

# Evaluation of a Pre-Made Expanding-Retrieval Flashcard Tool and Medical Student Performance on Step 1

Alex Miller MS4; Geargin Wilson MS3; Howard Sachs MD

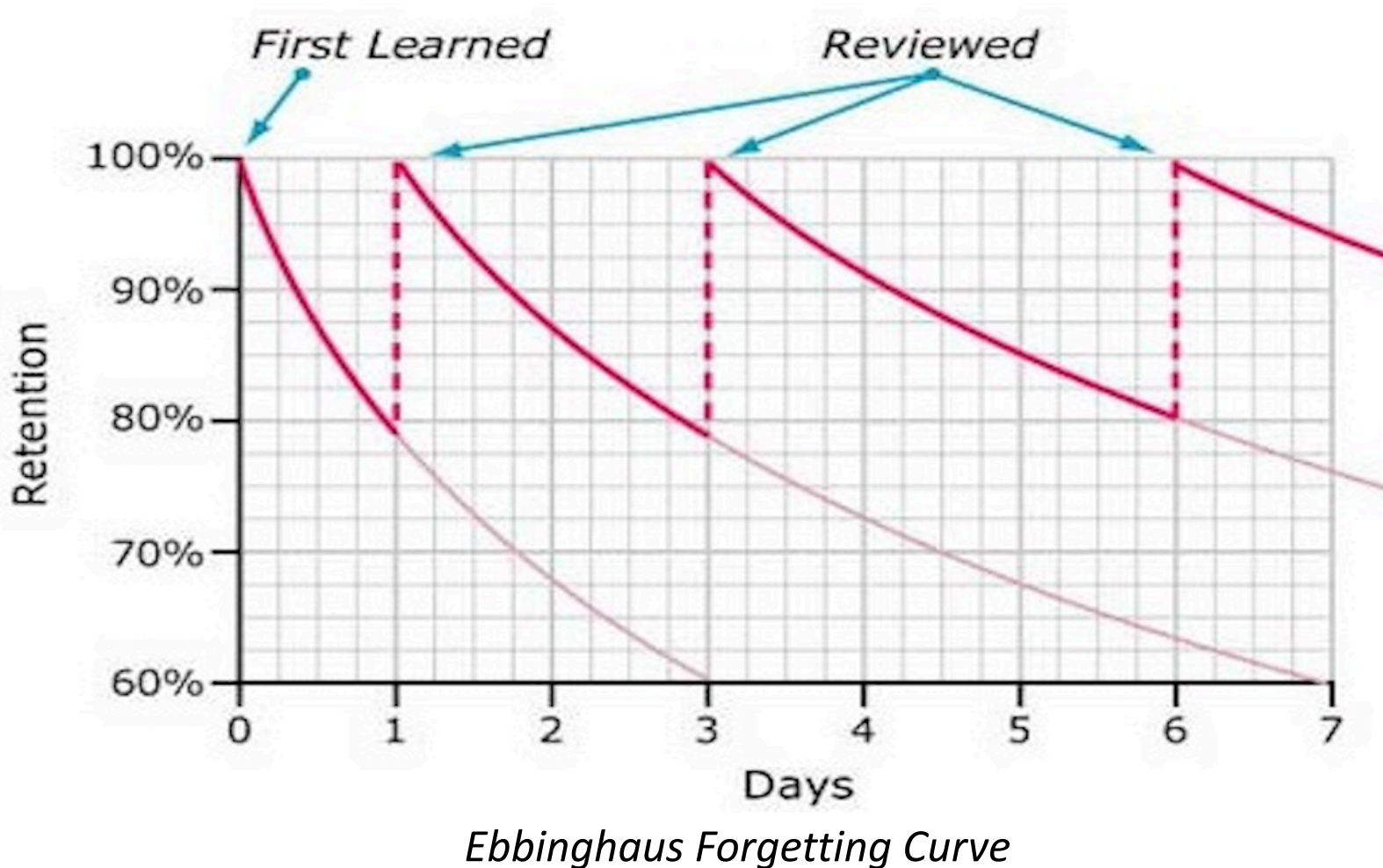
University of Massachusetts Medical School

## Introduction

As medical schools move away from pre-clinical grades, the United States Medical Licensing Examination (USMLE) Step 1 examination has become an increasingly important measure of student performance and candidacy for residency. A host of new and well-designed tools have emerged, one set of which are predicated on premade, spaced-repetition flashcards.

