

A Generalizable Approach to Predicting Performance on USMLE Step 2 CK

Jeffrey B Bird ¹, Doreen M Olvet ¹, Joanne M Willey ¹, Judith M Brenner ^{1,2}

¹Department of Science Education, Donald and Barbara Zucker School of Medicine at Hofstra/Northwell, Hempstead, NY, 11549, USA; ²Department of Medicine, Donald and Barbara Zucker School of Medicine at Hofstra/Northwell, Hempstead, NY, 11549, USA

Correspondence: Judith M Brenner, Donald and Barbara Zucker School of Medicine at Hofstra/Northwell, 500 Hofstra University, Hempstead, NY, 11549-1000, USA, Tel +1 516 463-7590, Email Judith.M.Brenner@hofstra.edu

Introduction: The elimination of the USMLE Step 1 three-digit score has created a deficit in standardized performance metrics for undergraduate medical educators and residency program directors. It is likely that there will be greater emphasis on USMLE Step 2 CK, an exam found to be associated with later clinical performance in residents and physicians. Because many previous models relied on Step 1 scores to predict student performance on Step 2 CK, we developed a model using other metrics.

Materials and Methods: Assessment data for 228 students in three cohorts (classes of 2018, 2019, and 2020) were collected, including the Medical College Admission Test (MCAT), NBME Customized Assessment Service (CAS) exams and NBME Subject exams. A linear regression model was conducted to predict Step 2 CK scores at five time-points: at the end of years one and two and at three trimester intervals in year three. An additional cohort (class of 2021) was used to validate the model.

Results: Significant models were found at 5 time-points in the curriculum and increased in predictability as students progressed: end of year 1 (adj $R^2 = 0.29$), end of year 2 (adj $R^2 = 0.34$), clerkship trimester 1 (adj $R^2 = 0.52$), clerkship trimester 2 (adj $R^2 = 0.58$), clerkship trimester 3 (adj $R^2 = 0.62$). Including Step 1 scores did not significantly improve the final model. Using metrics from the class of 2021, the model predicted Step 2 CK performance within a mean square error (MSE) of 8.3 points (SD = 6.8) at the end of year 1 increasing predictability incrementally to within a mean of 5.4 points (SD = 4.1) by the end of year 3.

Conclusion: This model is highly generalizable and enables medical educators to predict student performance on Step 2 CK in the absence of Step 1 quantitative data as early as the end of the first year of medical education with increasingly stronger predictions as students progressed through the clerkship year.

Keywords: assessment, prediction model, career advising, match, NBME

Introduction

The National Resident Matching Program (NRMP), or The Match, is a multistep process in which soon-to-be medical school graduates provide numerous pieces of information to residency programs, hoping to be invited for an interview and ultimately being favorably ranked, leading the way to a residency spot. Candidates provide personal statements, transcripts, medical student performance evaluations (MSPE), CVs, an Electronic Residency Application Service (ERAS) application, letters of recommendation, and of course, USMLE Step scores. To facilitate a successful match among its graduates, medical schools have established processes to advise medical students with the goal of guiding them to apply to programs for which they will successfully compete.

Despite its designation as a licensure exam, the three-digit USMLE Step 1 score has been among the most important components of the application that residency program directors consider. In the most recent NRMP Program Director's Survey,¹ USMLE Step 1 was most frequently (86%) cited as a factor in selecting applicants to interview. This represents a 13% increase since 2010.² For program directors seeing increasing numbers of applicants, the three-digit score represented a national quantitative metric that rose to primal importance. Thus, the elimination of the USMLE Step 1 three-digit score in January 2022 placed undergraduate medical educators and program directors at a deficit in their advisement and applicant selection roles, respectively. Nonetheless, many support this decision by the National Board of

Medical Examiners (NBME)³ as medical educators have witnessed students increasingly focused on this exam to the detriment of classroom engagement, personal well-being, as well as its impact on student finances as they purchase an assortment of study products.⁴

