The Impact of Digital Medical Resources on USMLE Step 2 CK Scores—A Retrospective Study of 1,985 US Medical Students

Eileen Lee, MA, Lead Researcher at AMBOSS & Consultant to the World Bank
Omar Cherkaoui, MD, MPH, Senior Researcher at AMBOSS
Zebulon Tolman, MD, Editor-in-Chief Emeritus at AMBOSS

Abstract

Purpose: We conducted this study to develop a deeper understanding of how studying with AMBOSS and other medical resources affects the USMLE Step 2 CK examination score.

Method: We surveyed 1,985 students from 159 US-based allopathic medical schools who took the USMLE Step 2 CK between 2020 and 2022 about resource usage via convenience sampling, snowball referral, and university outreach of both AMBOSS users and non-users. After conducting descriptive analyses, we ran a multiple regression model to assess the impact of various resources on USMLE Step 2 CK scores.

Results: On average, students preparing for Step 2 CK studied with 5.1 unique resources. Question banks (Qbanks) were the most frequently used category of resource followed by video, flashcards, review books, and podcasts. Using AMBOSS and UWorld Qbanks was consistently associated with higher scores on Step 2 CK. The multiple linear regression model showed that after controlling for Step 1 exam scores, Qbanks usage, flashcard usage, and length of dedicated study time, using AMBOSS, using UWorld, and using Anki to prepare were each associated with significantly higher USMLE Step 2 CK scores (AMBOSS: +10.4; UWorld: +8.3; Anki: +1.4; compared to not using). Using USMLE Rx and using BoardVitals were each significantly associated with lower scores on Step 2 CK (USMLE Rx: -2.6; BoardVitals: -3.2), while Kaplan had a minimal (+1.0), non-significant association. Respondents who only used the AMBOSS Qbank scored 2.1 points higher than those who used UWorld only, 13.0 points higher than those who used USMLE Rx only, and 13.6 points higher than those who used BoardVitals Qbank only. Using both AMBOSS and UWorld yielded a combined increase of 12.1 points or 3.8 points higher than using UWorld only or 1.7 points higher than using AMBOSS only.

Conclusion: Our results strongly indicate that using AMBOSS as a primary exam preparation resource can significantly increase USMLE Step 2 CK scores. Studying with AMBOSS in combination with UWorld and Anki can yield additional marginal increases to the USMLE Step 2 CK score, although the additional study time, stress, and financial costs may not be worthwhile

If you have any questions or comments, please reach out to our Research Team at research@amboss.com.

for many students. Other combinations of resources were significantly associated with lower scores or had a non-significant impact.

Background

The clerkship year of medical school has an especially steep learning curve because of the need to balance exam preparation alongside hospital rotations. With the change of the USMLE Step 1 exam to pass/fail, it is widely accepted that USMLE Step 2 CK will become the most important determinant for ensuring that applicants are shortlisted for interviews during the Residency Match process[1–4], since it will be the only standardized metric for residency applications. While the pressure to perform on Step 1 is expected to recede, students may feel pressured to manage more rigorous exam preparation for Step 2 CK.[4]

Compared with the research on the timing, intensity, and frequency of use of Step 1 exam preparation resources,[5–7] there is less information on resource predictors for high Step 2 CK exam scores. A review of available peer-reviewed literature on Step 2 CK exam performance shows that multiple-choice exam questions,[8] flashcards,[6] and prior standardized exam scores (Step 1 and Shelf exam scores in particular[9]) have the biggest positive impact, while a length of dedicated study time beyond two weeks had a slight negative effect for one school with a longitudinal interleaved clerkship curriculum.[10]

Research Purpose

AMBOSS' mission is to empower clinicians to provide the best possible care. This journey begins with a solid foundation in basic sciences and clinical specialties and continues with clinical experience.[11,12] Ultimately, the competencies of those intending to be physicians in the US are assessed via medical licensing examinations, including the USMLE Step 2 CK for the clerkship disciplines. The purpose of this study is to develop a deeper understanding of the drivers and successes (as measured by Step 2 CK scores) related to the use of AMBOSS and other digital medical resources for Step 2 CK exam preparation. These results are intended to support medical schools and AMBOSS in determining how to better prepare students for Step 2 CK and improve the clerkship year experience and preparedness of students as future clinicians.

Methodology

Target Population

The target population for this study was 4th-year medical students enrolled in enrolled in allopathic medical schools in the US, who had already passed Step 1, and completed their USMLE Step 2 CK exam. Students who met the above criteria and provided informed consent[a] via online survey were eligible to participate. Students received $15 for their participation and an optional $5 bonus for uploading a verified Step 2 CK score report.[b] We employed four waves of convenience

[a] See Appendix for informed consent and survey questionnaire documentation.
[b] 50.8% (n=1,008) of respondents uploaded a verified score report.

sampling by reaching out to AMBOSS users, followed by five waves of snowball referral and university outreach to increase cohort participation at universities and reach non-AMBOSS users.

Research Design

For this study, we deployed an online questionnaire consisting of 20 questions in May 2022. Respondents quantified their usage of study materials, including Qbanks, videos, printed resources, and podcasts, as either "Used for Step 2 CK," "Used for Shelf exams,"[c] "Used for Shelf exams and Step 2 CK," or "Did not use." The resources included in this study (found in Table 4) were based on mentions in peer-reviewed literature,[5–8] review of social media recommendations,[13–17] and prior internal research conducted with MD students in their clinical clerkships.

Only respondents who answered "Used for Step 2 CK" or "Used for Shelf exams and Step 2 CK" were counted as having used the resource.[d] The survey data was supplemented with survey respondents' usage data obtained from the AMBOSS platform. This step was taken to reduce the total number of survey questions and to minimize respondent and recall bias in self-reported Qbank usage statistics.

Prior to conducting the multiple regression analysis, we summarized the univariate descriptive statistics for unique resources used, standardized exam scores, the number of Step 2 CK questions answered, and the length of dedicated study time. In addition, we performed a series of bivariate cross-tabulations to assess the potential impact of unique resources used, total number of resources used, and resource type on average Step 2 CK scores. Next, we conducted a correlation analysis with standardized exam scores and length of dedicated study time to assess the strength and direction of the continuous variables for the multiple regression models. Lastly, we ran a series of multiple regression models to assess the impact of Qbank and flashcard resources on Step 2 CK scores while controlling for length of dedicated study weeks, Shelf exams scores, and Step 1 scores.

Results + Analysis

We received 2,119 responses and, after eliminating duplicate responses (n=45), incomplete or missing data in survey responses (n=59), and unqualifying or unlikely survey responses[e] (n=30), we arrived at 1,985 completed responses. A total of 159 MD universities[f] were represented in this study.

---

[c] Subject-based, standardized exams created by the NBME that students typically take after a clinical clerkship. Subjects include: Internal medicine, surgery, pediatrics, obstetrics and gynecology, psychiatry, neurology, and often family medicine and ambulatory care.

[d] This classification is consistent with Drake et al's[8] study on resources used for Step 2 CK.

[e] Self-reported (responses were removed if Step 1 scores < 190 or > 300 n=15 removed). Self-reported (48.2%) and verified (50.8%) via Step 2 CK score report (responses removed if Step 2 CK scores > 300 n=4)

[f] See Table 11 in Appendix for full list of universities.

Descriptive statistics of standardized exam scores, questions answered, and study time

This section highlights the univariate descriptive statistics for standardized exam scores, questions answered, and study time. Table 1 shows the mean, standard deviation (SD), and the range of respondents' Step 1 and Step 2 CK scores; number of Step 2 CK questions answered in AMBOSS; and length of dedicated study time. The average scores for Step 1 and Step 2 CK were higher than the national averages by +5.9 and +4.8 points, respectively.[g] Respondents reported an average of 4.2 weeks of dedicated studying for Step 2 CK. The majority (74.1%) of respondents had taken their Step 2 CK exam between May 2021 and September 2021 (see Chart 3 in Appendix for exam months and dates).

Table 1: Descriptive Statistics for Standardized Exam Scores, Step 2 CK Questions Answered, and Study Weeks (n=1,985)

Exam Scores and Study Preparation	*Mean (SD)*	*Range*
Step 1 Score	237.9 (16.4)	194–281
Step 2 CK Score	250.7 (13.7)	196–284
No. of Step 2 CK Questions[a]	641.0 (957.0)	0–3200
Dedicated Study Weeks	4.2 (2.0)	0–16

[a] Total number of questions answered in AMBOSS

Descriptive statistics of total resources used, by category, and average Step 2 CK score

It is often assumed that using too few or too many resources has been shown to have a negative impact on standardized exam scores.[27,28] Since we were not able to find peer reviewed literature to support these assumptions, we conducted a series of descriptive analyses to analyze this relationship in more detail. This section summarizes the bivariate descriptive statistics for the total number of unique resources used and the category of resource used on Step 2 CK scores. There are a wide range of recommendations on the best ways to study for Step 2 CK including using one Qbank to studying with multiple different resources.[19, 27,28] We wanted to analyze the non-linear relationship between resource usage and Step 2 CK scores to determine what (if any) associations exist between the quantity of unique resources used and average Step 2 CK score.

Table 2 shows the descriptive statistics for the average of Step 2 CK scores by total unique resources used. Students used an average of 5.1 resources, which consisted of 1.6 Qbanks, 1.5 videos, 0.8 review books, 0.6 podcasts, and 0.6 flashcards for Step 2 CK. Respondents who used 2 resources averaged the highest scores on Step 2 CK (253.1±13.5), followed by 3

[g] Average Step 1 Score in 2021 was 231 (SD 19) and average Step 2 CK score in 2021 was 246 (SD 15)[13]

resources (252.3±13.5), and 4 resources (252.6±13.9). Of the respondents who had used 1 resource (n=60), 90% (n=54) used UWorld only (see Table 6) and averaged 250.3 (±13.1) on Step 2 CK. Using 5 resources was associated with an average score of 250.9 (±13.5). From 5 resources onward, the average score continues to drop with each additional resource added. Respondents using 7 or more made up 23.9% of all respondents and had the lowest average Step 2 CK score (247.9±13.6).[h]

Table 2: Average Step 2 CK Score by Total Number of Qbank, Video, Review Books, Flashcard, and Podcast Resources Used (n=1,985)

Total Resources	*n (%)*	*Average Step 2 CK Score (SD)*	*Range*
1 resource	60 (3.0%)	250.3 (13.1)	218–273
2 resources	175 (8.8%)	253.1 (13.5)	219–278
3 resources	318 (16.0%)	252.3 (13.5)	210–281
4 resources	352 (17.7%)	252.6 (13.9)	200–284
5 resources	322 (16.2%)	250.9 (13.5)	209–280
6 resources	282 (14.2%)	249.9 (13.4)	196–278
7 or more resources	475 (23.9%)	247.9 (13.6)	210–280

Table 3 shows the descriptive statistics for the average Step 2 CK score by total Qbanks used. Almost all respondents (99.5%) used at least one Qbank for Step 2 CK. Respondents who used 2 Qbanks had higher average Step 2 CK scores (251.4±13.3) compared to respondents who used 0,[i] 1, or 3 or more Qbanks.