They are a departure from the process of studying, re-reading, re-highlighting, and re-annotating; flashcards necessitate some measure of active recall and afford corrective-feedback, test-enhanced learning. The inclusion of expanding-interval spaced-repetition learning algorithms to schedule the reviews of these cards further enhances these effects.

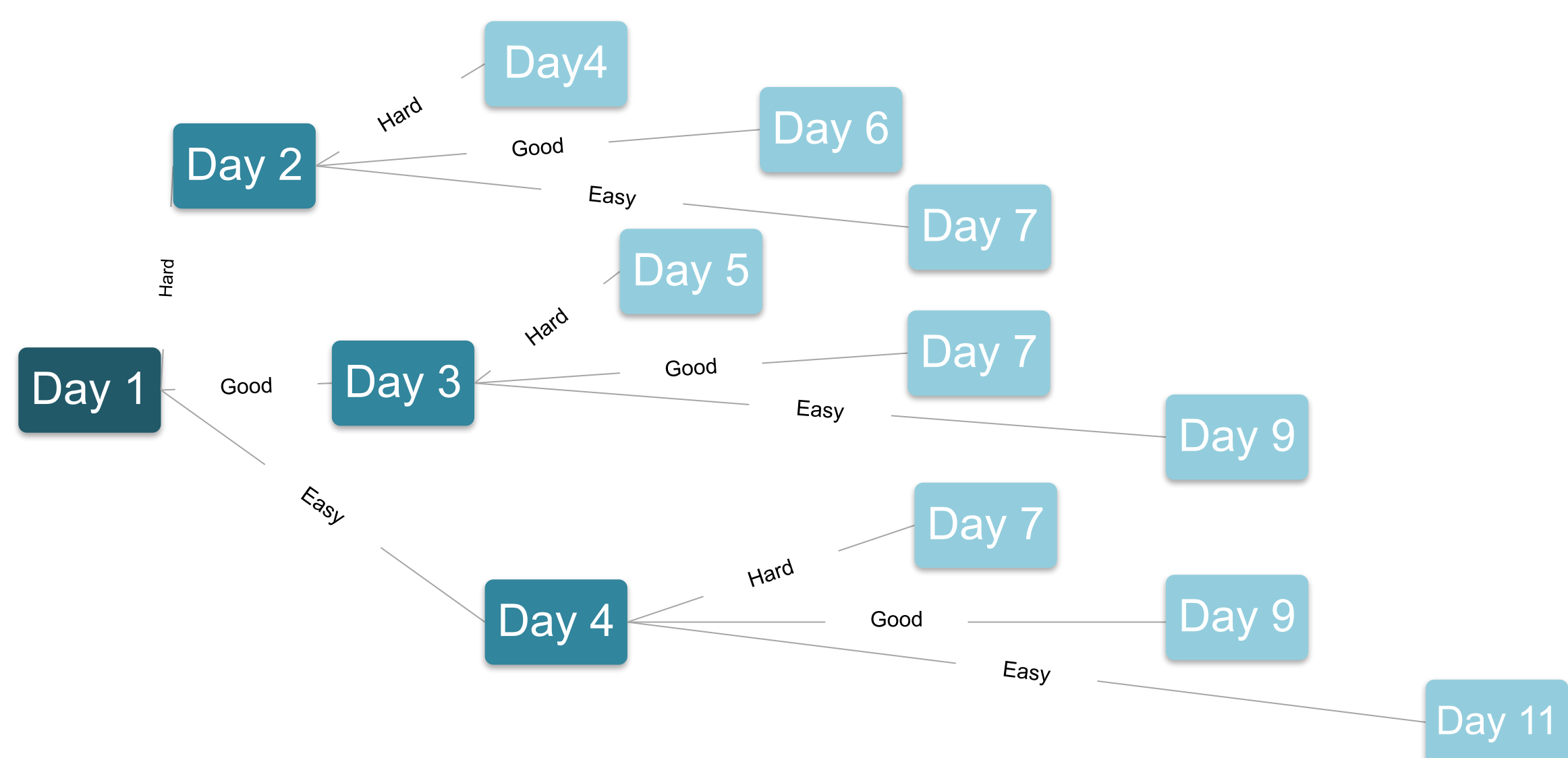
Expanding-interval repetition is predicated on the Ebbinghaus forgetting curve, which demonstrates the rate at which learned information decays if there is not subsequent review, but that with each instance of re-learning, the rate of decay decreases [b16] Complicating spaced-repetition learning further, its benefit appears greatest when materials are relearned at the precipice of being forgotten, when maximal effort is required for recall, but recall is still possible



## What is Anki...

Anki is an expanding-retrieval electronic flashcard program available for desktop and smartphone, capable of word processing, deck organization, and search functionality. It enables content creation, organization, and review in a single application.