With the elimination of the three-digit USMLE Step 1 score, it is likely that residency program directors will increasingly turn their attention to scores on USMLE Step 2 Clinical Knowledge (CK) exam as a standardized metric of students' proficiency. Among programs in the 2021 NRMP survey (when three-digit Step 1 scores were still available), 79% of program directors report using the Step 2 CK when considering applicants to interview and it was listed as third highest in frequency when ranking students.¹ In a recent study by our group, 90% of MSPE end-users agreed or strongly agreed that USMLE Step 2 CK will become more important for applicants lacking three-digit Step 1 scores.⁵ Turning to USMLE Step 2 CK is not without merit. The Step 2 CK has been found to be validated in its clinical relevance.⁶ Sharma et al⁷ described USMLE Step 2 CK as the strongest individual predictor of performance across all residency performance domains measured in an Internal Medicine residency program. Patient outcomes have also been linked to Step 2 CK where Norcini et al⁸ found that higher Step 2 CK performance was associated with a decrease in patient risk for mortality. The predictive validity of Step 2 CK is not without debate, however. McGaghie et al⁹ reviewed nine studies that included almost 400 students and found Step 1 and Step 2 CK scores failed to correlate with reliable measures of resident and fellow performance.

Medical educators will be met with a new challenge as they begin to contemplate how best to accommodate students who must prepare to sit for Step 2 CK at the end of their clerkship year, as there is little data to guide decision-making in this new learning environment. To appropriately advise medical students, medical educators need to develop methods that accurately predict USMLE Step 2 CK scores, ideally early enough in medical school to be meaningful. For some students, this prediction may mean adjusting study patterns or it may be the first indication that they are unlikely to succeed when applying to a highly competitive specialty. Previous models developed to predict USMLE Step 2 CK relied on internal measures, the Medical College Admission Test (MCAT) and the USMLE Step 1 score^{10,11}. Without the USMLE Step 1 as a prediction metric, these models have been rendered obsolete. We therefore developed a model to predict student performance on USMLE Step 2 CK using metrics other than USMLE Step 1 score. These data are available to any medical school beginning as early as the end of the first year.

Materials and Methods

This study was conducted at the Donald and Barbara Zucker School of Medicine at Hofstra/Northwell (ZSOM). The ZSOM employs an integrated preclinical curriculum in which students are assessed by a combination of summative essay exams and formative NBME exams as described below. The clerkship years are assessed using a combination of NBME subject exams, performance on written and oral assignments, and preceptor input. To increase the generalizability of our model, we used the standardized elements of our assessment portfolio in its development. Data from medical students who provided informed consent to use their assessment data are included in this analysis. This study was deemed exempt from review by the Hofstra University Institutional Review Board.

In addition to the MCAT, two types of nationally standardized assessments administered post-matriculation were used in the analysis: NBME Customized Assessment Service (CAS) exams and NBME Subject exams. The construction of CAS exams is a service provided by the NBME, which provides medical school faculty with a database of retired USMLE Step 1 questions from which faculty can select and use for end of course exams. At the ZSOM, faculty typically select 60–70 questions from this database that align with course content. NBME CAS exams were administered in a secure, proctored environment at the end of each of seven integrated preclinical courses over the first two years of the educational program.¹² During the clerkship year, a total of six NBME subject exams were administered as components of assessment for each clerkship, two at the end of each trimester; each trimester included two clerkships. Students completed the NBME CAS exam for formative purposes and the NBME Subject exams were summative assessments.

Assessment data for three cohorts of students (classes of 2018, 2019, and 2020) were examined retrospectively, including MCAT scores, scores from customized NBME CAS exams during the preclinical years, and NBME Subject exams during the clerkship year. Exam scores were transformed to a z-score, and a cumulative average score was calculated at matriculation and at five different time-points in the curriculum: end of year 1, end of year 2, after clerkship

trimester 1, after clerkship trimester 2, and at the end of year 3. A regression model was conducted to predict USMLE Step 2 CK performance at each of the five points. We ran these models both with and without the USMLE Step 1 score to determine the impact that removing the Step 1 score would have in our model. The data analysis for this paper was generated using SAS University Edition Software. Copyright © [2016] SAS Institute Inc. SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc., Cary, NC, USA.

Results

Across the three cohorts, there were complete data for 228 students (97%). The mean MCAT percentile across the three cohorts was 90.0 (SD = 9.9), and the mean score for the USMLE Step 1 and USMLE Step 2 CK was 239.1 (SD = 14.0) and 252.5 (SD = 13.1), respectively. Table 1 and Figure 1 show the model statistics at each time-point for models with and without Step 1 score. In our model, the MCAT was found not to be a significant predictor of performance ($P = 0.72$) and was removed from the analysis.