Less than half (42.9%) of respondents used only 1 Qbank for Step 2 CK, and this group, on average, studied with 3.1 additional resources (1.2 video resources, 0.8 review books, and 0.5 podcasts). Just over half (50.6%) of respondents used 2 Qbanks for Step 2 CK and this group on average studied with 3.6 additional resources (1.5 video resources, 0.8 review books, and 0.7 podcasts).

[h] Respondents who used 7 or more resources, used on average 2.0 Qbanks, 3.0 videos, 1.6 review books, and 1.1 podcasts.

[i] 100% of respondents (n=10) who had used 0 Qbanks for Step 2 CK had used 1 or more Qbanks for Shelf exam preparation. Respondents had an average Step 1 score of 240.2 (SD 21.9).

Table 3: Average Step 2 CK Score by Total Number of Qbanks Used (n=1,985)

Qbanks	*n (%)*	*Average Step 2 CK Score (SD)*	*Range*
0 Qbanks	10 (0.5%)	242.9 (22.7)	200–267
1 Qbank	851 (42.9%)	250.5 (13.7)	204–284
2 Qbanks	1004 (50.6%)	251.4 (13.3)	196–282
3 or more Qbanks	120 (6.0%)	247.4 (14.9)	210–280

Tables 4–6 summarize the descriptive statistics for the average Step 2 CK score by total videos, review books, and podcasts used. Over a quarter (28.7%) did not use any video resources for Step 2 CK, and this group also had the highest Step 2 CK scores (253.0±13.3). Nearly half of respondents did not use any review books (45.2%) or podcasts (49.5%) for Step 2 CK. Respondents who used 0 videos, review books, or podcasts had higher average scores (253±13.3, 252.4±13.0, and 251.5±13.4 respectively) than respondents who used 1 or more of these resources. Adding 2 Qbanks to Step 2 CK study preparation was, on average, associated with higher scores and adding 1 or more video, review book, and/or podcast resources was associated with lower average scores.

Table 4: Average Step 2 CK Score by Total Number of Videos Used (n=1,985)

Video resources	*n (%)*	*Average Step 2 CK Score (SD)*	*Range*
0 videos	570 (28.7%)	253.0 (13.3)	196–284
1 video	547 (27.6%)	251.3 (14.4)	200–282
2 videos	455 (22.9%)	249.6 (13.2)	204–280
3 or more videos	413 (20.8%)	248.1 (13.2)	212–278

Table 5: Average Step 2 CK Score by Total Number of Review Books Used (n=1,985)

Review books	*n (%)*	*Average Step 2 CK Score (SD)*	*Range*
0 review books	897 (45.2%)	252.4 (13.0)	200–280
1 review book	763 (38.4%)	250.5 (13.8)	201–284

2 or more review books	325 (16.4%)	246.7 (14.3)	196–276

Table 6: Average Step 2 CK Score by Total Number of Podcasts Used (n=1,985)

Podcasts	*n (%)*	*Average Step 2 CK Score (SD)*	*Range*
0 podcasts	982 (49.5%)	251.5 (13.4)	201–282
1 podcasts	847 (42.7%)	250.6 (13.6)	196–284
2 or more podcasts	156 (7.9%)	247.0 (15.1)	200–277

Descriptive statistics of unique resources used and average Step 2 CK score

Table 7 shows the resources used for Step 2 CK exam preparation. UWorld was the most widely used resource at 95.7%, followed by Anki at 59.1% and AMBOSS at 57.7%.[j] Almost half of respondents had used First Aid for the USMLE Step 2 CK (45.1%). The most frequently used video resources were Sketchy Medical and OnlineMedEd at 44.3% and 42.8%, respectively. Divine Intervention was the most frequently used podcast resource at 46.6%.

The highest average Step 2 CK scores were observed among those respondents who used Clinical Problem Solvers (255.8±13.8), The Curbsiders (252.7±15.1), Anki (251.8±13.2), and AMBOSS (251.3±13.5), followed by UWorld with 250.8±13.6, which was very similar to the overall average of all respondents (250.7±13.7). Notably, usage of 22 out of 27 specific resources was associated with Step 2 CK scores that were lower than the reported average.

Table 7: Resources Used for Step 2 CK Exam Preparation n=1,985

Resources Used for Step 2 CK Exam	*n (%)*	*Average Step 2 CK Score (SD)*
Qbanks		
UWorld	1900 (95.7%)	250.8 (13.6)
AMBOSS	1146 (57.7%)	251.3 (13.5)

[j] This figure refers only to Step 2 CK use. When considering use throughout the clerkship year including for Shelf and Step 2 CK exams preparation, use of AMBOSS increases to 76.2% and UWorld to 98.0%. For the full list of resources used throughout the clerkship year, please refer to Table 12.

Kaplan	84 (4.4%)	245.8 (15.3)
USMLE Rx	76 (3.4%)	247.3 (14.9)
BoardVitals	50 (2.5%)	246.5 (14.3)
Flashcards		
Anki	1173 (59.1%)	251.8 (13.2)
Review books		
First Aid for the USMLE Step 2 CK	896 (45.1%)	249.4 (13.8)
Step-Up to Medicine	222 (11.2%)	247.0 (14.6)
Lecture and course material	198 (10.0%)	249.9 (13.6)
Master the Boards for USMLE Step 2 CK	107 (5.4%)	245.4 (15.6)
Case Files review books	52 (2.6%)	248.4 (13.2)
Blueprints review books	22 (1.1%)	248.0 (14.3)
Video Resources		
Sketchy Medical	880 (44.3%)	249.8 (13.4)
OnlineMedEd	849 (42.8%)	248.9 (13.4)
Boards & Beyond	436 (22.0%)	248.9 (13.8)
Emma Holliday	405 (20.4%)	249.8 (13.0)
Osmosis	199 (10.0%)	247.1 (13.7)
Ninja Nerd Lectures	70 (3.5%)	247.2 (13.8)
Physeo	42 (2.1%)	245.4 (15.3)
Lecturio	29 (1.5%)	249.4 (15.0)
Podcasts		
Divine Intervention	925 (46.6%)	249.8 (14.0)
USMLE Step 2 CK Secrets	145 (7.3%)	246.3 (14.2)
The Curbsiders	85 (4.3%)	252.7 (15.1)
The Clinical Problem Solvers	56 (2.8%)	255.8 (13.8)
CoreEM	26 (1.3%)	248.4 (14.7)

Emergency Medical Minute	12 (0.6%)	250.2 (10.0)
Surgery 101	12 (0.6%)	242.8 (13.8)

Table 8 provides an overview of the descriptive statistics for the average Step 2 CK score by dedicated study weeks. A minority (2.0%) of students allocated 0 weeks for dedicated Step 2 CK study preparation, and this group scored, on average, 4.8–5.1 points lower than students who had allocated 1 to 3 weeks for dedicated study preparations. Students who had studied for 4 or more weeks did not, on average, score higher Step 2 CK scores than students who studied for 1 to 3 weeks.

Table 8: Average Step 2 CK Score by Dedicated Study Weeks (n=1,985)

Dedicated Study Weeks	*n (%)*	*Average Step 2 CK Score (SD)*	*Range*
0 weeks	40 (2.0%)	249.3 (16.1)	209–272
1 week	42 (2.1%)	254.1 (12.5)	227–273
2 weeks	209 (10.5%)	254.3 (13.0)	212–278
3 weeks	367 (18.5%)	254.4 (12.6)	211–280
4 weeks	730 (36.8%)	250.8 (12.9)	204–284
5 or more weeks	597 (30.1%)	247.1 (14.4)	196–281

Bivariate correlations for standardized exam scores, number of Step 2 CK questions answered, and length of dedicated study time on statistics of average Step 2 CK score

Table 9 shows the bivariate correlations for all continuous variables: the standardized exam scores, number of Step 2 CK questions answered in AMBOSS, and dedicated study weeks. In this study, the number of Step 2 CK questions answered and Step 1 score showed a strong positive correlation with the Step 2 CK score while dedicated study weeks showed a slight negative correlation.[k] The total resources used for Step 2 CK was negatively correlated with the Step 1 and Step 2 CK scores which means that students who scored lower on Step 1 may have opted for a longer dedicated study period although the additional resources used were not associated with an increase in Step 2 CK score.

[k] The strength and direction of the correlations were not statistically different from those found in prior studies.[10,14–16]

Step 1 scores and length of dedicated study time were negatively correlated which means that survey respondents who had scored lower on Step 1 were slightly more likely to have had a longer dedicated study period for Step 2 CK. Giordano et al[24] found a similar relationship when investigating the impact of first-year grades on Step 1 score. On average, lower performing students allocated more dedicated study days for Step 1 exam preparation.[24] Dedicated study weeks and number of Step 2 CK questions answered in AMBOSS were negatively correlated which means that respondents who had a longer dedicated study period did not answer more questions in AMBOSS.[l]

Table 9: Bivariate Correlations of Standardized Exam Scores, Questions Answered in AMBOSS, and Dedicated Study Weeks n=1,985

Variable n=1,985	*Step 1 Score*	*Step 2 CK Score*	*No. of Step 2 CK Qs*[a]	*Dedicated Study Weeks*	*Total Resources Used*[b]
Step 1 Score	-	.724**	.059**	-.269*	-.159**
Step 2 CK Score		-	.100**	-.226**	-.124**
No. of Step 2 CK Questions[1]			-	- .048*	-.012
Dedicated Study Weeks				-	.167**
Total Resources Used					-

Note: **p < .01 and *p < .05
[a] Total number of questions answered in AMBOSS
[b] Includes Qbanks, video, review books, and podcast resources

Multiple regression model assessing the impact of resources on Step 2 CK scores

We used a multiple linear regression model to assess the independent contribution of each resource to the Step 2 CK score while taking into account the use of other resources. Using a multiple regression analysis enabled us to untangle the complex relationship between related variables. Qbanks, podcasts,[m] and flashcards were included since their use was positively associated with a higher average score compared to not using them. Video and review books

[l] When measuring the total number of Step 2 CK questions, authors Drake et al[5] also limited the total number of questions for a single (unspecified) Qbank. Model 2 in the Appendix shows the analysis for unique resources without the number of Step 2 CK questions.
[m] Students who had used The Curbsiders and The Clinical Problem Solvers podcasts had higher average Step 2 CK scores (253.1±16.0 and 257.9±12.6) than the survey average (250.7±13.7).

were not included, since the use of these resources was not associated with higher average scores compared to not using them.