After activating cards, students are presented with flashcards daily according to the SuperMemo-2 spaced-learning algorithm. With each card they are given the option of “Hard,” “Good,” “Easy” and “Again.” Each corresponds to a different multiplier of the spaced repetition algorithm, with ‘Again’ restarting the algorithm entirely.



## ...and what is the Maimonideck?

Drawing from the most common Step 1 preparation materials, two UMass Medical School students created a 3000-flashcard Anki deck covering the full breadth of Step 1 material, known as “The Maimonideck.” Designed for daily use over the 9-month course of second year, students utilizing the deck were instructed to activate cards as their topics were encountered; doing this consistently, they could expect from 50-150 cards of varying ages to review per day. The potential of daily study espoused in the Oath of Maimonides lent the project its name:

“Grant me the strength, time and opportunity always to correct what I have acquired, always to extend its domain; for knowledge is immense and the spirit of man can extend indefinitely to enrich itself daily with new requirements.”

## Objectives

1. Are there testing gains in students who used the Maimonideck versus those who did not, and how do those gains vary with patterns of use?
2. Did users consider the Maimonideck an effective, efficient method of Step 1 preparation?

## Methods

The 125 members of the UMass Class of 2019 were emailed before the start of second-year courses. The messages included a description of Anki and the Maimonideck and a link to request access. Those who requested access received a subsequent email that included a study fact sheet, instructions for downloading the 3000 cards, a setup guide for new users, a card index and topic organizer, and a link to a shared website for reporting errata.

### Baseline Survey

- Administered prior to the start of classes
- Evaluated baseline academic information, namely MCAT scores.

### NBME survey

- Administered ~12 weeks into second year, following score reporting of a customized NBME exam covering pulmonary, cardiovascular, and renal organ systems.
- Evaluated Maimonideck usage and interaction, online question bank use prior to the exam, and unadjusted NBME score.
- “Usage grade” and “Interaction Grade” were given based on table below
- “Primacy grade” was based on a 5-point scale asking if user agreed with the statement, “The Maimonideck is your primary study tool for reviewing previously covered material? Categories included Primary Tool (Agreed / strongly agreed), Secondary tool (Neutral), or Reference tool (Disagreed / Strongly disagreed).

| USAGE GRADING                |                  |                  | INTERACTION GRADING |                    |
|------------------------------|------------------|------------------|---------------------|--------------------|
| Duration of use              | Frequency of use | Completion %     | Editing             | Adding             |
| 5 All weeks                  |                  |                  |                     |                    |
| 4 Most weeks throughout      | 7 days / week    | >75% / session   | Substantial amount  | Substantial amount |
| 3 Some weeks throughout      | 5-6 days / week  | 50-74% / session | Fair amount         | Fair amount        |
| 2 Several weeks before exam  | 3-4 days / week  | <50% / session   | Small amount        | Small amount       |
| 1 A couple weeks before exam | 1-2 days / week  | Quizzes only     | Not at all          | Not at all         |

USAGE GRADE: High use:  $\geq 11$  points; Moderate use: 7-10 points; Low use:  $\leq 6$  points

INTERACTION GRADE: High interaction:  $\geq 6$ ; Moderate interaction: 4-5; Low interaction:  $\leq 3$

### Step 1

- Administered following Step 1 score reporting.
- Evaluated primacy of Maimonideck among other study resources in the months preceding the Step 1 reading period. Responses were sorted into the same Primary-Secondary-Reference tool groups afor the NBME exam survey.
- Evaluated what percentage of the 2400 available UWorld questions respondents had completed prior to Step 1 and given options of: <50%, 50-70%, 71-90%, 91-100%, 100% of questions + <50% of questions again, and 100% of questions + >50% of questions again. Responses were grouped into Incomplete (<70% of questions), Near Complete (71-90% of questions), Complete (91-100% of questions), and Complete+ (>100% of questions).
- Several questions were asked about the Maimonideck itself: Was it effective?; Was it efficient?; Would they recommend it to a fellow student?

