Medical Science Educator (2021) 31:1125–1131
https://doi.org/10.1007/s40670-021-01286-y

ORIGINAL RESEARCH

Spaced Repetition Flashcards for Teaching Medical Students Psychiatry

Michael Sun[1] · **Shelun Tsai**[2] · **Deborah L Engle**[3] · **Shelley Holmer**[4]

Accepted: 29 March 2021 / Published online: 6 April 2021

Abstract

Objective Retrieval practice, often using electronic flashcards, is increasingly utilized among medical students for self-study. In this study, the authors evaluated usage and satisfaction with electronic flashcards based on a medical school psychiatry curriculum.

Methods First-year medical students at one institution consented to participate and received access to a set of pre-made flashcards. Surveys were distributed that collected demographic information along with measures of prior performance, test anxiety, and prior experience with electronic flashcards. The total number of flashcard reviews and time spent on the platform for each student were collected using statistics internally generated by the platform. Each student's final exam score was also collected.

Results A total of 114 of 129 first year medical students (88%) consented to participate, and 101 students were included in the final analysis. Fifty-eight (56%) were flashcard users with a median of 660 flashcards studied over 2.95 h. A total of 87% of flashcard users found the flashcards to be helpful, and 83% of flashcard users would recommend the flashcards to someone else. Flashcard usage was not associated with final exam scores.

Conclusions This novel electronic study resource was well-received by first-year medical students for psychiatric instruction in medical school, though usage was not associated with higher exam scores.

Keywords Retrieval practice · Flashcards · Technology · Psychiatric instruction in medical school

Introduction

Testing in medical education largely takes an evaluative rather than didactic role. Despite several studies demonstrating the effectiveness of test-enhanced learning, or retrieval practice, medical curricula instead often implement testing in the form of high-stakes summative assessments such as the United Stated Medical Licensing Examination (USMLE) Step examinations [1–4].

Recently, learners have increasingly utilized retrieval practice for self-study [5], and both students and faculty have called for greater integration of retrieval practice into the general medical curriculum [6, 7]. Retrieval practice allows learners to benefit from the spacing effect, which states that retention improves with repeated exposure to information over time, and the testing effect, that actively stimulating memory by testing improves performance [1, 8]. By employing these concepts in cognitive psychology, retrieval practice strengthens memory and increases the likelihood of subsequent recall [9].

The principle that retaining or remembering a piece of information depends upon the number of times it is reviewed and the temporal distribution of those reviews was first demonstrated by Ebbinghaus in 1885 [10]. Since that time, much effort has been expended in developing systems to maximize the probability of retaining a piece of information for a given time frame and a limited amount of time to review [11, 12]. Overall, data over thousands of study participants studying

✉ Michael Sun
mysun@mgh.harvard.edu

[1] Department of Psychiatry, Massachusetts General Hospital and McLean Hospital, Boston, MA, USA

[2] Department of Obstetrics and Gynecology, Duke University Medical Center, Durham, NC, USA

[3] Department of Medical Education, Duke University School of Medicine, Durham, NC, USA

[4] Department of Psychiatry, Duke University School of Medicine, Durham, NC, USA

retention intervals from a few seconds to weeks or more demonstrate clear benefits for spaced retrieval practice [8].

In the past few years, medical students have adopted the use of electronic flashcard platforms such as Anki to instantiate retrieval practice [5]. Anki is an open-source electronic flashcard platform that prompts users to self-test while varying testing frequency based on user-perceived question difficulty using a simple algorithm [13]. The platform can be accessed online using a computer or mobile device, and users can synchronize their data across multiple devices. User data can also be extracted from the platform to provide objective measures of use.

Past studies have demonstrated correlations between flashcard usage and factual knowledge in medical education as well as USMLE Step 1 performance [4, 5, 14]. However, no study to date has assessed the effectiveness of retrieval practice for medical education in psychiatry. In addition, the literature has not described self-directed use of electronic flashcards created as a complement to a more traditional, lecture-based medical curriculum. In this study, we evaluated whether retrieval practice using electronic flashcards based on a medical school curriculum for psychiatry was feasible and effective. The primary aim of this study was to evaluate medical students' usage of and satisfaction with this novel study resource. Our secondary aim was to evaluate whether usage was correlated with higher scores in students' exam scores in psychiatry.