We used a linear regression model with a predicted Step 2 CK score as the dependent variable and AMBOSS Qbank use, UWorld Qbank use, AMBOSS × UWorld Qbanks interaction,[n] USMLE Rx Qbank use, Kaplan Qbank use, BoardVitals Qbank use, Anki use, The Curbsiders podcast use, The Clinical Problem Solvers podcast use, number of Step 2 CK questions answered in AMBOSS, Step 1 score, and dedicated study weeks as the independent variables. AMBOSS Qbank use, UWorld Qbank use, AMBOSS × UWorld Qbank interaction, USMLE Rx Qbank use, Kaplan Qbank use, BoardVitals Qbank use, Anki use, The Curbsiders podcast use, and The Clinical Problem Solvers podcast use were binary variables (0 = did not use for Step 2 CK; 1 = used for Step 2 CK) and Step 1 score, number of Step 2 CK questions,[o] and dedicated study weeks were continuous variables.

AMBOSS use is measured twice (once as a binary variable: "used for Step 2 CK or not;" and once as a continuous variable: "number of Step 2 CK questions answered in AMBOSS"). Since AMBOSS includes both a Qbank and a medical library that contains over 1200 articles with hundreds of videos, the use of the platform is more heterogeneous and varied than for a question-only platform. This means that our model assumes that non-AMBOSS Qbank usage is more homogeneous[p]. We reasoned that it was necessary to capture the different uses via two variables.

Table 10 shows the predicted change in Step 2 CK score that can be attributed to resource usage while controlling for Step 1 score and dedicated study weeks. The full model details (e.g. coefficients, residuals, and measures of goodness of fit) can be found in Table 14, Chart 1, and Chart 2. AMBOSS, UWorld Qbank, AMBOSS and UWorld Qbank, and Anki usage were significantly associated with higher Step 2 CK scores. BoardVitals and USMLE Rx were significantly associated with lower Step 2 CK scores. The use of Kaplan Qbank, The Curbsiders podcast, and The Clinical Problem Solvers were not significantly associated with Step 2 CK scores.

Students who answered 3200 questions in AMBOSS experienced the greatest impact on Step 2 CK score, scoring 10.4 points higher than students who did not use AMBOSS and 2.1 points higher than students who used UWorld only. Respondents using both AMBOSS and UWorld scored 12.1 points higher than those who did not and 3.8 points higher than those that used UWorld only. Students who used Anki scored on average 1.4 points higher than those who did not. Students who used Kaplan scored on average 1.0 points higher than those who did not,

---

[n] We included an interaction term for AMBOSS × UWorld since 50.7% of respondents used both AMBOSS and UWorld for their Step 2 CK studies. Table 12 summarizes the mean Step 2 CK score for respondents who used both resources which is 6.8 points higher than the use of UWorld alone.

[o] Prior to the Step 2 CK exam, the average AMBOSS user answered 641.0 (SD 957.0) questions. The large standard deviation reflects the variation in usage. We conducted a log transformation of the data in order to correct for the potential skew and the findings remained significant (models available upon request).

[p] We concluded that the general recommendation for UWorld usage is at least 1 pass with an average of 70% correct.[20-23]

although this was not significant (p=.358). This is likely due to the wider fluctuations in Step 2 CK scores.[q] Students who used USMLE Rx and BoardVitals scored on average 2.6 points and 3.1 points lower, respectively, than those who did not use these Qbanks.

Table 10: Model 1—Assessing the Impact of Using AMBOSS, UWorld, USMLE Rx, Kaplan, BoardVitals, and Anki on Step 2 CK Controlling for Step 1 Score and Dedicated Study Weeks n=1,985

Model 1—Impact of Variables on Step 2 CK Score n=1,985	
Resources Used for Step 2 CK	*Change in Step 2 CK Score*
Used AMBOSS (3200 questions)	+10.4**
Used UWorld	+8.3**
Used AMBOSS + UWorld	+12.1*
Used USMLE Rx	-2.6*
Used Kaplan	+1.0
Used BoardVitals	-3.2*
Used Anki	+1.4**
Used The Curbsiders	-0.21
Used The Clinical Problem Solvers	+2.2

Note: ** p< .01 *p < .05

Regression model 1 accounted for 53.8% of the variance[r] in the dataset. As a measure of the goodness of fit of the model and to rule out potential systematic biases, we conducted an analysis of the difference between the predicted score and actual scores. The differences were not correlated with the fitted values which therefore did not suggest a systematic bias in the model.[s]

[q] The standard deviation for Step 2 CK score for Kaplan Qbank use (SD 15.3) was larger than for the other Qbanks (see Table 14).
[r] Measured by adjusted R squared.
[s] For a detailed explanation of the multiple linear regression model equation and regression coefficients, see Table 14.

Chart 1 plots the actual Step 2 CK score (x-axis) and predicted Step 2 CK score (y-axis) for Model 1. The Pearson's correlation coefficient of .736 indicates a strong positive relationship which is also significant (p<.001), which suggests a good model fit.

Chart 1: Plot of Predicted Values for Model 1 vs. Actual Step 2 CK Scores n=1,985

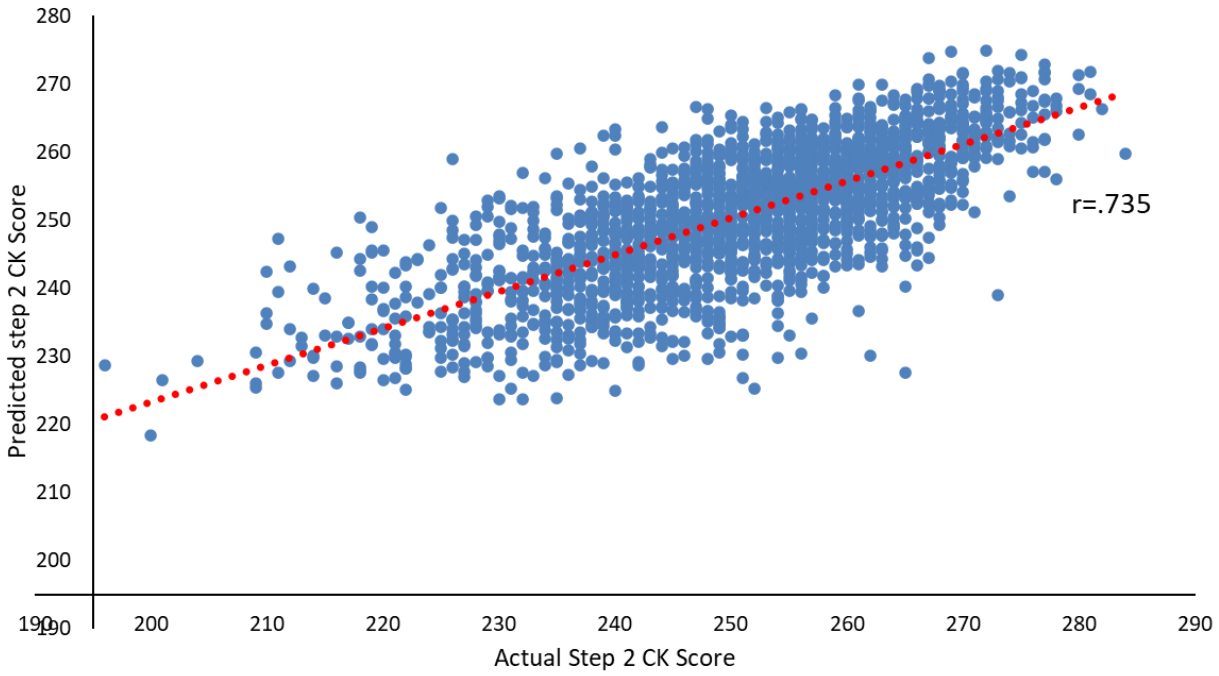
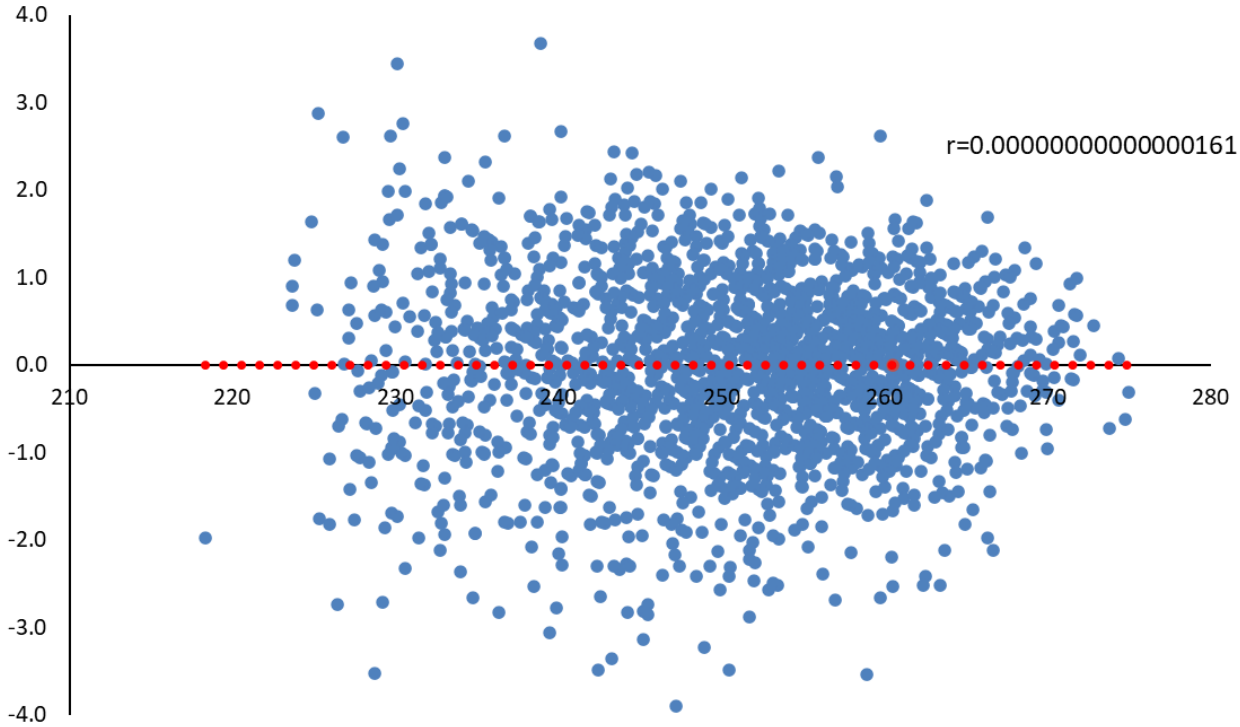