## NBME Results (n=44)

- There was no significant difference between MCAT scores of users and non-users, or between the Primacy, Usage, or Interaction user groups. There was no difference in pre-test question bank completion.
- There was a statistically significant difference in test score when comparing High Use vs Moderate + Low Use ( $p=0.03$ , power=0.71), as well as High + Moderate Interaction vs Low Interaction ( $p=0.02$ , power=0.64)
- The average of all 125 UMass test-takers was 74.5 (11.75). The 44 Maimonideck users averaged 79.4 (9.5); removing them resulted in a non-user average of 71.9.

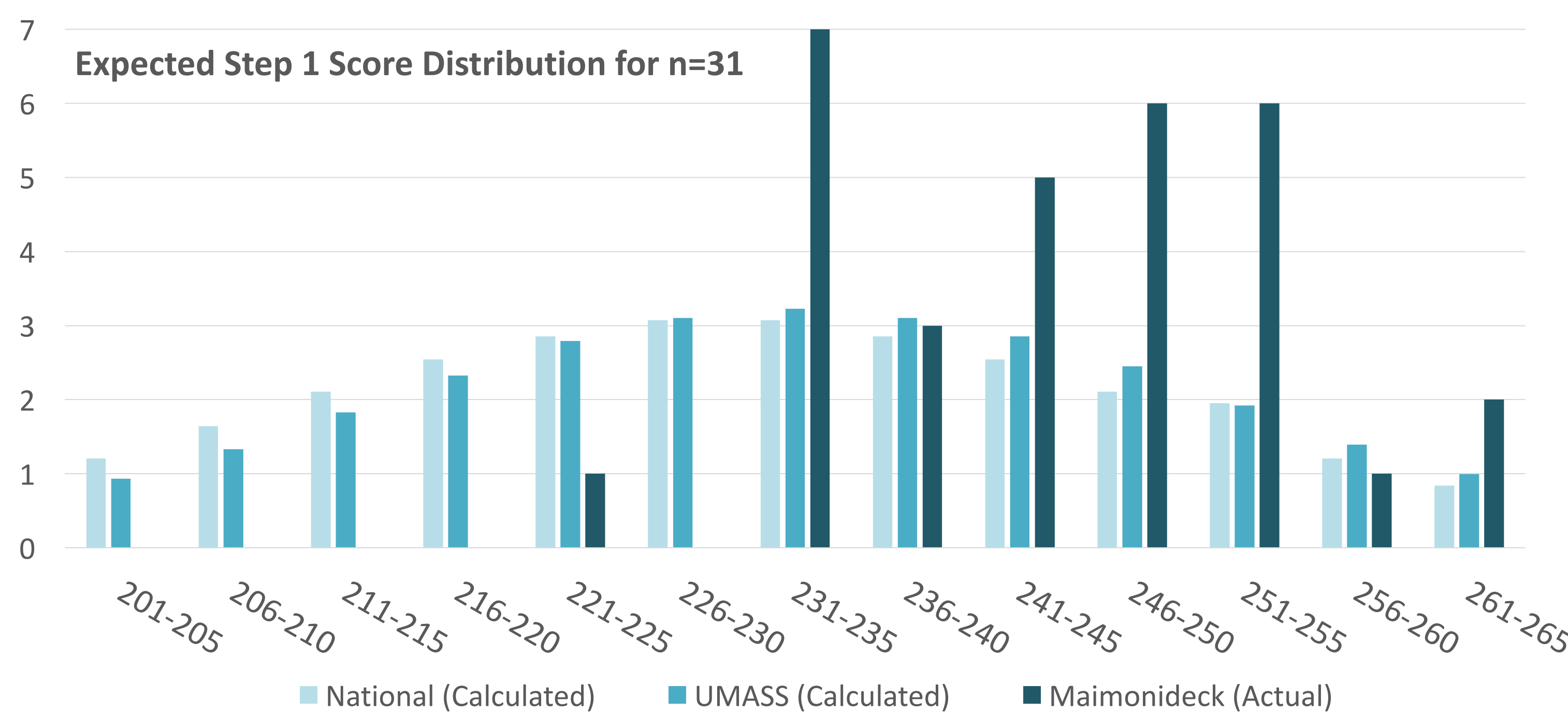
|          | Primacy Grade |          |           | Usage Grade |              |         | Interaction Grade |               |           | All Users  | Non Users  |
|----------|---------------|----------|-----------|-------------|--------------|---------|-------------------|---------------|-----------|------------|------------|
|          | Primary Tool  | Neutral  | Reference | High Use    | Moderate Use | Low Use | High Edit         | Moderate Edit | Low Edit  |            |            |
| n        | 23            | 5        | 16        | 15          | 19           | 10      | 13                | 14            | 17        | 44         | 81         |
| MCAT (n) | 31.6 (20)     | 31.5 (4) | 32.1 (14) | 32.4 (13)   | 31.2 (17)    | 32 (8)  | 31.9 (9)          | 32.3 (14)     | 31.3 (15) | 31.84 (38) | 31.76 (50) |
| NBME     | 81            | 81.4     | 76.4      | 83.6        | 78.1         | 75.6    | 81.6              | 78.8          | 75.3      | 79.4       | 71.9       |
| p=0. 24* |               |          |           | p=0.03†     |              |         | p=0. 41‡          |               |           |            |            |
|          | p=0. 12**     |          |           | p=0. 15††   |              |         | p=0.02‡‡          |               |           |            |            |