Methods

This project was granted exempt status by the university's Institutional Review Board. First year medical students at a single academic institution were eligible for this study. As part of the regular first year medical school curriculum, students took part in a brain and behavior course in January 2019 as part of their second semester curriculum. Only the first week of this month-long course were devoted to psychiatry (January 2nd through January 8th), during which time the students attended lectures and participated in small group sessions. The remainder of the course was dedicated to neuroanatomy and the neurology. On the first day of the brain and behavior course, an educational session was held to provide a tutorial, explain the concepts of retrieval practice and active recall, and obtain written informed consent. Participants filled out an initial survey collecting demographic information. They then received the username and password to a pre-generated Anki account containing 488 flashcards that were created based on parts of the introductory psychiatry text assigned as readings as well as learning objectives defined for the course [15]. Four weeks later, on the same day as the course's final exam (January 28th), a second survey was distributed assessing participant study habits and user experience with the flashcards. Usage statistics between the start of the study and the final exam were obtained from each user's account. Each participant's final exam score was also obtained. The psychiatry portion of the final exam question content was based on National Board of Medical Examiners Psychiatry Shelf-style questions. Final exam question content was not shared with the makers of the flashcards, though topic-level content was obtained after the final exam (see Table 1). Course instructors did not participate in data collection or analysis and had no knowledge of which students were study participants or any study participant's usage statistics over the course of the study.

Flashcards were mostly in the form of short questions and answers covering key terms and relationships in psychiatry, including but not limited to diagnostic criteria, the mental status exam, and psychopharmacology (Fig. 1). Please refer to Table 1 for a breakdown of the study flashcards and final exam questions by topic. Study flashcards varied in difficulty and assessed and reinforced certain definitions, mnemonics, and clinical pearls. While most flashcards were written to prioritize simplicity and 1:1 relationships [16], a minority were designed to provoke more open-ended thought about, for example, the possible components of the appearance part of the mental status exam. Users were instructed to read the question, attempt to answer, and then hit a button to reveal the correct answer. After revealing the correct answer, the user can evaluate the card as "Again", "Hard", "Good", or "Easy". Then, the flashcard is scheduled to be reviewed in the future using an algorithm, with the review date dependent upon the difficulty evaluation. For example, evaluating a flashcard as "Good" increases the interval before the next review date, while self-grading as "Again" resets the interval and prompts the user to review the card again after 1 to 10 min. Selecting "Hard" or "Easy" decreases or increases the review interval, respectively. Flashcard users were defined as participants with >0 flashcard reviews and determined by accessing the internally generated Anki usage data. Participants were excluded from analysis if their Anki flashcard data was unattainable.

Continuous variables were summarized with median and interquartile range (IQR) and compared between groups using Wilcoxon rank sum tests. Chi-square tests were used for categorical variables. A linear regression model was fit to assess the association between the final exam scores and flashcard usage.

Results

A total of 114 of 129 first year medical students (88%) consented to participate in the study. Thirteen students were excluded due to unobtainable flashcard usage. These were students who indicated that they moved the study flashcards

Table 1 Study flashcards and final exam questions by topic

Topic	Number of flashcards	Number of final exam questions
The psychiatric interview	24	0
Childhood disorders (attention deficit hyperactivity disorder, autism spectrum disorder, intellectual disability, tic disorders)	35	0
Psychotic disorders (brief psychotic disorder, delusional disorder, schizoaffective disorder, schizophrenia, schizophreniform disorder)	34	4
Mood disorders (bipolar disorder, depressive disorders)	35	2
Anxiety disorders (panic disorder, agoraphobia, generalized anxiety disorder, phobias, separation anxiety disorder, selective mutism)	28	4
Obsessive disorders (body dysmorphic disorder, obsessive compulsive disorder, trichotillomania)	10	3
Traumatic disorders (acute stress disorder, post-traumatic stress disorder, reactive attachment disorder, adjustment disorder)	23	3
Somatic disorders (conversion disorder, dissociative disorder, factitious disorder, illness anxiety disorder, somatic symptom disorder)	17	2
Eating disorders (anorexia nervosa, bulimia nervosa)	11	0
Parasomnias (obstructive sleep apnea, normal sleep, restless legs syndrome)	20	1
Sexual disorders (gender dysphoria, paraphilic disorders, sexual dysfunction)	16	0
Impulse control disorders (conduct disorder, intermittent explosive disorder, oppositional defiant disorder)	11	0
Substance use disorders	28	1
Neurocognitive disorders (dementia, delirium)	26	0
Personality disorders	15	0
Treatment modalities (pharmacotherapy, psychotherapy, electroconvulsive therapy)	91	0
Other	64	1
Total	488	21