Chart 2 shows the goodness of fit of the model by plotting the predicted values on the x-axis and the standardized difference between the predicted values and the actual scores (standardized residuals) on the y-axis. The residuals are not correlated as indicated by the Pearson's correlation coefficient (r=.00000000000000161), which suggests that there is no systematic bias in the model. Due to a lower representation (in our study) of scores in the bottom 50th percentile, the spread is larger around lower values less than 250.

Chart 2: Plot of Goodness of Fit of Model Predicted Values for Model 1 vs. Standardized Residuals n=1,985



Discussion

This study provides an overview of the variety of resources and study habits used by students to prepare for Step 2 CK. A series of univariate (descriptive statistics), bivariate (cross-tabulations and correlation), and multivariate (multiple regression) analyses were used to explore and understand the strength and direction of the relationship between resource usage and Step 2 CK score.

Using fewer resources was associated with higher scores

The results of the bivariate analyses lend support to the assumption[27,28] that using too few or too many resources may have a negative impact on Step 2 CK scores. Students used on average 5.1 unique resources for Step 2 CK, which correlated with an average Step 2 CK score of 250.7

(±13.7). This average score is higher than the national average which is discussed further in the study limitations below.

The average number of resources reflects a wide variation in unique resources used with 3.0% using 1 resource, 42.5% using 2–4 resources and 38.1% using 6 resources or more (Table 2). While the average of resources used was 5.1, those who used 2 resources had the highest average score at 253.1 (±13.5). Consequently, it may appear that the average student is studying with too many resources.

This hypothesis is consistent with what the authors have heard, anecdotally, about study recommendations,[27,28] namely, that maximizing learning efficiency means focusing on fewer resources rather than splitting one's attention across using too many. According to educational psychology, the Cognitive Load Theory suggests that juggling various resources likely increases the extraneous cognitive load.[28] This effect may hinder overall learning since more time is spent on synthesizing information from incongruent resources rather than on repeating content. Despite this widespread recommendation on the benefits of using fewer resources, we were unable to identify any studies to support this claim as it relates to USMLE board exams preparation.

Quantity matters and so does quality

Furthermore, our results in Table 13 also showed detailed breakdowns of unique Qbank, video, review book, and podcast resources used by students in preparing for Step 2 CK. Of the 17 combinations, the top 5 scoring combinations used 3–4 resources. This finding indicates that some students can do well if using more than 2 resources. In these cases, the main determinants of the score were likely the type and quality of resources used rather than the quantity.

For example, within the 7 highest scoring groups, both the AMBOSS and UWorld Qbank were consistently utilized. The group that used only the AMBOSS and UWorld Qbanks averaged 256.2 (±12.8) on Step 2 CK.[t] Adding Anki and Divine Intervention resulted in an increase of 1.4 points to the average score at 257.6 (±10.6) and adding Anki and Sketchy resulted in a 1.1 point increase at 257.3 (±10.4). Students who used these combinations of 4 resources averaged an additional 1.4 and 1.1 points higher, respectively, than students who used AMBOSS and UWorld only. Nevertheless, it is worth considering if the extra time and effort invested in studying with 2 additional resources, in addition to 2 Qbanks, would be worth the minimal increase.

When looking at the types of resources (Qbank, video, review book, or podcast) used overall, Table 3 shows that the highest Step 2 CK average was observed among those who had used 2 Qbanks with 251.4 (±13.3), although 1 Qbank was just slightly behind at 250.5 (±13.7). Those using 3 or more Qbanks only averaged 247.4 (±14.9), while those using 0 averaged 242.9 (±22.7). This finding indicates that those who scored the highest focused on using 2 Qbanks, and those studying with only 1 Qbank scored on average 0.9 points lower. Although these findings are descriptive, this association begs the question of whether, for the average student, the mild

[t] which was 5.4 points higher than respondents who used UWorld Qbank only who had an average of 250.8 (±12.7)

increase to the average score would be worth the additional effort in time, costs, and stress of studying with an additional Qbank.

Looking at other types of resources in Table 4–6, respondents who had used 0 videos review books, or podcasts, respectively, had the highest average scores as compared to using 1 or more of these resources. This downward trend continues with each additional video, review book, or podcast resource added. These findings suggest that using video, review books, or podcasts re likely to result in a lower average score; however, it is important to note that there was a wide fluctuation in the individual scores and within the specific resources. For example, as seen in Table 7, students who used The Clinical Problem Solvers or The Curbsiders podcasts scored higher than the average at 255.8 (±13.8) and 252.7 (±15.1), respectively. Upon closer inspection of the individual resources, 22 out of 27 resources were associated with lower than average scores.[u]

All in all, these findings suggest that students preparing for Step 2 CK may want to prioritize studying with 1–2 Qbanks rather than video, review books, or podcasts. It is important to note that we cannot definitively assess the impact of using one resource over another by looking at the univariate or bivariate statistics in isolation. There may be other self-selecting factors which cannot be adequately measured in either analysis

Not too long not too short - dedicated study weeks on Step 2 CK score

One potential self-selecting factor which we had examined in more detail was the length of the dedicated study time on Step 2 CK score. In Table 9, there is a significant negative correlation between dedicated study weeks and Step 1 and Step 2 CK score. This means that students who scored lower on Step 1 were more likely to have allocated more time to preparing for Step 2 CK. Nevertheless, this additional time does not necessarily result in a higher score (Table 8). In Tables 7 and 9, we observed a negative association between length of dedicated study time and Step 2 CK score. Students who studied 1–3 weeks had higher average scores (1 week=254±12.5; 2 weeks=254.3±13.0; 3 weeks=254.4±12.6) compared to those who studied 0 weeks (249.3±16.1) or 4 or more weeks (247.1±14.4).

Prior studies have found that there is a positive correlation between the number of questions answered[5,8] and standardized exam scores.[6] Authors Antonio et al[7] found that a length of study time over 2 weeks was associated with lower scores on Step 2 CK. Compared to the peer-reviewed literature on Step 1 and dedicated study time,[10,24] there are fewer studies that assess the impact of the length of dedicated study time on Step 2 CK scores. Giordano et al[24] found that students who received straight A's during their first year of medical school did not necessarily improve their Step 1 score by adding more study days to their exam prep. However, students who did not receive straight A's during their first year were able to improve their Step 1 score by adding more study days to their exam preparation. The authors proposed that by allocating more

[u] The survey average for Step 2 CK score was 250.7 (+13.7).

dedicated study days to Step 1 preparation may enable lower performing students to perform as well on Step 1 as higher performing peers which may also be the case for Step 2 CK preparation.

In this study, the length of dedicated study time was negatively associated with the Step 2 CK score. In particular, studying for 4 or more weeks was associated with lower scores (Table 8). This relationship may be explained, partially, by the total number of resources used which is also positively associated with the length of dedicated study time (Table 9). The relationship between more unique resources used for Step 2 CK and lower average scores may also be explained by self-selection whereby students who had a lower Step 1 score opted for a longer dedicated period of study. To reduce the potential for confounding biases, these associations would need to be studied further.

Recommendations from multiple regression analysis

To help untangle some of the complex relationships between resource usage, dedicated study weeks, and standardized exams scores, we used a multiple linear regression model to assess the independent contribution of each on the Step 2 CK score while accounting for the presence of the other factors. The resources included in the multiple regression analysis were chosen based on whether their use resulted in a higher than average increase to the Step 2 CK score (Table 7).

As shown in Table 10, AMBOSS, UWorld Qbank, and Anki were the only unique resources that showed statistically significant increases to the average score as compared to not using these resources. Students who used AMBOSS Qbank experienced the highest increase at 10.4 points over those who did not use AMBOSS and 3.8 points higher than respondents who used UWorld only. Students who used Anki scored on average 1.4 points higher than those who did not. Students who used Kaplan scored on average 1.0 points higher than those who did not, although this was not significant (p=.358). This is likely due to the wider fluctuations in Step 2 CK scores. Students who used USMLE Rx and BoardVitals scored on average 2.6 points and 3.1 points lower, respectively, than those who did not use these Qbanks.

Study Strengths

The strengths of this study are in its large sample size (n=1,985) and inclusion of multiple institutions (159 US MD universities). We did not find any peer-reviewed studies that analyzed the effect of standardized scores and resource usage on Step 2 CK scores which also included more than one institution. Other studies involved single institutions[5,25] in which case, only one study analyzed the impact of resource usage on Step 2 CK scores.[5] The studies which involve multiple institutions[14,15] assessed the impact of grades and prior standardized exams scores on Step 2 CK. This study builds upon the findings from other Step 2 CK studies[5,7,14,15,25] and accounts for resources used by type and quantity (Qbanks and flashcard resources), standardized exam scores, and dedicated study time. All together, these factors provide us with a representative cross-section of medical students who have taken Step 2 CK between 2020 and 2022. This has enabled us to build a comprehensive and actionable understanding of which resources were associated with higher Step 2 CK scores.

