysis is also shown on the graphs.

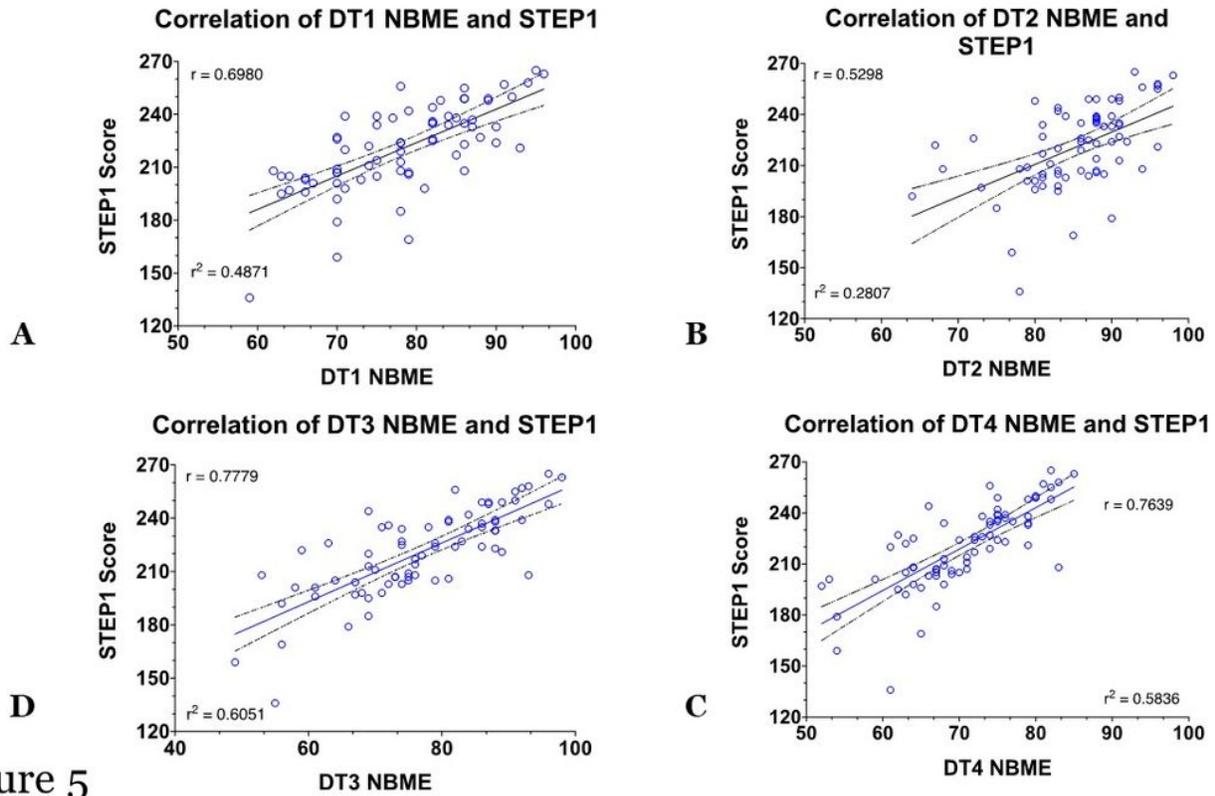


Figure 5

Figure 5

Correlation and regression analysis of the students' NBME scores and STEP1 scores for the Class of 2021 (n = 72).

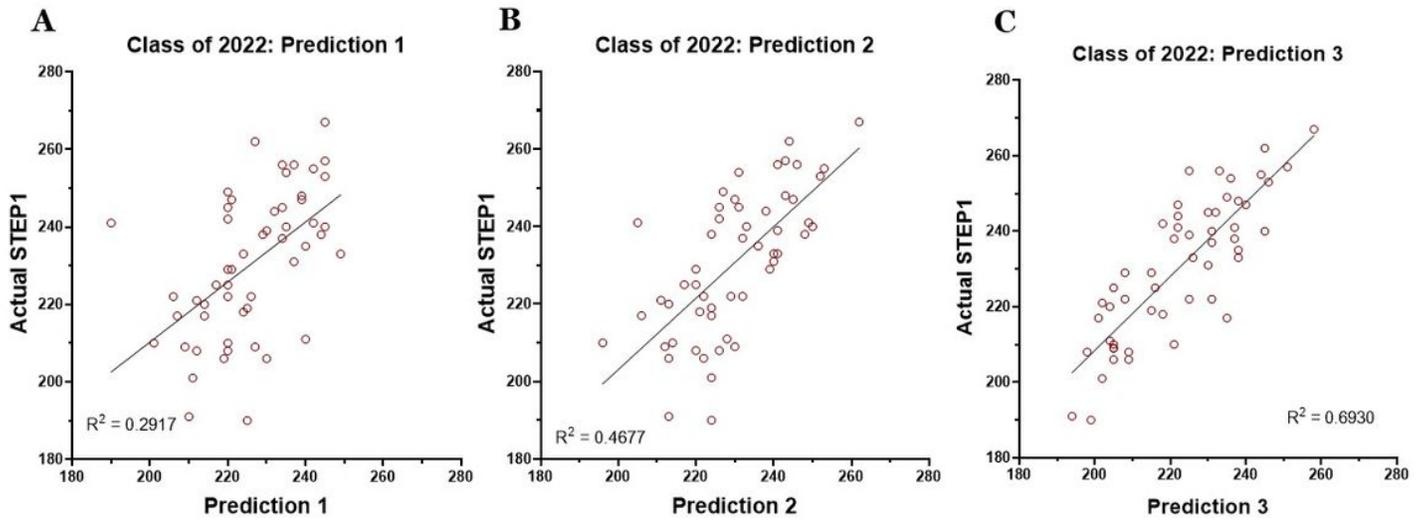


Figure 6

Figure 6

Regression analysis for the STEP1 scores and the three predictions generated for the students of the Class of 2022 (n = 53).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [FigureS1.jpg](#)