ORIGINAL RESEARCH

Implementation of Spaced Repetition by First-Year Medical Students: a Retrospective Comparison Based on Summative Exam Performance

Anila Mehta[1] · Nathaniel Brooke[1] · Anessa Puskar[1] · Mary Clare Crochiere Woodson[1] · Barbara Masi[2] · Robert C. Wallon[1] · Donald A. Greeley[1]

Abstract

Many medical students use spaced repetition as a study strategy to improve knowledge retention, and there has been growing interest from medical students in using flashcard software, such as Anki, to implement spaced repetition. Previous studies have provided insights into the relationship between medical students' use of spaced repetition and exam performance, but most of these studies have relied on self-reports. Novel insights about how medical students use spaced repetition can be gleaned from research that takes advantage of the ability of digital interfaces to log detailed data about how students use software. This study is unique in its use of data extracted from students' digital Anki data files, and those data are used to compare study patterns over the first year of medical school. Implementation of spaced repetition was compared between two groups of students who were retrospectively grouped based on average performance on three exams throughout the first year of medical school. Results indicate that students in the higher scoring group studied more total flashcards and implemented spaced repetition via Anki earlier in the year compared to the lower scoring group. These findings raise the possibility that implementing spaced repetition as a study strategy early in medical school may be related to improved knowledge retention and exam performance. Additional research should be performed at more sites to further examine the relationship between spaced repetition implementation and exam performance.

Keywords Spaced repetition · Exam performance · Medical students · Flashcards · Study methods

Introduction

Spaced repetition digital applications are a popular study tool used by medical students to study basic sciences material [1–6]. Preclinical study time in medical school is critical to students as this is the time to build foundational medical knowledge to be able to perform well on exams and in clinical rotations and eventually to provide the best possible patient care.

One spaced repetition application that is commonly used by medical students is Anki, which is a free, customizable flashcard program [7]. This spaced repetition application allows students to self-test while changing the testing frequency of flashcards based on self-rating of flashcard difficulty. If a student rates a flashcard as answered incorrectly, the card will be shown more frequently to reinforce the material. These increasing intervals of time are intended to strengthen memory and reduce forgetting [2, 8, 9]. To keep track of learners' progress for the spaced repetition algorithm, the Anki software maintains a log of every time a user reviews a flashcard and stores this log in a standard SQLite

✉ Anila Mehta
  anilam2@illinois.edu

  Nathaniel Brooke
  nbrooke2@illinois.edu

  Anessa Puskar
  tawakol2@illinois.edu

  Mary Clare Crochiere Woodson
  mcc7@illinois.edu

  Barbara Masi
  barbm602@gmail.com

  Robert C. Wallon
  rwallon2@illinois.edu

  Donald A. Greeley
  donald.greeley@carle.com

[1] Carle Illinois College of Medicine, Urbana-Champaign, IL, USA

[2] Icahn School of Medicine at Mount Sinai, New York City, NY, USA

database file on the user's computer. Each entry in this log includes the contents of the flashcard, the date and time that the card was reviewed, how long the student viewed the card, and how the student rated the difficulty of the card.

Several studies have identified a positive relationship between spaced repetition usage and exam performance [10–14]. For example, previous research indicated that increased Anki usage was associated with higher USMLE Step 1 scores [1]. Another study found that completing an additional 1700 unique Anki flashcards was associated with an increase of one additional point on USMLE Step 1 [2]. A similarity in the methods used in these studies is that they utilized self-reports of Anki usage rather than logs of user activity from the software database, and a limitation of self-report is that it is subject to recall bias. In the literature, there have been very few studies of spaced repetition implementation that used flashcard data logged directly via software. One study that did utilize a data log-based approach involved collecting flashcard usage data generated by residents in obstetrics and gynecology. This study created a pre-made flashcard deck for the residents, and the flashcards were in a specific format that included a question stem and multiple-choice answers. This study found that using the flashcards to study correlated with improvement in scores, but the results were not statistically significant [15].

Based on the literature, we identified an important area in which a unique contribution could be made. There is a need to study spaced repetition implementation using data extracted directly from software in the context of investigating learners' usage patterns of their own flashcards and pre-made decks. Therefore, in this study, we explored whether spaced repetition usage patterns are associated with first-year summative exam performance and whether early implementation of spaced repetition is associated with positive performance on successive exams.

The goal of our study was to identify and compare spaced repetition usage patterns and implementation among first-year medical students retrospectively grouped by exam performance. Specifically, we addressed the following research question: in what ways and to what extent does spaced repetition implementation through Anki usage vary between a group of students with above median exam performance compared to a group of students with below median exam performance?