Study Limitations

We acknowledge several limitations to the study. To reduce the potential burden in time and length of the survey on participants, we relied on self-reported measures for Step 1 and Shelf exam scores—with an option to upload a verified Step 2 CK score report for an extra incentive. Over half (50.6%) of the Step 2 CK scores were verified via uploaded score reports. We verified 100% of participants' score reports and matched 99.99% of these self-reported scores to their official reports. In the cases where the self-reported responses did not match the score report , the self-reported responses were updated accordingly.[v] This extremely high rate of matched self-reported and verified scores provides at least some reassurance that the self-reported scores among all participants, even those who did not upload a verified score report, were likely accurate.

Another limitation was that we did not assess the level of usage of resources used for Step 2 CK. Other studies on Step 1 scores have qualified the extent[10,26] to which a resource was used (e.g. light vs. heavy use), and in this study, we did not assess level of usage for each resource except for the AMBOSS Qbank. This was done to minimize recall bias and to reduce the survey burden.

We employed a convenience study design and observed that there was a self-selection bias in survey participation. The average scores for Step 1 and Step 2 CK were higher than the national averages.[w] Nevertheless, this is consistent with other studies, albeit for Step 1, that have used survey data to supplement the standardized scores. Authors Burk-Rafel et al[10] deployed a study behaviors survey to students who had taken Step 1 and found that respondents' average Step 1 score was higher than that of non-respondents.

It is also important to note that respondents took Step 2 CK prior to Step 1 changing to pass/fail. While this provides us with a useful baseline, we can expect that future cohorts taking Step 2 CK may have different study habits and behaviors. For example, universities may start experimenting with a dedicated Step 2 CK study time and thereby reduce the heterogeneity in study weeks, which was previously largely self-determined. Nevertheless, these insights may prove useful to current and future Step 2 CK exam takers.

Lastly, we did not specify if the Step 1 score was a first or second exam try so we reasoned that respondents would likely enter their highest Step 1 score. Responses for Step 1 score that were less than 194 or greater than 300 were removed from the analysis, as the former score was the cutoff for passing for all respondents and the latter score was outside the possible scoring scale

Conclusion

[v] Only n=5 self-reported scores did not match the verified score report. These errors were within +5 points of the verified scores and in one case where the error was larger, this was likely due to a typo (515 instead of 215).

[w] Average Step 1 Score in 2021 was 231 (SD 19) and average Step 2 CK score in 2021 was 246 (SD 15).

This study was intended to quantify the current resources used for Step 2 CK study preparation and to assess their potential impact on Step 2 CK scores. These results may help medical students and universities better assess the quality and quantity of various resources that are used for study preparations. Our results strongly indicate that using too few or too many resources was associated with lower scores on Step 2 CK.

In addition, the results of the bivariate and multivariate analyses strongly suggest that using AMBOSS as a primary exam preparation resource can significantly increase Step 2 CK scores. When assessing the impact of individual resources on Step 2 CK score, the use of the AMBOSS Qbank was significantly associated with the highest increase to Step 2 CK score, followed by the UWorld QBank. And using both AMBOSS and UWorld yielded a higher score than using either resource alone. The use of Anki was also significantly associated with a minor increase to Step 2 CK score. Utilizing almost all other combinations of resources will most likely be associated with lower score or have a non-significant impact.

Therefore, we would recommend that students who are looking to score higher on Step 2 CK strongly focus on studying with AMBOSS, and possibly UWorld Qbank and Anki during their dedicated Step 2 CK study preparation. Students wanting to economize on time, effort, and cost, which will be the case for many students during their busy and stressful clerkship year, should consider adding AMBOSS as their sole Qbank since the use of AMBOSS was associated with the highest increase to Step 2 CK score. Students who have additional time and budget and who are looking to maximize their Step 2 CK score by using 2 Qbanks should consider using both AMBOSS and UWorld as primary Qbanks since the use of these two Qbanks yielded a higher score than using either alone.

Appendix

Informed Consent

Survey participants were asked to provide informed consent at the start of the survey. Those that did not consent by selecting “I disagree” were not eligible to participate.

> *Thank you for taking the time to answer this survey about your clerkship year study experiences and scores.*
>
> *Before we start, it may be helpful to have your MCAT, Step 1, Shelf and Step 2 CK scores on hand.*
>
> *Use of Survey Data: This survey will link your responses to an internal user ID that connects your responses to your AMBOSS data. The survey data and your personal info will remain anonymous throughout the entire study! You may request that the personal data collected in this survey be sent to you as a copy, edited, or deleted by emailing us at research@amboss.com. We delete all personal data after two years. We will never rent or sell your information to anyone, ever. If you have any questions, you can contact us at research@amboss.com.*
>
> *By clicking "I agree" below, you are indicating that you are at least 18 years old, have read and understood this consent form, and agree to participate in this survey.*
>
> - *I agree*
> - *I disagree*

Survey

1. Please select your current year of study
   - ◯ PGY-2 //if selected then disqualify
   - ◯ PGY-1 //if selected then disqualify
   - ◯ 4th year (M4)
   - ◯ 3rd year (M3) //if selected then disqualify
   - ◯ 2nd year (M2) //if selected then disqualify
   - ◯ 1st year (M1) //if selected then disqualify

2. Have you taken the Step 2 CK exam?
   - ◯ Yes
   - ◯ No //if selected then disqualify

3. When did you take the Step 2 CK exam? (date response field)

4. How many weeks did you dedicate solely to preparing for Step 2 CK? (open-ended number field)

5. Please describe your use of these Qbanks and flashcard resources for your clerkship year study prep:

	Used for Shelf exams only	Used for Step 2 CK only	Used for Shelf and Step 2 CK	Did not use	Never heard of it
Kaplan	○	○	○	○	○
UWorld	○	○	○	○	○
AMBOSS	○	○	○	○	○
USMLE Rx	○	○	○	○	○
Board Vitals	○	○	○	○	○
Anki	○	○	○	○	○

//randomize row options

6. Please describe your use of these print / textbooks for your clerkship year study prep:

	Used for Shelf exams only	Used for Step 2 CK only	Used for Shelf and Step 2 CK	Did not use	Never heard of it
First Aid	○	○	○	○	○
Master the Boards	○	○	○	○	○
Step-Up to Medicine	○	○	○	○	○
Dr. Pestana's Surgery Notes	○	○	○	○	○
Blueprints	○	○	○	○	○
Case Files	○	○	○	○	○
Lecture and course material	○	○	○	○	○

//randomize row options

7. Please describe your use of these video resources for your clerkship year study prep:

	Used for Shelf exams only	Used for Step 2 CK only	Used for Shelf and Step 2 CK	Did not use	Never heard of it
Boards & Beyond	○	○	○	○	○
OnlineMedEd	○	○	○	○	○
Physeo	○	○	○	○	○
Lecturio	○	○	○	○	○
Osmosis	○	○	○	○	○
Ninja Nerd Lectures	○	○	○	○	○
Emma Holliday	○	○	○	○	○
Sketchy	○	○	○	○	○

//randomize row options

8. Please describe your use of these podcasts for your clerkship year study prep:

	Used for Shelf exams only	Used for Step 2 CK only	Used for Shelf and Step 2 CK	Did not use	Never heard of it
Surgery 101	○	○	○	○	○
Divine Intervention	○	○	○	○	○
Emergency Medical Minute	○	○	○	○	○
The Curbsiders	○	○	○	○	○
Core EM	○	○	○	○	○
USMLE Step 2 Secrets / Inside the Boards	○	○	○	○	○
The Clinical Problem Solvers	○	○	○	○	○

//randomize row options

10. Would you please share your score on these exams?

Please fill this out as accurately as possible. Your information will remain confidential. Personally identifiable information will be anonymized or removed. Your responses will help us improve AMBOSS for future students to reach your target score.

MCAT USMLE Step 1 USMLE Step 2 CK Medicine Shelf Surgery Shelf Family Medicine Shelf Pediatrics Shelf OB-GYN Shelf Psychiatry Shelf	(open-ended numeric response)

Representation of Survey Respondents in Study

Table 11 lists the US MD universities represented in the study. Responses by university range from n=1 to n=45.

Table 11: List of Universities Represented in the Study n=1,985

Universities n=1,985	n	%
Wake Forest University School Of Medicine	45	2.3%
Washington University in St. Louis School of Medicine	38	1.9%
Columbia University College of Physicians and Surgeons	35	1.8%
Thomas Jefferson University Sidney Kimmel Medical College	33	1.7%
Wayne State University School of Medicine	33	1.7%
George Washington University School of Medicine and Health Sciences	30	1.5%
Case Western Reserve University School of Medicine	29	1.5%
University of Illinois College of Medicine - Chicago	29	1.4%
University of South Florida Morsani College of Medicine (USF)	26	1.3%
University of Florida College of Medicine	26	1.3%
Indiana University School of Medicine	26	1.3%
University of Iowa Carver College of Medicine	26	1.3%
Northeast Ohio Medical University College of Medicine	25	1.3%
University at Buffalo (SUNY) Jacobs School of Medicine and Biomedical Sciences	25	1.3%

Texas A&M University Health Science Center College of Medicine	23	1.2%
University of Hawaii at Manoa John A. Burns School of Medicine	23	1.2%
Northwestern University Feinberg School of Medicine	23	1.1%
University of Toledo College of Medicine and Life Sciences	21	1.1%
Icahn School of Medicine at Mount Sinai	21	1.1%
University of Maryland - Baltimore School of Medicine	21	1.0%
Howard University College of Medicine	20	1.0%
Albert Einstein College of Medicine	20	1.0%
Yale School of Medicine	19	1.0%
Zucker School of Medicine at Hofstra/Northwell	19	1.0%
University of Miami Leonard M. Miller School of Medicine	19	1.0%
University of Michigan Medical School	19	1.0%
Rutgers New Jersey Medical School	18	0.9%
Medical College of Georgia at Augusta University	17	0.9%
University of Louisville School of Medicine	17	0.9%
University of Washington School of Medicine	17	0.9%
SUNY Downstate Medical Center College of Medicine	17	0.9%
University of Texas Health Science Center at Houston (UT-Houston) McGovern Medical School	17	0.8%
Eastern Virginia Medical School	16	0.8%
Medical University of South Carolina College of Medicine	16	0.8%
Rush Medical College Of Rush University	16	0.8%
Michigan State University College of Human Medicine	15	0.8%
New York Medical College	15	0.8%
Renaissance School of Medicine at Stony Brook University	15	0.8%
Western Michigan University Homer Stryker MD School of Medicine	15	0.8%
University of Central Florida College of Medicine	15	0.8%