Primacy of Maimonideck: \*Primary vs (Secondary & Reference); \*\* (Primary & Secondary) vs Reference

Usage Grade: †High use vs (Moderate use & Low use); †† (High use & Moderate use) vs Low use

Editing Grade: ‡High edit vs (Moderate edit & Low edit); ‡‡ (High edit & Moderate Edit) vs Low edit

|                          | Primary   |    | Secondary |   | Reference |   | All Users | Non Users |
|--------------------------|-----------|----|-----------|---|-----------|---|-----------|-----------|
| N                        | 20        |    | 5         |   | 6         |   | 31        | 90        |
| UWorld Complete + (n)    | 6         | 12 | 2         | 4 | 2         | 3 | 10        |           |
| UWorld Complete (n)      | 6         |    | 2         |   | 1         |   | 9         |           |
| UWorld Near-Complete (n) | 6         |    | 0         |   | 1         |   | 7         |           |
| UWorld Incomplete (n)    | 2         | 8  | 1         | 1 | 2         | 3 | 5         |           |
| MCAT (n)                 | 31.7 (17) |    | 30 (4)    |   | 31.7 (6)  |   | 31.4 (27) | 31.9 (61) |
| Step 1 Score             | 245.4     |    | 244.4     |   | 238.2     |   | 243.8     | 229.3     |



## Step 1 Results (n=31)

- There was no significant difference between MCAT scores of users and non-users, or between the MCAT scores or UWorld completion between the Primacy, Secondary, and Reference tool groups among users.
- The average of all UMass test-takers was 233 (19). The 31 Maimonideck users averaged a 243.8 (9.7). Removing them resulted in a non-user average of 229.3
- Nearly all (93%) of respondents agreed or strongly agreed that the Maimonideck was an effective learning tool, with 74% agreeing or strongly agreeing that it was a time-efficient learning tool. 91% of respondents would recommend the Maimonideck to a classmate.
- A selection of free-text responses are copied below:

“When I finally was able to go through all of First Aid, I already knew almost all of it because I had fully reviewed the Maimonideck throughout the year.”

“Though it didn’t work well for every subject (i.e., Biostats mostly) I would say that Maimonideck was definitely the best tool I had in study and learning material for Step 1. I don’t think I would’ve done nearly as well with class or the exam without it. It’s different than other shared Anki decks because it’s not just a word recall.”