Methods

Participants and Materials

This study was conducted at Carle Illinois College of Medicine (CIMED) in Urbana-Champaign, IL, and was part of an IRB-approved protocol (#187787). Forty-eight students who completed the first year of the preclinical curriculum in academic year 2020–2021 were invited to take part in the study, and 29 students participated (60% response rate). Participants used flashcards of their own selection (i.e., a flashcard deck was not provided to students in this study). This study did not track which of the many premade decks available online were used by students, or if students created their own flashcards to study.

Context

The preclinical phase at CIMED includes an organ systems–based curriculum. The curriculum begins with a Foundations course which includes material on genetics, biochemistry, statistics, and histology. This course is subsequently followed by organ system courses. In the approximately 11-month timeframe of data included in this study, these courses included Foundations, Cardiology, Respiratory, Renal, Clinical Neuroscience, and Musculoskeletal and Integumentary.

Data Collection

Potential participants were asked to submit a copy of their Anki software usage database via Qualtrics Software in June 2021. A Python program was developed to access the usage log within each Anki database and extract the record of every time the student studied a flashcard. A subset of flashcards that the student had previously learned and answered correctly when the card was most recently studied were marked as review cards. Exact parameters determining how many times in a row and how far apart in time the card must have been answered correctly to be considered a review card are set within the Anki software itself and vary slightly between users based on personal preference. We used the Python program to aggregate the extracted flashcard study records for each day to compute a daily number of cards studied, daily number of review cards studied, and daily total time spent studying. The program processed each database file separately and output results in a CSV file for further processing and analysis.

All eligible study participants completed three summative exams generated via the National Board of Medical Examiners (NBME) Customized Assessment System over the course of the year prior to Anki database submission. Exams were completed by students in approximately 3-month intervals as they completed the first-year curriculum, and scores on the three exams were averaged for each student. Then participants were retrospectively assigned into one of two groups for post hoc analyses based on whether their average exam score was above or below the median of the entire class of 48 students. These groups will henceforth be referred to as above median group (AMG) and below median group (BMG). There were 15 participants in AMG and 10 participants in BMG.

Data Analysis

The data extracted from the Anki usage logs were aggregated into the two exam performance groups for further analysis. Several variables were generated for AMG and BMG, including total number of all flashcards studied over the data collection period and weekly averages, the total number of review flashcards studied, the rate at which all flashcards were studied (seconds per card), the number of days using Anki, and the number of days skipped while using Anki.

When calculating total numbers of flashcards studied, duplicates were not removed, so if a student studied the same card multiple times, every time that card was studied is included in the total. For example, if one student studied 10 cards once each, and another student studied 2 cards 5 times each, both students' total flashcards studied would be counted as 10. This approach to flashcard counts is used throughout the entire study whenever a number of flashcards is discussed.

When calculating total numbers of flashcards for each student over the study period, if no flashcards were studied on a respective date or if the student had not started using Anki yet, these scenarios were counted as "0" flashcards for that date. However, when calculating weekly averages, if the student had not yet started using Anki, they were excluded from the calculation; students who started using Anki and then studied no flashcards on a given date were counted as having studied "0" flashcards on that date. This decision was made to differentiate between students who had started Anki and skipped days versus student who had not yet started using the software.

The aforementioned variables were aggregated in the statistical program Statistical Package for the Social Sciences (SPSS). Mann–Whitney tests were used to compare means between AMG and BMG for the variables of interest. The Mann–Whitney test was the most appropriate approach to comparing means given the relatively small group sizes because it provided a non-parametric method to compare means (as opposed to a parametric method via t-test). The graph was created in Microsoft Excel and imported to Adobe Illustrator for further formatting.

Results

Out of all respondents ($n = 29$), 26 used Anki at some point during their first year of medical school and submitted a database file (90% of respondents). One submitted database file was corrupted and could not be opened, so it was excluded from the study, resulting in 25 valid database files included in the study.

In Fig. 1, the 7-day average of the number of total daily flashcards is shown for each exam performance group over 41 time periods. Note that students are not included in the weekly averages in Fig. 1 until they started using Anki. The exam dates and vacation time are shown as vertical bars on the figure. The courses in the curriculum are depicted across the top of the figure.