Louisiana State University (LSU) School of Medicine - New Orleans	15	0.7%
Chicago Medical School at Rosalind Franklin University of Medicine and Science	14	0.7%
University of Nebraska College of Medicine	14	0.7%
University of Texas School of Medicine at San Antonio (UT-San Antonio)	14	0.7%
University of Arizona College of Medicine - Tucson	14	0.7%
University of Texas Southwestern Medical Center at Dallas, Southwestern Medical School (UT-Southwestern)	14	0.7%
University of South Dakota Sanford School of Medicine	14	0.7%
University of California, Los Angeles David Geffen School of Medicine (UCLA)	13	0.7%
University of Rochester School of Medicine and Dentistry	13	0.7%
University of Utah School of Medicine	13	0.7%
Brown University Warren Alpert Medical School	13	0.7%
University of Colorado Anschutz Medical Campus	13	0.7%
Weill Cornell Medical College	13	0.7%
Meharry Medical College School of Medicine	13	0.6%
Drexel University College of Medicine	12	0.6%
University of Mississippi School of Medicine	12	0.6%
University of Wisconsin School of Medicine and Public Health	12	0.6%
Creighton University School of Medicine	12	0.6%
The Ohio State University College of Medicine	12	0.6%
University of Arizona College of Medicine - Phoenix	12	0.6%
University of Minnesota Medical School	12	0.6%
University of Texas Medical Branch at Galveston School of Medicine	12	0.6%
University of Kansas School of Medicine - Kansas City	12	0.6%
University of Kentucky College of Medicine	12	0.6%
University of Virginia School of Medicine	12	0.6%
Johns Hopkins University School of Medicine	12	0.6%

Baylor College of Medicine	12	0.6%
Temple University Lewis Katz School of Medicine	11	0.6%
California Northstate University College of Medicine	11	0.6%
University of Nevada, Reno School of Medicine	11	0.6%
Virginia Commonwealth University School of Medicine	11	0.6%
University of Pittsburgh School of Medicine	11	0.6%
University of Nevada, Las Vegas School of Medicine	11	0.6%
Wright State University Boonshoft School of Medicine	11	0.6%
Creighton University Health Sciences Campus - Phoenix	10	0.5%
Saint Louis University School of Medicine	10	0.5%
University of Oklahoma College of Medicine	10	0.5%
University of Vermont Robert Larner, MD College of Medicine	10	0.5%
University of Arkansas for Medical Science (UAMS) College of Medicine	10	0.5%
Emory University School of Medicine	10	0.5%
UMass Chan Medical School	10	0.5%
Duke University School of Medicine	10	0.5%
University of Cincinnati College of Medicine	10	0.5%
University of North Carolina (UNC) Chapel Hill School of Medicine	10	0.5%
CUNY School of Medicine	10	0.5%
Albany Medical College	9	0.5%
Dartmouth College Geisel School of Medicine	9	0.5%
SUNY Upstate Medical University College of Medicine	9	0.5%
Uniformed Services University of the Health Sciences F. Edward Hébert School of Medicine	9	0.5%
University of Texas Rio Grande Valley School of Medicine	9	0.5%
Texas Tech University Health Sciences Center Paul L. Foster School of Medicine	9	0.5%
University of Missouri - Kansas City School of Medicine	9	0.5%

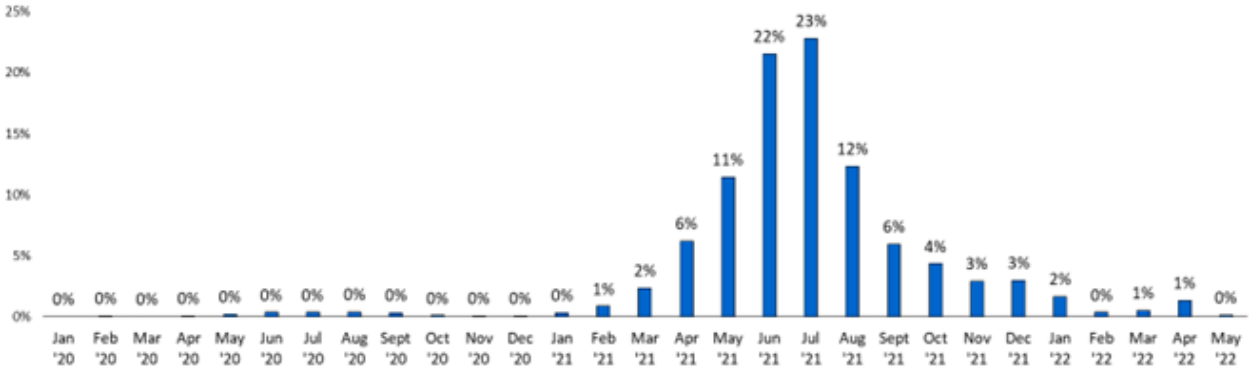
West Virginia University School of Medicine	9	0.5%
California University of Science and Medicine	9	0.5%
University of Illinois College of Medicine - Rockford	8	0.4%
University of Tennessee Health Science Center College of Medicine	8	0.4%
Loma Linda University School of Medicine	8	0.4%
Vanderbilt University School of Medicine	8	0.4%
Tulane School of Medicine	8	0.4%
University of North Dakota School of Medicine and Health Sciences	7	0.4%
University of South Carolina - Greenville School of Medicine	7	0.4%
University of Kansas School of Medicine - Wichita	7	0.4%
Geisinger Commonwealth School of Medicine	7	0.4%
University of Missouri-Columbia School of Medicine	7	0.4%
Oregon Health & Science University School of Medicine	7	0.4%
University of Chicago Pritzker School of Medicine	7	0.3%
Boston University School of Medicine	7	0.3%
Tufts University School of Medicine	7	0.3%
Florida State University College of Medicine	6	0.3%
Mayo Clinic School of Medicine - Minnesota	6	0.3%
Oakland University William Beaumont School of Medicine	6	0.3%
University of California, Davis School of Medicine (UC Davis)	6	0.3%
Medical College of Wisconsin - Milwaukee	6	0.3%
University of California, Riverside School of Medicine (UC Riverside)	6	0.3%
University of South Carolina School of Medicine	6	0.3%
East Tennessee State University James H. Quillen College of Medicine	5	0.3%
Frank H. Netter MD School of Medicine at Quinnipiac University	5	0.3%
Texas Tech University Health Sciences Center School of Medicine	5	0.3%
Virginia Tech Carilion School of Medicine	5	0.3%

Washington State University Elson S. Floyd College of Medicine	5	0.3%
Marshall University Joan C. Edwards School of Medicine	5	0.3%
Rutgers Robert Wood Johnson Medical School	5	0.3%
Dr. Kiran C. Patel College of Allopathic Medicine at Nova Southeastern University	5	0.3%
Central Michigan University College of Medicine	5	0.3%
Stanford University School of Medicine	5	0.3%
University of California, Irvine School of Medicine (UC Irvine)	5	0.3%
Harvard Medical School	5	0.3%

Note: Universities with fewer than 5 responses were Charles R. Drew University/UCLA Medical Education Program (CDU UCLA), Florida Atlantic University Charles E. Schmidt College of Medicine, Florida International University Herbert Wertheim College of Medicine, Georgetown University School of Medicine, Hackensack Meridian School of Medicine, Indiana University School of Medicine - Fort Wayne, Indiana University School of Medicine - Terre Haute, Loyola University Chicago Stritch School of Medicine, Mayo Clinic Alix School of Medicine–Arizona, Medical College of Wisconsin - Green Bay, New York University Long Island School of Medicine, Penn State College of Medicine, Southern Illinois University School of Medicine, Temple/St. Luke's School of Medicine, The OU-TU School of Community Medicine, The University of Texas at Austin Dell Medical School, University of California, San Diego School of Medicine (UCSD), University of Connecticut School of Medicine, University of Illinois at Urbana-Champaign Carle Illinois College of Medicine, University of Illinois College of Medicine–Peoria, University of Kansas School of Medicine–Salina, University of Kentucky College of Medicine–Bowling Green Campus, University of Minnesota Medical School - Duluth Campus, University of Missouri-Columbia School of Medicine, University of New Mexico School of Medicine, University of Pennsylvania Perelman School of Medicine, University of South Alabama College of Medicine, University of South Carolina School of Medicine, and University of Southern California Keck School of Medicine (USC).

Chart 3 shows the Step 2 CK exam months for survey respondents. The range of exam dates spanned January 2020 to May 2022.

Chart 3: Step 2 CK Exam Months for Survey Respondents n=1,985



Descriptive statistics of unique combination of resources used and average Step 2 CK score

Table 12 expands on Tables 3–7 by providing a detailed breakdown of the combinations[x] of Qbank, video, review book, and flashcard resources used and the average Step 2 CK score. Respondents who used both UWorld and AMBOSS were consistently shown to have a higher average Step 2 CK score than respondents who used UWorld alone or UWorld in combination with other resources.

UWorld, AMBOSS, Anki, and Divine Intervention for Step 2 CK averaged the highest Step 2 CK score at 257.6±10.6 and respondents who used UWorld, AMBOSS, Anki, Sketchy had a similar average Step 2 CK score at 257.3±10.4. Respondents who used UWorld and AMBOSS had an average score of 256.2±12.8. Adding one additional resource to the UWorld and AMBOSS combination like Anki or First Aid yielded a mild increase to the average score at 0.2 and 0.6 points respectively. Adding Anki and Divine Intervention or Anki and Sketchy produced a moderate increase in average score of 1.5 and 1.1 points, respectively.

Table 12: Resources Used for Shelf and/or Step 2 CK Exams Preparation n=1,985

Resources Used for Shelf Exams and Step 2 CK	*n (%)*
Qbanks	
UWorld	1939 (98.0%)
AMBOSS	1512 (76.2%)
USMLE Rx	140 (7.1%)
Kaplan	118 (5.9%)
BoardVitals	88 (4.4%)
Flashcards	
Anki	1402 (70.6%)
Review books	
First Aid for the USMLE Step 2 CK	1073 (54.1%)
Lecture and course material	939 (47.3%)
Case Files review books	875 (44.1%)
Step-Up to Medicine	684 (34.5%)
Blueprints review books	336 (16.9%)
Master the Boards for USMLE Step 2 CK	142 (7.2%)

[x] Combinations range from 1 to 4 unique resources. Table 2 illustrates how the use of 5 or more resources was associated with lower average scores. Average Step 2 CK score for combinations of 5 or more resources are available upon request.