to their own private account and did not agree to share their private review history with the study authors. After excluding these students, there were a total of 101 study participants for whom review data and initial survey responses were available.

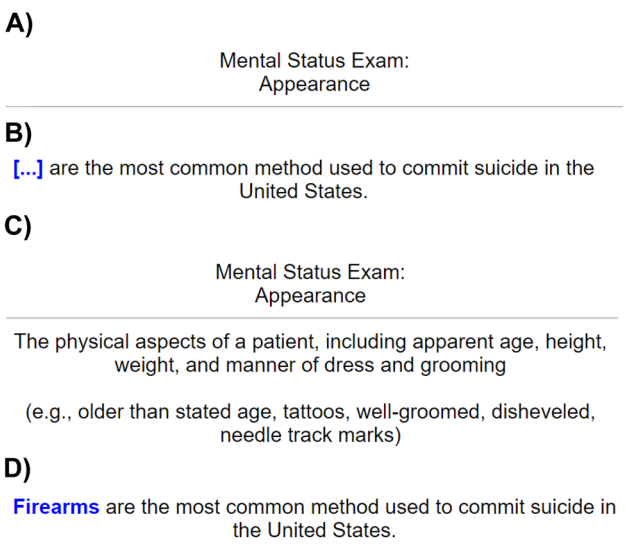

Fig. 1 Example study flashcards. **a**, **b** The question or prompt is presented to the user. **c**, **d** The answer along with any explanatory material is revealed at the user's discretion and subsequently self-graded to determine the interval until the next review

Of the 101 study participants, 58 (56%) were flashcard users and 43 (44%) were flashcard non-users (Table 2). Flashcard users studied a median of 660 flashcards (IQR 125.75–785.75) over a median of 2.95 h (IQR 0.44–3.39). The two groups were relatively similar across a number of demographic variables, including sex, age, race, Medical College Admission Test (MCAT) scores, and college grade point average (GPA). There was also no difference in prior experience with Anki (81% of users; 79% of non-users). Flashcard users did not statistically significantly differ from flashcard non-users with regard to exam scores ($p = 0.54$). The correlation between number of flashcard reviews and final exam score was also not statistically significant ($p = 0.47$).

Fifty-one participants (50%) completed the survey following the final exam, of which 30 (59%) were flashcard users and 21 (41%) were flashcard non-users (Table 3). Survey responders did not differ significantly from survey non-responders with regard to any of the measured variables.

Of the post-exam survey respondents, there was no difference between flashcard users and flashcard non-users regarding their interest in pursuing psychiatry as a career (23% of users; 24% of non-users). Reported total study time

Table 2 Demographics of participants

	Flashcard non-user ($N = 43$)	Flashcard user ($N = 58$)	p value
Age (years)			0.39
Median (IQR)	23.7 (23–25)	23.4 (23–24)	
Sex			0.16
Male	14 (32.5%)	27 (46.6%)	
Female	29 (67.4%)	31 (53.4%)	
Race/ethnicity			0.51
White	21 (48.8%)	24 (41.4%)	
Hispanic	1 (2.3%)	2 (3.4%)	
African American	5 (11.6%)	9 (15.5%)	
Asian	10 (23.3%)	18 (31.0%)	
American Indian or Alaskan Native	0 (0%)	2 (3.4%)	
Mixed	6 (14.0%)	3 (5.2%)	
Prior Anki experience	34 (79.1%)	47 (81.0%)	0.81
Best MCAT score			0.17
519–528	20 (46.5%)	18 (31.0%)	
509–518	22 (51.1%)	35 (60.3%)	
499–508	1 (2.3%)	5 (8.6%)	
< 499	0 (0%)	0 (0%)	
Undergraduate GPA			0.37
3.9–4.0	16 (37.2%)	26 (44.8%)	
3.8–3.89	8 (18.6%)	15 (25.9%)	
3.7–3.79	10 (23.3%)	7 (12.1%)	
3.6–3.69	4 (9.3%)	2 (3.4%)	
3.5–3.59	3 (7.0%)	5 (8.6%)	
3.4–3.49	1 (2.3%)	1 (1.7%)	
3.3–3.39	1 (2.3%)	1 (1.7%)	
3.2–3.29	0 (0%)	0 (0%)	
3.1–3.19	0 (0%)	1 (1.7%)	
< 3.1	0 (0%)	0 (0%)	