## Discussion

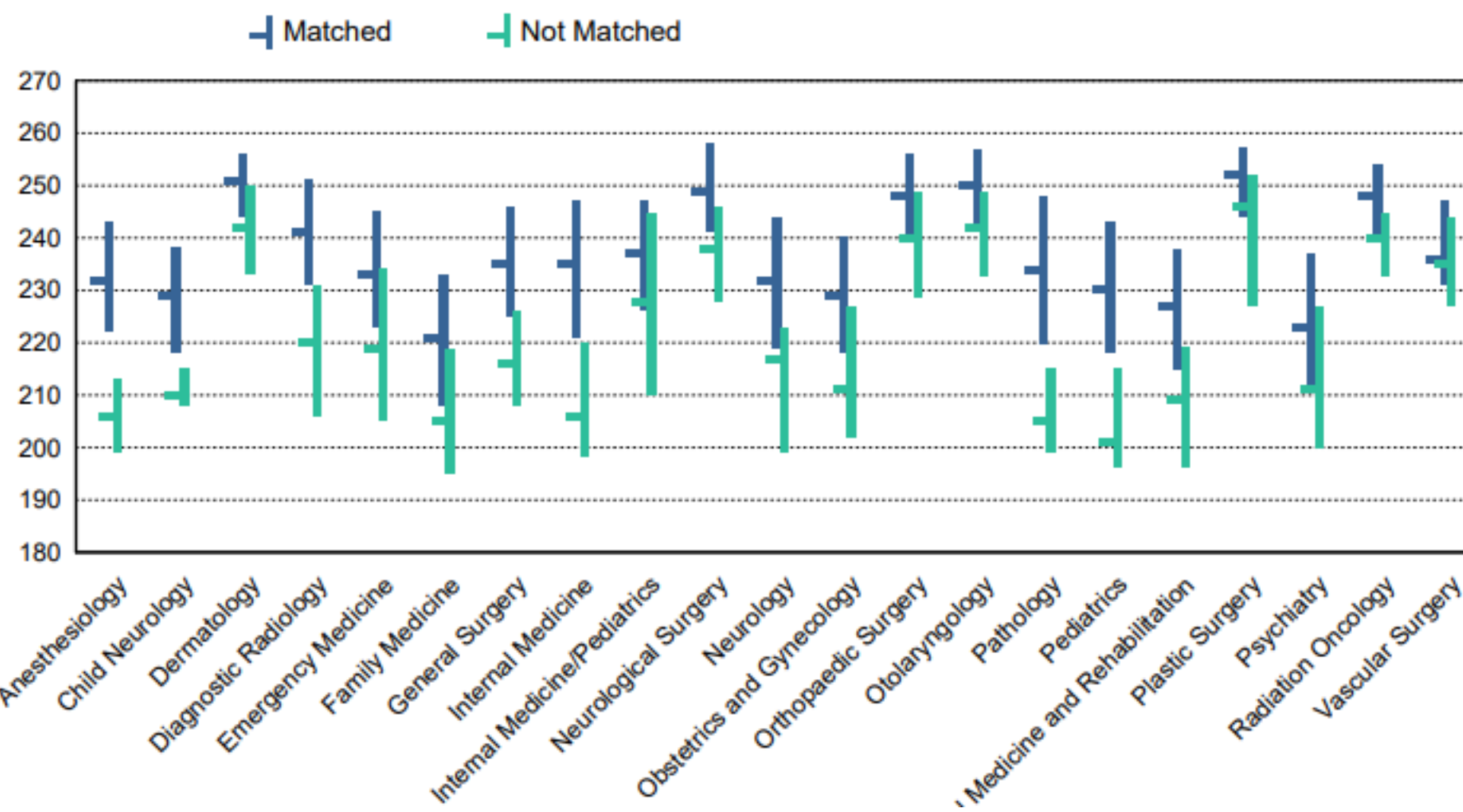
There has been a proliferation of Step 1 resources, but few address a fundamental difficulty of preparing for the exam: study scheduling. The Maimonideck – a resource created by two medical students – was designed for straightforward, daily use over the entirety of the second-year curriculum, and appears to be effective in that capacity. The 31 users averaged a 243.8 on Step 1 (~75<sup>th</sup> percentile of all test takers), compared to the 229.3 of non-responding UMMS students (~50<sup>th</sup> percentile of all test takers).

A second iteration of this study is underway for the Class of 2020 and will seek to address the primary weaknesses of the current project, namely:

- Study power: We anticipate equal if not increased interest from the 150 students in the Class of 2020
- Improved Control Data: Obtain UWorld and Step 1 data from non-users
- Selection Bias: Evaluate both users and controls for specialty preference and target score as possible markers for self-selection

For better or worse, the stakes of Step 1 are not decreasing. The value of this exam as a measuring stick for residency qualification, and its effect on medical education in general, are very much up for debate. What is not, however, is its importance to current medical students with specific specialty and geographic preferences. If students are to weather this high stakes exam they should have the resources – both in content and delivery – to do so as effectively and efficiently as possible.

### USMLE Step 1 Scores of U.S. Allopathic Seniors by Preferred Specialty and Match Status



**Acknowledgements:** Special thanks to Judy Savageau, MPH and Christine Woolf, PhD of UMass and B. Price Kerfoot, MD of BWH for their consultation and encouragement.