In Table 1, cumulative flashcard use variables are compared between exam performance groups. The comparison shows several statistically significant differences in usage patterns between the exam performance groups. Participants in the above median exam performance group (AMG) studied
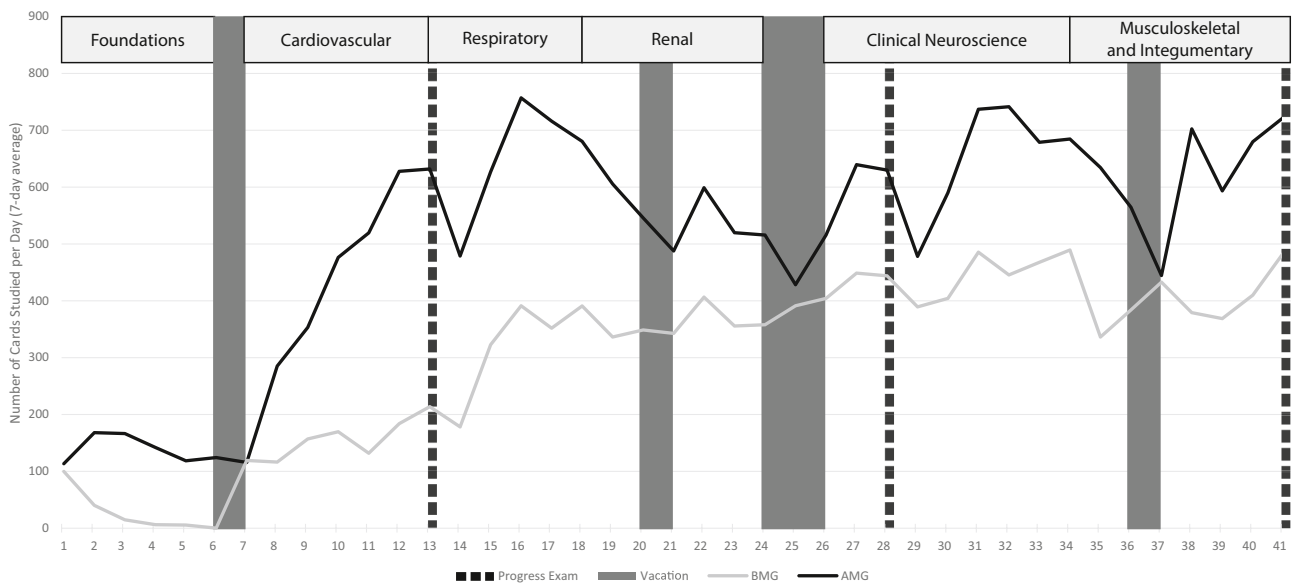

Fig. 1 Seven-day average of the total daily of all flashcards studied divided into two exam performance groups (AMG and BMG)

Table 1 Comparison of various Anki usage parameters between the students in AMG and BMG

Variable	Group	Mean	Std. deviation	*p*-value
Number of all flashcards studied	BMG	81,209.30	48,162.26	0.016*
	AMG	146,144.60	65,614.95	
Number of "review" flashcards studied	BMG	46,620.60	30,213.39	0.036*
	AMG	88,887.60	50,172.36	
Rate of which flashcards were studied (seconds/card)	BMG	12.23	3.53	0.091
	AMG	9.91	2.05	
Number of days Anki used	BMG	193.00	59.78	0.004**
	AMG	248.20	47.04	
Number of days since starting to use Anki	BMG	237.90	49.58	0.007**
	AMG	278.00	18.00	
Number of days skipped since starting Anki	BMG	44.90	32.75	0.189
	AMG	29.80	46.59	
Number of flashcards studied per day Anki was used	BMG	388.96	165.17	0.041*
	AMG	565.46	213.89	

*$p < 0.05$; **$p < 0.01$

more total flashcards compared to their peers in the below median exam performance group (BMG). Additionally, participants in the AMG studied more review flashcards than those in the BMG. Participants in the AMG also started studying flashcards earlier than those in the BMG. Participants in the AMG had significantly higher number of days of Anki use and a significantly higher number of days since starting Anki compared to the BMG, measured as days between first Anki use and the end of data collection. Participants in the AMG studied more total flashcards and used Anki more days than their peers in the BMG, and when controlling for the number of days that the software was used, participants in the AMG still studied significantly more flashcards per day of Anki use compared to their peers in the BMG. Of note, no significant difference between groups was seen in the rate at which students studied flashcards, measured in seconds per card or in the number of days that the user did not use Anki after their initial use of the software.

Discussion

In this study, we addressed the ways and the extent to which spaced repetition implementation through Anki compares between students in AMG and BMG. The first-year medical students in this study relied heavily on the digital spaced repetition application, Anki. However, the two exam performance groups (AMG and BMG) used the application to different extents. We compared various usage parameters in several areas including quantity of flashcards, the rate at which flashcards were studied, and the timeline of implementation of the application.