Significant models were found at each of the 5 post-matriculation points in the curriculum and increased in predictability as students progressed (Table 1 and Figure 1); end of year 1 (adj $R^2 = 0.29$), end of year 2 (adj $R^2 = 0.34$), clerkship trimester 1 (adj $R^2 = 0.52$), clerkship trimester 2 (adj $R^2 = 0.58$), end of year 3 (adj $R^2 = 0.62$). The power to predict USMLE Step 2 CK scores effectively doubled from end of year 1 to end of year 3, with the greatest single-period increase end of year 2 to the early third year.

In comparison, models that included Step 1 scores showed a higher level of predictability at the end of year 2 (adj $R^2 = 0.49$ compared to adj $R^2 = 0.34$ without Step 1) once students had completed all preclinical exams and Step 1. The gap

Table 1 Results of Multiple Regression Models Used to Predict Student Performance on the USMLE Step 2 CK Exam (N = 228) by Medical Students with and without Using USMLE Step 1 as a Predictor

Predictor Variables	Model without Step 1			Model with Step 1		
	Regression Coefficient b (95% CI)	p-value	Overall Model Adjusted R^2	Regression Coefficient b (95% CI)	p-value	Overall Model Adjusted R^2
End of Year 1						
NBME CAS (4 exams)	9.22 (7.34–11.09)	< 0.0001	0.29	-	-	-
End of Year 2						
NBME CAS (7 exams)	10.44 (8.56–12.32)	< 0.0001	0.34	3.45 (1.29–6.00)	< 0.005	0.49
USMLE Step 1 Score	-	-	-	0.50 (0.38–0.63)	< 0.0001	
Clerkship Trimester 1						
NBME CAS (7 exams)	5.68 (3.76–7.60)	< 0.0001	0.52	2.52 (0.32–4.72)	< 0.05	0.56
NBME Subject Exam (2 exams)	6.48 (5.01–7.90)	< 0.0001		4.82 (3.33–6.31)	< 0.0001	
USMLE Step 1 Score	-	-	-	0.32 (0.20–0.45)	< 0.0001	
Clerkship Trimester 2^a						
NBME CAS (7 exams)	3.81 (1.93–5.69)	< 0.0001	0.58	-	-	0.60
NBME Subject Exam (4 exams)	8.01 (6.68–9.45)	< 0.0001		6.86 (5.34–8.38)	< 0.0001	
USMLE Step 1 Score	-	-	-	0.29 (0.19–0.40)	< 0.0001	
Clerkship Trimester 3^a						
NBME CAS (7 exams)	2.76 (0.90–4.62)	< 0.005	0.62	-	-	0.63
NBME Subject Exam (6 exams)	8.88 (7.51–10.25)	< 0.0001		7.80 (6.27–9.32)	< 0.0001	
USMLE Step 1 Score	-	-	-	0.23 (0.12–0.34)	< 0.0001	

Note: ^aNBME CAS was not significant and was removed from the USMLE Step 1 Score model.

Abbreviation: NBME CAS, National Board of Medical Education Customized Assessment Service.