did not differ between the two groups. Flashcard users rated the effect of having access to the flashcards as significantly more helpful for alleviating test anxiety, with 69% reporting either "some decrease" or "a large decrease" in anxiety. A total of 87% of flashcard users found the flashcards to be helpful, and 83% of flashcard users would recommend the flashcards to someone else.

In the final exam survey, the most common primary study method overall was the textbook (45% of survey respondents), followed by the Anki flashcards (31%). Twenty-nine of the 30 (97%) flashcard users reported using Anki as a study resource, and 11 (37%) reported using Anki as their primary study resource. Interestingly, even 12 of the 21 (57%) participants classified as flashcard non-users based on data generated by interacting with the flashcard platform indicated that they had used the study flashcards, and 5 (24%) indicated that Anki was their primary study resource. As a sensitivity analysis, these flashcard non-users were moved to the user group, though after recalculation, similar results to the primary analysis were found.

The final exam survey also solicited free text comments regarding the strengths of the flashcards as well as things that could be changed or improved (Table 4). Study participants highlighted ease of use and concordance with other study materials including the assigned textbook readings. Potential improvements identified by the study participants include reducing wordiness and expanding the flashcards to cover more course content.

Discussion

This study evaluated medical student satisfaction with and usage of a novel electronic flashcard self-study resource. Overall, the flashcards were well received by participants, with almost all flashcard users reporting that they found the flashcards helpful and would recommend the flashcards to someone else. Flashcard use was not associated with final exam scores.

Fifty-eight study participants were flashcard users, as defined by downloading the review data associated with each student's study account which was provided at the beginning of the study. This study likely underestimates the extent of flashcard usage among study participants, as 35% of respondents to the post-exam survey indicated that they had used the study flashcards by transferring them to a private account. This study was only able to access review data from the accounts created for this study, which helps to explain several curious results involving significant Anki use in the "flashcard non-user" group. Additional explanations could involve failure of the data synchronization process after each review session or use of the study flashcards in a group setting in which review history was only attributed to a single account.

Unfortunately, permission to publish final exam performance statistics was not obtained due to the sensitive nature of revealing aggregate medical school performance on standardized exams. Nevertheless, flashcard use was not associated with final exam scores, and the number of flashcard reviews did not correlate with final exam score. Taking the study design into consideration, it is possible that students self-selected the study methods which have worked well for them in the past. Bolstering this consideration is the result that prior experience with Anki did not differ between the two groups—i.e., many students have had experience with Anki in the past but did not choose to utilize this study method for the Brain and Behavior course. It is interesting