The AMG studied a significantly higher number of all flashcards and of review flashcards throughout the year

compared to the BMG. Although Fig. 1 shows increases in total number of cards used for both AMG and BMG in the time periods leading up to exams, there was less evidence of massed repetition ("cramming") for students in AMG as they studied significantly more flashcards per day compared to students in BMG. Interestingly, Fig. 1 shows that students in AMG generally slightly decreased the number of flashcards completed in time intervals leading up to and during vacation, while students in BMG generally maintained a similar number of flashcards completed in time intervals leading up to and during vacation.

Even though the AMG completed a higher number of all flashcards and review flashcards than the BMG, this was not attributable to rushing through the flashcards because there was no significant difference in the studying rate for all flashcards and review flashcards between the two groups. This indicates that the AMG spent more overall time utilizing Anki compared to their peers due to their larger number of flashcards completed. This suggests that a relatively consistent, high amount of study time early on in medical school may be associated with higher exam performance among medical students. This finding aligns with a previous study that reported that USMLE Step 1 scores were significantly higher for students who studied 8–11 h per day compared with those who studied 0–3 h per day. However, there was no significant improvement in USMLE Step 1 scores observed in students who studied more than 11 h per day. This indicated that additional time spent studying enhanced exam performance up until a certain limit [16].

The AMG started using the spaced repetition application earlier than the BMG. Early on in the first year prior to the first exam, the AMG increased their weekly average of flashcards significantly more than BMG as seen in Fig. 1. The AMG used Anki on more days throughout the first

year of medical school significantly more than their peers. They utilized Anki on more days out of the year and started using the application earlier. However, there was no difference between groups in the number of days skipped using the software once started. This finding suggests that both groups were consistent in their spaced repetition usage once they started implementing the spaced repetition application. A previous study on medical students who were learning medical pharmacology reported that the optimal strategy is to implement flashcards in the medical curricula early on during preclinical years in order to support memorization, to understand concepts, and to provide a structure for long-term revision [17]. The findings from our study align with this finding and can possibly be explained by the use of flashcards early on in the year, allowing for students to revise and modify their flashcards as directed by their own personal studying.

Overall, the results of this study indicate that students who performed higher on average than their peers on summative exams implemented spaced repetition early in their first year of medical school and continued to use this study strategy consistently. These findings that showed a positive relationship between implementation of spaced repetition and exam performance align with results of previous studies in the literature that showed a positive relationship between implementation of spaced repetition and improved performance on USMLE Step 1 [1, 2]. Other prior research found that retrieval practice and spaced repetition were two study strategies that were consistently utilized by the majority of high-performing students included in the study [18]. Our findings align with this finding, and this suggests that regular use of self-testing throughout studying may be a factor that contributes to high performance on exams. The results of this study suggest that implementation of spaced repetition may help medical students improve recall of previously learned material throughout the first year of medical school.

Limitations

A limitation of this study is the relatively small sample size and the context of one medical school cohort at a single institution, which may not be representative of medical students across the country. With a 60% response rate of the study, we are unable ascertain if the remaining students in the class utilized Anki in their studying or did not. Another limitation is that we were unable to control for the medical knowledge that students had prior to beginning medical school. We assumed that our participants had a similar baseline knowledge prior to medical school. It is important to note that there may be systematic differences between students in AMG and BMG that we were unable to elicit in this study that may explain some of the results we obtained. For example, some students'

study habits may have been influenced by friends, and our data sources cannot account for potential social influences related to implementation of spaced repetition. Furthermore, we did not assign a pre-made deck of flashcards for this study so we are unable to discern whether the design of specific cards may have contributed to the results. It is also important to recognize that using a spaced repetition flashcard application may not be an effective self-study tool for every medical student. It is important to note that there are additional study strategies and tools that students utilize aside from the spaced repetition application that was the focus of this paper. Due to the inability to account for all aspects of participants' studying, the association between spaced repetition implementation and exam performance should not be interpreted as a causal relationship. Additionally, in this study, the average scores across the three exams were utilized to protect participant confidentiality, and individual students in the study may have scored above or below the median on individual exams. This limits the conclusions that can be drawn about changes in flashcard usage over time from this study. Nevertheless, this study suggests that implementing spaced repetition early-on in the first year of medical school and remaining consistent throughout the year at a high amount could be beneficial in improving students' recall and exam performance.

Conclusion

A spaced repetition digital application, Anki, was widely used by the first-year medical students who participated in this study. Students in the AMG implemented a spaced repetition study pattern early in the curriculum and studied more flashcards overall. There are many other factors among students that could contribute to higher exam performance such as various baseline medical knowledge prior to medical school, other study materials utilized by students, or differences in test-taking strategies. These findings suggest that much remains to be learned from investigating medical students' spaced repetition implementation and how it may relate to exam performance. This study is an initial step in exploring the relationship between spaced repetition implementation and exam performance using log data, and we suggest that additional research utilizing log data from study tools should be performed across larger cohorts of medical students and at other medical schools to further explore this relationship.