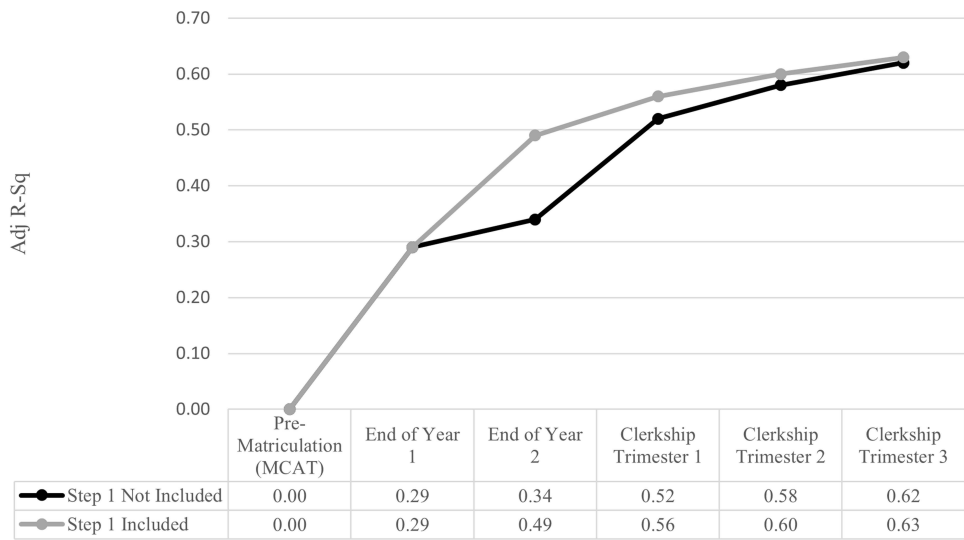


Figure 1 Cumulative predictive value of MCAT, NBME CAS, and NBME Subject Exams with and without Step 1 in determining USMLE Step 2 CK Scores.
Abbreviations: NBME CAS indicates National Board of Medical Education Customized Assessment Service; MCAT, Medical College Admission Test; USMLE, United States Medical Licensing Examination.

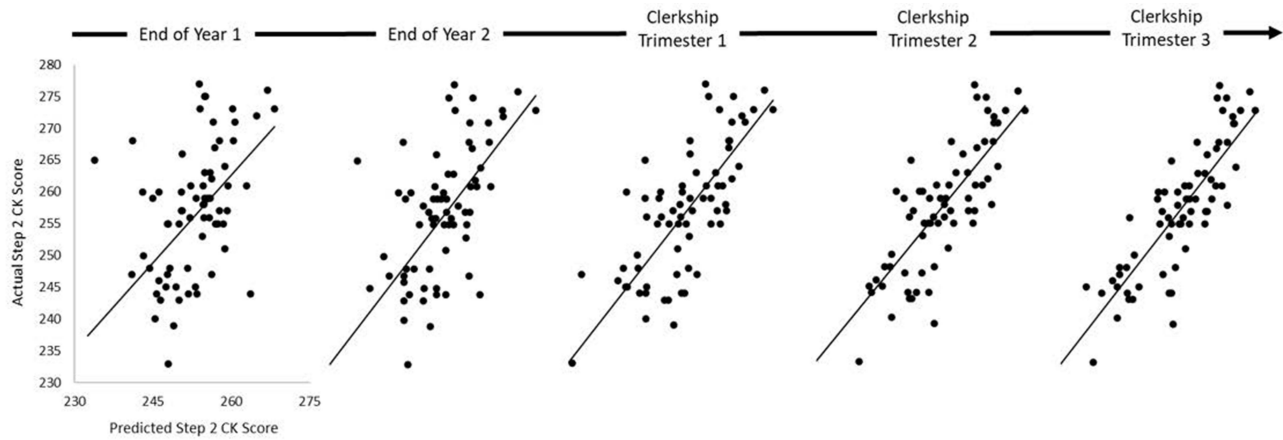


Figure 2 Association between actual and predicted USMLE Step 2 CK scores at five points in the curriculum. Predicted values were determined using models that excluded USMLE Step 1.

in predictability, however, was greatly narrowed by the end of clerkship trimester 1 ($\text{adj } R^2 = 0.56$), clerkship trimester 2 ($\text{adj } R^2 = 0.60$), and virtually identical by the end of year 3 ($\text{adj } R^2 = 0.63$).

To validate our findings, assessment data from an additional cohort (class of 2021) were analyzed using the models that excluded Step 1. Step 2 CK scores were predicted for students who had full assessment data and had taken the Step 2 CK exam ($N = 72$). The mean squared error (MSE) was calculated between the predicted and actual scores for each of the 5 post-matriculation points in the curriculum (Figure 2). Results revealed that the models were able to predict within an MSE of 8.3 points ($SD = 6.8$) at the end of year 1 increasing in predictability incrementally to within an MSE of 5.4 points ($SD = 4.1$) by the end of the third year. The lower the mean and deviation, the stronger the prediction.