Table 3 Post-exam survey results

	Flashcard non-user ($N = 21$)	Flashcard user ($N = 30$)	p value
Interested in psychiatry as a career	5 (23.8%)	7 (23.3%)	0.97
Total study time			0.26
Less than 1 hour	10 (47.6%)	9 (30.0%)	
Between 1 and 2 hours	3 (14.3%)	11 (36.7%)	
Between 2 and 3 hours	5 (23.8%)	8 (26.7%)	
More than 3 hours	3 (14.3%)	2 (6.7%)	
Study resources			
Study flashcards	12 (57.1%)	29 (96.7%)	< 0.01
Lecture materials	10 (47.6%)	17 (56.7%)	0.52
Textbook (Black and Andreasen)	18 (85.7%)	23 (76.7%)	0.32
First aid	9 (42.9%)	9 (30%)	0.28
Primary study resource			
Study flashcards	5 (23.8%)	11 (36.7%)	
Lecture materials	2 (9.5%)	3 (10.0%)	
Textbook (Black and Andreasen)	10 (47.6%)	13 (43.3%)	
First aid	1 (4.8%)	1 (3.3%)	
Other	3 (14.3%)	2 (6.7%)	
Effect of study flashcards on test anxiety			
A large decrease in anxiety	N/A	6 (20.7%)	
Some decrease in anxiety	N/A	14 (48.3%)	
No change in anxiety	N/A	8 (27.6%)	
Some increase in anxiety	N/A	1 (3.4%)	
A large increase in anxiety	N/A	0 (0%)	
Found study flashcards useful	N/A	26 (86.7%)	
Would recommend study flashcards to someone else	N/A	25 (83.3%)	

to note that according to survey results, total study time did not differ between the two groups. It is possible that students that did not spend time on the flashcards substituted more time studying the lecture materials and textbook.

A strength of this electronic flashcard resource is its availability across several types of devices including cell phones, which allows for convenient access for short bursts of reviewing in addition to more formal studying. Another strength is its algorithmic implementation of the spacing and testing effects, greatly reducing user effort to capture the gains in memory associated with these effects. Finally, this electronic flashcard platform collects objective flashcard usage data generated each instance the user interacts with the platform interface. The inclusion of this objective usage

Table 4 Free text comments from post-exam survey

"What were the strengths of this project?"	Examples
Ease of use (× 10)	"I like being able to pick it up whenever and use it to maximize even short walks or bus rides."
Concordance with other course materials (× 9)	"I like that the Anki deck correlated with our chapters from the textbook."
Saves time (× 3)	"Made studying faster about 30-45 min a day, as opposed to several hours if you do the reading"
Active recall (× 2)	"Can actively test myself to see if I was remembering concepts from previous days."
Facilitates learning new material	"It allows me to focus in on most confusing details of every topic and straighten them out."
"What would you like to see changed or improved in future iterations of this project?"	Examples
Cards were too wordy (× 6)	"The cards were a little text heavy. Perhaps parsing down verbiage would be helpful"
Expand content covered (× 2)	"Not just a psychiatry deck but a 'brain and behavior' deck"
Different types of cards	"It would be nice to see different types of Anki cards, including images for visual learning in the deck."

data avoids the recall bias inherent in self-reported usage. However, the limitations of relying on objective usage data include the technological difficulties in securing each user's data as described above.

In contrast to this study, prior studies have demonstrated correlations between implementations of retrieval practice and retention of information. Larsen et al. and Kerfoot et al. studied residents and medical students, respectively, designing modest didactic curricula and randomizing students to a spaced repetition protocol [3, 17]. Both found significant advantages to utilizing retrieval practice. Deng et al. assessed usage of Anki specifically and found that users of Anki performed better on the USMLE Step 1 licensing examination [5].

While this study did not find an association between flashcard use and exam scores, differences include the intended subject of study (specific concepts in psychiatry versus broader coverage for a comprehensive licensing examination), method of usage tracking (objective data collection versus self-report), and duration of flashcard usage (1-month-long brain and behavior course vs 20 weeks or longer of step 1 preparation). Of these, perhaps most relevant is the duration of flashcard usage. Concrete advantages in test scores may not have materialized in the 1-month time frame over which flashcard usage was assessed. For example, in Kerfoot et al., only students who utilized the spaced education protocol for greater than 6 months showed a significant increase in test scores.

A limitation of the study is its limited sample size as well as the significantly lower response rate of the post-exam survey (though survey respondents did not differ significantly from non-respondents). Study measures of satisfaction and test anxiety could be improved by implementing validated measures in future surveys. This study could be extended by following students for a longer duration, given the expected benefits for long-term retention of information. Furthermore, the scope of electronic flashcard complements to the traditional curriculum could extend to the breadth of the medical education curriculum as well as the depth of graduate medical education in psychiatry.