Availability of Data and Material Not available.

Declarations

Ethics Approval and Consent to Participate Authors consent.

Consent for Publication Authors consent.

Competing Interests The authors declare no competing interests.

References

1. Lu M, Farhat JH, Beck Dallaghan GL. Enhanced learning and retention of medical knowledge using the mobile flash card application Anki. Med Sci Educ. 2021;31(6):1975–1981. Published 1 Sep 2021. https://doi.org/10.1007/s40670-021-01386-9.
2. 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. 18 Nov 2016;:]. Perspect Med Educ. 2015;4(6):308–13. https://doi.org/10.1007/s40037-015-0220-x.
3. Wu JH, Gruppuso PA, Adashi EY. The self-directed medical student curriculum. JAMA. 2021;326(20):2005–6. https://doi.org/10.1001/jama.2021.16312.
4. Harris DM, Chiang M. An analysis of Anki usage and strategy of first-year medical students in a structure and function course. Cureus. 2022;14(3):e23530. Published 27 Mar 2022. https://doi.org/10.7759/cureus.23530.
5. Rana T, Laoteppitaks C, Zhang G, Troutman G, Chandra S. An investigation of Anki flashcards as a study tool among first year medical students learning anatomy. FASEB J. 2020;34(S1):1.
6. Sun M, Tsai S, Engle DL, Holmer S. Spaced repetition flashcards for teaching medical students psychiatry. Med Sci Educ. 2021;31(3):1125–31. Published 6 Apr 2021. https://doi.org/10.1007/s40670-021-01286-y.
7. Anki. https://apps.ankiweb.net/. Accessed 1 June 2021.
8. Lambers A, Talia AJ. Spaced repetition learning as a tool for orthopedic surgical education: a prospective cohort study on a training examination. J Surg Educ. 2021;78(1):134–9. https://doi.org/10.1016/j.jsurg.2020.07.002.
9. Larsen DP, Butler AC, Roediger HL 3rd. Test-enhanced learning in medical education. Med Educ. 2008;42(10):959–66. https://doi.org/10.1111/j.1365-2923.2008.03124.x.
10. Schneid SD, Pashler H, Armour C. How much basic science content do second-year medical students remember from their first year? Med Teach. 2019;41(2):231–3. https://doi.org/10.1080/0142159X.2018.1426845.
11. Dobson J, Linderholm T, Perez J. Retrieval practice enhances the ability to evaluate complex physiology information. Med Educ. 2018;52(5):513–25. https://doi.org/10.1111/medu.13503.
12. Roediger HL 3rd, Butler AC. The critical role of retrieval practice in long-term retention. Trends Cogn Sci. 2011;15(1):20–7. https://doi.org/10.1016/j.tics.2010.09.003.
13. Karpicke JD, Roediger HL 3rd. The critical importance of retrieval for learning. Sci. 2008;319(5865):966–8. https://doi.org/10.1126/science.1152408.
14. Pumilia CA, Lessans S, Harris D. An evidence-based guide for medical students: how to optimize the use of expanded-retrieval platforms. Cureus. 2020;12(9):e10372. Published 11 Sep 2020. https://doi.org/10.7759/cureus.10372.
15. Tsai S, Sun M, Asbury ML, Weber JM, Truong T, Deans E. Novel spaced repetition flashcard system for the in-training examination for obstetrics and gynecology [published correction appears in Med Sci Educ. 2021;31(4):1559]. Med Sci Educ. 2021;31(4):1393–99. Published 19 May 2021. https://doi.org/10.1007/s40670-021-01320-z.
16. Kumar AD, Shah MK, Maley JH, Evron J, Gyftopoulos A, Miller C. Preparing to take the USMLE Step 1: a survey on medical students' self-reported study habits. Postgrad Med J. 2015;91(1075):257–61. https://doi.org/10.1136/postgradmedj-2014-133081.
17. Jape D, Zhou J, Bullock S. A spaced-repetition approach to enhance medical student learning and engagement in medical pharmacology. BMC Med Educ. 2022;22(1):337. Published 2 May 2022. https://doi.org/10.1186/s12909-022-03324-8.
18. Landoll RR, Bennion LD, Maggio LA. Understanding excellence: a qualitative analysis of high-performing learner study strategies. Med Sci Educ. 2021;31(3):1101–8. Published 25 Mar 2021. https://doi.org/10.1007/s40670-021-01279-x.