Video Resources	
OnlineMedEd	1650 (83.1%)
Emma Holliday	1329 (67.0%)
Sketchy Medical	1051 (52.9%)
Boards & Beyond	598 (30.1%)
Osmosis	321 (16.2%)
Ninja Nerd Lectures	139 (7.0%)
Physeo	67 (3.4%)
Lecturio	59 (3.0%)
Podcasts	
Divine Intervention	1072 (54.0%)
The Curbsiders	257 (12.9%)
USMLE Step 2 CK Secrets	178 (9.0%)
CoreEM	148 (7.5%)
The Clinical Problem Solvers	145 (7.3%)
Surgery 101	41 (2.1%)
Emergency Medical Minute	28 (1.4%)

Table 13: Average Step 2 CK Score for Respondents by Total Resources Used (1–4 Resources only) n=1,985

Descriptive Statistics	*Total Resources*	*n*	*Average Step 2 CK Score (SD)*	*Range*
UWorld + AMBOSS + Anki + Divine Intervention	4	36 (1.8%)	257.6 (10.6)	234–278
UWorld + AMBOSS + Anki + Sketchy	4	27 (1.4%)	257.3 (10.4)	233–277
UWorld + AMBOSS + First Aid	3	25 (1.3%)	256.8 (11.3)	239–281
UWorld + AMBOSS + Anki	3	47 (2.3%)	256.4 (12.2)	211–275
UWorld + AMBOSS + Divine	3	19 (1.0%)	256.4 (9.9)	241–276
UWorld + AMBOSS	2	40 (2.0%)	256.2 (12.8)	225–278
UWorld + AMBOSS + Emma Holliday + Divine Intervention	4	10 (0.5%)	256.1 (6.0)	247–265
UWorld + Divine Intervention	2	33 (1.7%)	254.7 (12.8)	228–277
UWorld + First Aid	2	22 (1.1%)	254.4 (9.7)	234–270
UWorld + Anki + Sketchy	3	17 (0.9%)	254.4 (13.5)	226–273
UWorld + First Aid + Anki	3	16 (0.8%)	254.0 (13.3)	228–274
UWorld + OnlineMedEd + Sketchy + First Aid	4	12 (0.6%)	253.6 (11.2)	233–270
UWorld + Anki	2	31 (1.6%)	253.2 (13.0)	219–273
UWorld + Anki + Divine Intervention	3	29 (1.4%)	253.1 (13.9)	214–273
UWorld + OnlineMedEd	2	17 (0.9%)	252.3 (17.5)	219–274
UWorld + Sketchy	2	15 (0.8%)	251.3 (15.4)	221–272

UWorld only	1	54 (2.7%)	250.8 (12.7)	222–273

Note: Only combinations for groups where n ≥ 10 (or ≥ 0.5%) are shown.

Results of the Linear Regression Model 1

The results of the linear regression for Model 1 are represented in this equation.

$$y' = B_0 + AMBOSSx_1 + UWorldx_2 + AMBOSS\ and\ UWorldx_1x_2 + USMLE\ Rx\ x_3 +$$
$$+ Kaplan\ x_4 + Board\ Vitals\ x_5 + Anki\ x_6 + The\ Curbsiders\ x_7$$
$$+ The\ Clinical\ Problem\ Solvers\ x_8 + No.Step\ 2\ CK\ Qs\ answered\ in\ AMBOSSx_9$$
$$+ Step\ 1\ Score\ x_{10} + Dedicated\ Study\ Weeksx_{11}$$

$$y' = 101.464 + 7.104x_1 + 8.272x_2 - 6.609x_1x_2 - 2.603x_3 + 1.031x_4 - 3.171x_5$$
$$+ 1.448x_6 - 0.207x_7 + 2.215x_8 + 0.001x_9 + 0.590x_{10} - 0.220x_{11}$$

The impact of using AMBOSS is 7.104 + 0.001 × 3200 = 10.404 points while controlling for: use of UWorld Qbank, use of both AMBOSS and UWorld Qbanks, use of USMLE Rx Qbank, use of Kaplan Qbank, use of BoardVitals Qbank, use of Anki, use of The Curbsiders podcast, use of The Clinical Problem Solvers podcast, number of Step 2 CK questions answered in AMBOSS, Step 1 score, and dedicated weeks spent studying for Step 2 CK. Regression model 1 accounted for 53.8% of the variance[y] in the dataset.

The additional impact of using both AMBOSS and UWorld is modified by the negative interaction term. This means that using both resources will not produce an additive effect of an 15.375 point increase (main effects for AMBOSS + UWorld are 10.404 + 8.272 = 18.675) rather in our dataset, using both resources resulted in a 12.067 (main effects for AMBOSS + UWorld - AMBOSS × UWorld interaction is 10.404 + 8.272 – 6.609 = 12.067) additional point increase. Using both Qbanks represents an additional gain of 1.663 points over using AMBOSS alone and a 3.795 point gain over using UWorld alone. Table 14 depicts the regression coefficients for Model 1.

[y] measured by the adjusted R^2. For Model 1 R=.736 and (unadjusted) R^2=.541.

Table 14: Regression Coefficients for Model 1–Assessing the Impact of Using Qbanks and Anki on Step 2 CK Controlling for Step 1 Score and Dedicated Study Weeks n=1,985

Regression Coefficients	*B unstandardized coefficients* (SE)	*β standardized coefficients*	*p-value*
Used AMBOSS	7.104 (2.682)	0.257	.008
Used UWorld	8.272 (2.431)	0.123	.001
Used AMBOSS and UWorld	-6.609 (2.699)	-0.241	.014
Used USMLE Rx	-2.603 (1.162)	-0.037	.025
Used Kaplan	1.031 (1.121)	0.015	.358
Used BoardVitals	-3.171 (1.407)	-0.036	.024
Used Anki	1.448 (0.431)	0.052	.001
Used The Curbsiders	-0.207 (1.144)	-0.003	.856
Used The Clinical Problem Solvers	2.215 (1.406)	0.027	.116
No. of Step 2 CK Questions	0.001 (<.001)	0.072	<.001
Step 1 Score	0.590 (0.013)	0.708	<.001
Dedicated Study Weeks	-0.220 (0.110)	-0.032	.046
Adjusted R^2	.538		

Note: With the exception of the interaction term and its component effects (which would be expected), collinearity was not significant. When the interaction term was removed from the model, no collinearity effects were detected.
[1] Total number of questions answered in AMBOSS

Multiple regression model 2 assessing the impact of resources on Step 2 CK scores

In model 2, we used a linear regression model with a predicted Step 2 CK score as the dependent variable and AMBOSS use, UWorld use, AMBOSS × UWorld interaction,[z] USMLE Rx use, Kaplan use, BoardVitals use, Anki use, The Curbsiders use, The Clinical Problem Solvers use, Step 1 score, and dedicated study weeks as the independent variables. AMBOSS use, UWorld use, AMBOSS × UWorld interaction, USMLE Rx use, Kaplan use, BoardVitals use, Anki use, The Curbsiders use, and The Clinical Problem Solvers use were binary variables (0 = did not use for Step 2 CK; 1 = used for Step 2 CK) and Step 1 score and dedicated study weeks were continuous variables.

Table 10 shows the predicted change in Step 2 CK score that can be attributed to resource usage while controlling for Step 1 score and dedicated study weeks. The full model details (e.g. coefficients, residuals, and measures of goodness of fit) can be found in Table 16. The AMBOSS Qbank, UWorld Qbank, AMBOSS and UWorld Qbanks, and Anki usage were significantly associated with higher Step 2 CK scores. BoardVitals and USMLE Rx were significantly associated with lower Step 2 CK scores. The use of Kaplan Qbank, The Curbsiders podcast, and The Clinical Problem Solvers were not significantly associated with Step 2 CK scores.

Students who used AMBOSS experienced the greatest impact on Step 2 CK score, scoring 8.7 points higher than students who did not use AMBOSS and 0.4 points higher than students who used UWorld only. Respondents using both AMBOSS and UWorld scored 9.4 points higher than those who did not and 1.1 points higher than those that used UWorld only. Students who used Anki scored on average 1.5 points higher than those who did not. Students who used Kaplan scored on average 1.1 points higher than those who did not, although this was not significant (p=.325). Students who listened to The Curbsiders or the Clinical Problem Solvers podcasts scored on average 0.02 and 2.0 points higher, respectively, than students who did not use these resources. Neither of the podcasts were significant (p=.988 and p=.155 respectively). Students who used USMLE Rx and BoardVitals scored on average 2.7 points and 3.3 points lower, respectively, than those who did not use these Qbanks.

[z] We include an interaction term for AMBOSS × UWorld since 50.7% of respondents used both AMBOSS and UWorld for their Step 2 CK studies. Table 13 summarizes the mean Step 2 CK score for respondents who used both resources which is 6.8 points higher than the use of UWorld alone.

Table 15: Model 2—Assessing the Impact of Using AMBOSS, UWorld, USMLE Rx, Kaplan, BoardVitals, Anki, the Curbsiders, and the Clinical Problem Solvers on Step 2 CK Controlling for Step 1 Score and Dedicated Study Weeks n=1,985

Model 2—Impact of Variables on Step 2 CK Score n=1,985	
Resources Used for Step 2 CK	*Change in Step 2 CK Score*
Used AMBOSS	+8.7**
Used UWorld	+8.3**
Used AMBOSS + UWorld	+9.4**
Used USMLE Rx	-2.7*
Used Kaplan	+1.1
Used BoardVitals	-3.3*
Used Anki	+1.5**
Used The Curbsiders	+0.02
Used The Clinical Problem Solvers	+2.0

Note: ** p< .01 *p < .05

Regression model 2 accounted for 53.7% of the variance[aa] in the dataset. As a measure of the goodness of fit of the model and to rule out potential systematic biases, we conducted an analysis of the difference between the predicted score and actual scores. The differences were not correlated with the fitted values which therefore did not suggest a systematic bias in the model.