Lastly, survey results show the extent of prior experience with electronic flashcards, with roughly 80% of students having already encountered Anki prior to the start of the study. Anecdotally, many educators are not aware of the rapid uptake of such study strategies among medical students. This study helps to show that students are enthusiastic for and will use study aids that complement their more traditional instruction. Integration of electronic flashcards into the medical curricula can lead to decreased test anxiety in students without burdening students with increased total study time or any potential decrease in test scores. Thus, medical educators should be aware of the potential of electronic flashcards to help meet their goals and the goals of their students, and researchers should be aware of the potential for electronic flashcards to study the discretization, uptake, and retention of information pertinent to medical education.

Data Availability The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

Declarations

Conflict Interests The authors declare no competing interests.

References

1. Roediger HL, Karpicke JD. The power of testing memory: basic research and implications for educational practice. Perspect Psychol Sci. 2006;1(3):181–210. https://doi.org/10.1111/j.1745-6916.2006.00012.x.
2. Larsen DP, Butler AC, Roediger HL. Test-enhanced learning in medical education. Med Educ. 2008;42(10):959–66. https://doi.org/10.1111/j.1365-2923.2008.03124.x.
3. Larsen DP, Butler AC, Roediger HL. Repeated testing improves long-term retention relative to repeated study: a randomised controlled trial. Med Educ. 2009;43(12):1174–81. https://doi.org/10.1111/j.1365-2923.2009.03518.x.
4. Larsen DP, Butler AC, Roediger HL. Comparative effects of test-enhanced learning and self-explanation on long-term retention. Med Educ. 2013;47(7):674–82. https://doi.org/10.1111/medu.12141.
5. Deng F, Gluckstein JA, Larsen DP. Student-directed retrieval practice is a predictor of medical licensing examination performance. Perspect Med Educ. 2015;4(6):308–13. https://doi.org/10.1007/s40037-015-0220-x.
6. Parmelee D, Roman B, Overman I, et al. The lecture-free curriculum : Setting the stage for life-long learning : AMEE Guide No . 135. Med Teach. 2020;0(0):1–8. https://doi.org/10.1080/0142159X.2020.1789083.
7. Ahmed OM, Juthani PV, Green ML, et al. Optimizing preclinical learning with retrieval practice: A call to action. Med Teach. 2020;1–3. https://doi.org/10.1080/0142159X.2020.1788212.
8. Cepeda NJ, Pashler H, Vul E, Wixted JT, Rohrer D. Distributed practice in verbal recall tasks: a review and quantitative synthesis. Psychol Bull. 2006;132(3):354–80. https://doi.org/10.1037/0033-2909.132.3.354.
9. Karpicke JD. Retrieval-Based Learning: A Decade of Progress. Vol 2. Second Edi. Elsevier. 2016. https://doi.org/10.1016/B978-0-12-809324-5.21055-9..
10. Ebbinghaus H. Memory: A Contribution to Experimental Psychology. 1885.
11. Dempster FN. Spacing effects and their implications for theory and practice. Educ Psychol Rev. 1989;1(4):309–30. https://doi.org/10.1007/BF01320097.
12. Lindsey RV, Shroyer JD, Pashler H, Mozer MC. Improving students' long-term knowledge retention through personalized review. Psychol Sci. 2014;25(3):639–47. https://doi.org/10.1177/0956797613504302.
13. https://apps.ankiweb.net/ Accessed Mar 1, 2020. https://apps.ankiweb.net.
14. Schmidmaier R, Ebersbach R, Schiller M, Hege I, Holzer M, Fischer MR. Using electronic flashcards to promote learning in medical students: retesting versus restudying. Med Educ.

2011;45(11):1101–10. https://doi.org/10.1111/j.1365-2923.2011.04043.x.
15. Black DW, Andreasen NC. Introductory Textbook of Psychiatry. 6th ed. Am Psych Pub; 2014.
16. Wozniak P. Effective learning: Twenty rules of formulating knowledge. Pub 1999. https://www.supermemo.com/en/archives1990-2015/articles/.
17. Kerfoot BP, DeWolf WC, Masser BA, Church PA, Federman DD. Spaced education improves the retention of clinical knowledge by medical students: a randomised controlled trial. Med Educ. 2007;41(1):23–31. https://doi.org/10.1111/j.1365-2929.2006.02644.x.