Discussion

Now that the USMLE Step 1 no longer provides a three-digit score, performance on USMLE Step 2 CK is poised to become more important for program directors to evaluate applicants for residency positions.⁵ Here, we present a predictive model for performance on USMLE Step 2 CK that does not rely on USMLE Step 1 quantitative scores.

The elegance of our model is its simplicity. Using metrics easily accessible to any medical school, this model demonstrated that predictions as early as the end of year one can be made by using NBME CAS exams, and increasingly stronger predictions are possible as students take NBME Subject exams throughout the clerkship year. Although Step 1 scores remain a significant predictor at the end of year 2, the gap in predictability is mitigated by clerkship trimester 1 and eliminated by the end of year 3.

The validation of the model using an additional cohort revealed the strengths and limitations when predicting scores for future students. It is important to understand the accuracy of the model when academic advisors are identifying learners at risk. At the end of year 1 the MSE is the highest, so advisors may choose to select a larger number of potentially at-risk students as compared to the end of the third year, when the model is the most accurate and the focus can be narrowed.

The paramount role of educators is to ensure the success of their learners. Given the potentially high stake nature of USMLE Step 2 CK, creating a model to predict the score on USMLE Step 2 CK independent of the USMLE Step 1 score is of practical importance. Performance on Step 2 CK has implications beyond undergraduate medical education. Although the data is limited, USMLE Step 2 CK questions were not only found to be valid in terms of their relevance to clinical practice,⁶ but scores on this exam can predict performance during residency⁷ and even patient mortality risk.⁸ Our predictive model should help educators identify students in need of targeted tutoring early in their medical education, which could have an impact on patient care throughout their careers. This model will also assist career advisors in having more informed conversations with their advisees.

A concern among medical educators is the impact on student learning in the preclinical years now that the three-digit score on USMLE Step 1 has been eliminated. Medical schools may change their preclinical curricula as they may feel less constrained to “teach to the test”. In addition, while USMLE Step 1 is currently a key driver in learning, it is unknown if this change in scoring will alter student motivation. Students may view the importance of basic science curricula less important and change their approach to studying in their preclinical years. Our model demonstrates that early learning, captured by NBME CAS exam scores, is a significant factor in predicting performance on measures taken towards the end of medical school (ie, USMLE Step 2 CK). Understanding this association should help schools as they grapple with curricular changes and issues of learner motivation. This should help educators in messaging the importance of foundational, basic science to their student bodies. Performance in the preclinical years or early years of medical school contributes to future success.

It is noteworthy that MCAT performance had no bearing on the model; rather, the entire model relies on post-matriculation performance. This finding is contrary to what others have reported, showing a weak to medium association between MCAT and USMLE Step 2 CK scores.¹³ However, schools may want to consider the addition of NBME CAS to their assessment toolkit.

This study is limited by the learning environment during 2018 to 2020, in which USMLE Step 1 remained an exam with a three-digit score. Educators do not yet know how students will modify their approach to learning now that USMLE Step 1 has changed to pass/fail and the model may need to be updated when this is better understood. Generalizability may be limited in that not all institutions use NBME CAS exams and questions may differ between institutions that do. However, all questions are validated by the NBME from a limited pool and findings in this study may encourage other institutions to adopt NBME CAS exams. The NBME CAS is also formative in our curriculum, which may impact scores compared to a summative exam. Further exploration of formative vs summative performance as well as analyzing demographic data may provide additional insight.

Undergraduate medical educators may need to develop tools to help predict student performance on USMLE Step 2 CK now that the USMLE Step 1 scoring has changed from three-digit to pass/fail. In the future, it will be important to understand how this model works once the changes of the switch to pass/fail are better understood and to consider other metrics to build the prediction model. It will also be important to understand how schools use this model in early advising of students.

Conclusion

In the absence of the USMLE Step 1 three-digit score, we have found that a model using variables available to any medical school is able to effectively predict performance on the USMLE Step 2 CK exam starting as early as the end of

the first year of medical school with increasingly stronger predictions as students move through their medical education. This has potential implications for advisement and future performance as physicians.

Ethical Approval

This study was deemed exempt by the Hofstra University Internal Review Board.

Acknowledgments

The authors would like to recognize and appreciate the following individuals who helped with the preparation of this manuscript: Krista Paxton and Saori (Wendy) Herman, MLIS, AHIP.