[aa] Measured by adjusted R squared

Table 16: Regression Coefficients for Model 2–Assessing the Impact of Using Qbanks and Anki on Step 2 CK Controlling for Step 1 Score and Dedicated Study Weeks n=1,985

Regression Coefficients	*B unstandardized coefficients (SE)*	*β standardized coefficients*	*p-value*
Used AMBOSS	8.677 (2.669)	0.314	.001
Used UWorld	8.274 (2.442)	0.123	.001
Used AMBOSS and UWorld	-7.547 (2.703)	-0.275	.005
Used USMLE Rx	-2.730 (1.166)	-0.038	.019
Used Kaplan	1.109 (1.126)	0.016	.325
Used BoardVitals	-3.304 (1.413)	-0.038	.019
Used Anki	1.469 (0.433)	0.053	.001
Used The Curbsiders	0.018 (1.148)	<0.001	.988
Used The Clinical Problem Solvers	2.008 (1.412)	0.024	.155
Step 1 Score	0.593 (0.013)	0.711	<.001
Dedicated Study Weeks	-0.243 (0.111)	-0.035	.028
Adjusted R^2	.534		

Note: With the exception of the interaction term and its component effects (which would be expected), collinearity was not significant. When the interaction term was removed from the model, no collinearity effects were detected.
[1] Total number of questions answered in AMBOSS

Results of the Linear Regression Model 2

The results of the linear regression for Model 2 are represented in this equation.

$$y' = B_0 + AMBOSSx_1 + UWorldx_2 + AMBOSS\ and\ UWorldx_1x_2 + USMLE\ Rx\ x_3 +$$
$$+ Kaplan\ x_4 + Board\ Vitals\ x_5 + Anki\ x_6 + The\ Curbsiders\ x_7$$
$$+ The\ Clinical\ Problem\ Solvers\ x_8 + Step\ 1\ Score\ x_9 + Dedicated\ Study\ Weeksx_{10}$$

$$y' = 101.173 + 8.677x_1 + 8.274x_2 - 7.547x_1x_2 - 2.730x_3 + 1.109x_4 - 3.304x_5$$
$$+ 1.469x_6 + 0.018x_7 + 2.008\ x_8 + 0.593\ x_9 - 0.243x_{10}$$

The impact of using AMBOSS is 8.677 points while controlling for: use of UWorld Qbank, use of both AMBOSS and UWorld Qbanks, use of USMLE Rx Qbank, use of Kaplan Qbank, use of BoardVitals Qbank, use of Anki, use of The Curbsiders podcast, use of The Clinical Problem Solvers podcast, Step 1 score, and dedicated weeks spent studying for Step 2 CK. Regression model 1 accounted for 53.4% of the variance[bb] in the dataset.

The additional impact of using both AMBOSS and UWorld is modified by the negative interaction term. This means that using both resources will not produce an additive effect of an 16.950 point increase (main effects for AMBOSS + UWorld are 8.677 + 8.274 = 18.840) rather in our dataset, using both resources resulted in a 9.403 (main effects for AMBOSS + UWorld - AMBOSS × UWorld interaction is 8.677 + 8.274 – 7.547 = 9.403) additional point increase. Using both Qbanks represents an additional gain of 1.611 points over using AMBOSS alone and a 3.763 point gain over using UWorld alone. Table 14 depicts the regression coefficients for Model 1.

[bb] Measured by the adjusted R^2. For Model 2 R=.733 and (unadjusted) R^2=.537.

References

1. Sergesketter A, Shammas RL, Langdell HC, Geng Y, Eberlin KR, Fox P, Ko JH, Lifchez SD, Momoh AO, Nguyen V, Tenenbaum M, Thanik V, Phillips BT. Predicting academic performance during plastic surgery residency: Can Step 2 scores reliably replace Step 1? *J Surg Educ*. 2022;79(3): 828-36. doi: https://doi.org/10.1016/j.jsurg.2021.11.015
2. Zhang RV, Awan OA, Resnik CS, Hossain R. Potential impact of a pass or fail United States Medical Licensing Exam Step 1 scoring system on radiology residency applications. Acad Radiol. 2022;29(1):158-65. doi: https://doi.org/10.1016/j.acra.2020.10.016
3. Wang A, Karunungan KL & Story JD. Residency program director perspectives on changes to US Medical Licensing Examination. *JAMA Netw Open.* 2021;4(1) :e2129557. doi:10.1001/jamanetworkopen.2021.29557
4. Pontell ME, Makhoul, AT, Ganesh Kumar N, Drolet, BC. The change of USMLE step 1 to pass/fail: perspectives of the surgery program director. *J Surg Educ*. 2021;78(1):91–8. doi:10.1016/j.jsurg.2020.06.034
5. Drake E, Phillips JP, Kovar-Gough I. Exploring Preparation for the USMLE Step 2 Exams to Inform Best Practices. *PRiMER*. 2021;5:26. Published 2021 Aug 31. doi:10.22454/PRiMER.2021.69310
6. Zahn CM, Saguil A, Artino AR Jr, et al. Correlation of National Board of Medical Examiners scores with United States Medical Licensing Examination Step 1 And Step 2 scores. *Acad Med*. 2012;87(10):1348-1354.doi:10.1097/ACM.0b013e31826a13bd
7. Antonio S, Kracw RA, Dizon W, Simanton E. To wait or two weeks: The relationship between Step 2 CK scores and the length of dedicated study time within a longtidunal interleaved clerkship curriculum. *Cureus*. 2022;14(7): e26599. doi:10.7759/cureus.26599
8. Jackson F, Duane E, Harmon R, Kollar RA, Rainville NM, Smith RM. Resources That Improve Medical Board Licensing Examination Performance. *Cureus*. 2019;11(10):e5927. Published 2019 Oct 16. doi:10.7759/cureus.5927
9. Deng F, Gluckstein JA, Larsen DP. Student-directed retrieval practice is a predictor of medical licensing examination performance [published correction appears in Perspect Med Educ. 2016 Nov 18;:]. *Perspect Med Educ*. 2015;4(6):308-313. doi:10.1007/s40037-015-0220-x
10. Burk-Rafel J, Santen SA, Purkiss J. Study Behaviors and USMLE Step 1 Performance: Implications of a Student Self-Directed Parallel Curriculum. *Acad Med*. 2017;92(11S Association of American Medical Colleges Learn Serve Lead: Proceedings of the 56th Annual Research in Medical Education Sessions):S67-S74. doi:10.1097/ACM.0000000000001916
11. Woods NN, Brooks LR, Norman GR. The value of basic science in clinical diagnosis: creating coherence among signs and symptoms. *Med Educ*. 2005;39(1):107-12. doi: 10.1111/j.1365-2929.2004.02036.x
12. Norris ME, Cachia MA, Johnson MI, Rogers KA, Martin CM. Expectations and perceptions of students' basic science knowledge: through the lens of clerkship directors. *Med Sci Educ*. 2020;*30*(1):355-65. doi: 10.1007/s40670-019-00913-z

13. USMLE. USMLE score interpretation guidelines. USMLE.org; 2022; 1-4. Available at: https://www.usmle.org/sites/default/files/2022-05/USMLE%20Step%20Examination%20Score%20Interpretation%20Guidelines_5_24_22_0.pdf. Accessed July 1, 2022.
14. Ogunyemi D, Taylor-Harris D. Factors that correlate with the US Medical Licensure Examination Step-2 scores in a diverse medical student population. *J Natl Med Assoc*. 2005;97(9):1258-1262.
15. Kies SM, Roth V, Rowland M. Association of Third-Year Medical Students' First Clerkship With Overall Clerkship Performance and Examination Scores. *JAMA.* 2010;304(11):1220–1226. doi:10.1001/jama.2010.1184
16. Ghaffari-Rafi, A., Lee, R.E., Fang, R. *et al.* Multivariable analysis of factors associated with USMLE scores across US medical schools. *BMC Med Educ*. 2019;19**(**154). https://doi.org/10.1186/s12909-019-1605-z
17. Manguvo A, Litzau M, Quaintace J, Ellison S. Medical students' NBME subject exam preparation habits and their predictive effects on actual scores. *J Contemp Med Edu*. 2015;3(4):143-9. doi:10.5455/jcme.20151213105832
18. Divinepodcaster. My Step 2CK dedicated study period recommendations (by Divine). Accessed Jul 1, 2022. https://www.reddit.com/r/Step2/comments/hsf822/my_step_2ck_dedicated_study_period/
19. Sekaheimer. How good a resource is UWorld by itself for Step 2 CK. Accessed Jul 1, 2022. https://www.reddit.com/r/Step2/comments/beobbk/how_good_a_resource_is_uworld_by_itself_for_step/
20. Thedtothea. How to start studying for Step 2? UWorld is 4,000+ questions now so getting through it all is not realistic. *Reddit*. Posted March 31, 2022. Accessed August 9, 2022. https://www.reddit.com/r/Step2/comments/tsv4dc/how_to_start_studying_for_step_2_uworld_is_4000/
21. Kimdivo. Step 2 CK Experience–262. *Reddit.* Posted August 26, 2021. Accessed August 9, 2022. https://www.reddit.com/r/Step2/comments/pbx9gf/step_2_ck_experience_262/
22. CRISPER21. How do you use Uworld for Step 2 CK? *Reddit*. Posted Jan 21, 2021. Accessed August 9, 2022. https://www.reddit.com/r/Step2/comments/l1qc61/how_do_you_use_uworld_for_step_2_ck/
23. Giordano C, Hutchinson D, Peppler R. A predictive model for USMLE Step 1 scores. *Cureus*. 2016;8(9): e769. doi:10.7759/cureus.769.
24. 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;19(8):385-391. doi: 10.2147/AMEP.S138557.
25. Bonasso P, Lucke-Wold B 3rd, Reed Z, Bozek J, Cottrell S. Investigating the Impact of Preparation Strategies on USMLE Step 1 Performance. *MedEdPublish*. 2015;4(1):5. doi:10.15694/mep.2015.004.0005
26. Nahid50. Overwhelming resources. *Reddit*. Posted Dec 16, 2021. Accessed Nov 27, 2022. https://www.reddit.com/r/Step2/comments/rhq6q2/overwhelming_resources/

27. UAM90. Please advise what resources to follow. *Reddit*. Posted May 13, 2021. Accessed Nov 27, 2022. https://www.reddit.com/r/Step2/comments/nbbnx8/please_advise_what_resources_to_follow/
28. Schnotz W, Kurschner C. A reconsideration of cognitive load theory. *Educ Psychol Rev*. 2007;19:469-508. DOI 10.1007/s10648-007-9053-4