Funding

There is no funding to report.

Disclosure

The authors report no conflicts of interest in this work.

References

1. National Resident Matching Program. Data release and research committee: results of the 2021 NRMP program director survey; 2021. Available from: <https://www.nrmp.org/wp-content/uploads/2021/11/2021-PD-Survey-Report-for-WWW.pdf>. Accessed August 17, 2022.
2. National Resident Matching Program. Data release and research committee: results of the 2010 NRMP program director survey; 2010. Available from: <http://www.nrmp.org/wp-content/uploads/2021/07/programresultsbyspecialty2010v3.pdf>. Accessed August 17, 2022.
3. Prober CG, Kolars JC, First LR, Melnick DE. A plea to reassess the role of United States medical licensing examination step 1 scores in residency selection. *Acad Med*. 2016;91(1):12–15. doi:10.1097/ACM.0000000000000855
4. Chen DR, Priest KC, Batten JN, Frago LE, Reinfeld BI, Laitman BM. Student perspectives on the “step 1 climate” in preclinical medical education. *Acad Med*. 2019;94(3):302–304. doi:10.1097/ACM.0000000000002565
5. Bird JB, Friedman KA, Arayssi T, Olvet DM, Conigliaro RL, Brenner JM. Review of the medical student performance evaluation: analysis of the end-users’ perspective across the specialties. *Med Educ Online*. 2021;26(1):1876315. doi:10.1080/10872981.2021.1876315
6. Cuddy MM, Dillon GF, Clauser BE, et al. Assessing the validity of the USMLE step 2 clinical knowledge examination through an evaluation of its clinical relevance. *Acad Med*. 2004;79(10Suppl):S43–45. doi:10.1097/00001888-200410001-00013
7. Sharma A, Schauer DP, Kelleher M, Kinnear B, Sall D, Warm E. USMLE step 2 CK: best predictor of multimodal performance in an internal medicine residency. *J Grad Med Educ*. 2019;11(4):412–419. doi:10.4300/JGME-D-19-00099.1
8. Norcini JJ, Boulet JR, Opalek A, Dauphinee WD. The relationship between licensing examination performance and the outcomes of care by international medical school graduates. *Acad Med*. 2014;89(8):1157–1162. doi:10.1097/ACM.0000000000000310
9. McGaghie WC, Cohen ER, Wayne DB. Are United States medical licensing exam step 1 and 2 scores valid measures for postgraduate medical residency selection decisions? *Acad Med*. 2011;86(1):48–52. doi:10.1097/ACM.0b013e3181ffacdb
10. Monteiro KA, George P, Dollase R, Dumenco L. Predicting United States medical licensure examination step 2 clinical knowledge scores from previous academic indicators. *Adv Med Educ Pract*. 2017;8:385–391. doi:10.2147/AMEP.S138557
11. Guot HM, Franqui-Rivera H. Predicting performance on the United States medical licensing examination step 1 and step 2 clinical knowledge using results from previous examinations. *Adv Med Educ Pract*. 2018;9:943–949. doi:10.2147/AMEP.S180786
12. Brenner JM, Bird JB, Willey JM. Formative assessment in an integrated curriculum: identifying at-risk students for poor performance on USMLE step 1 using NBME custom exam questions. *Acad Med*. 2017;92:S21–S25. doi:10.1097/ACM.0000000000001914
13. Donnon T, Paolucci EO, Violato C. The predictive validity of the MCAT for medical school performance and medical board licensing examinations: a meta-analysis of the published research. *Acad Med*. 2007;82(1):100–106. doi:10.1097/01.ACM.0000249878.25186.b7

Advances in Medical Education and Practice

Dovepress

Publish your work in this journal

Advances in Medical Education and Practice is an international, peer-reviewed, open access journal that aims to present and publish research on Medical Education covering medical, dental, nursing and allied health care professional education. The journal covers undergraduate education, postgraduate training and continuing medical education including emerging trends and innovative models linking education, research, and health care services. The manuscript management system is completely online and includes a very quick and fair peer-review system. Visit <http://www.dovepress.com/testimonials.php> to read real quotes from published authors.

Submit your manuscript here: <http://www.dovepress.com/advances-in-medical-education-and-practice-journal>