ORIGINAL RESEARCH

Novel Spaced Repetition Flashcard System for the In-training Examination for Obstetrics and Gynecology

Shelun Tsai[1] · Michael Sun[4] · Melinda L. Asbury[2,4] · Jeremy M. Weber[3] · Tracy Truong[3] · Elizabeth Deans[1]

Abstract

Introduction Electronic flashcards allow repeated information exposure over time along with active recall. It is increasingly used for self-study by medical students but remains poorly implemented for graduate medical education. The primary goal of this study was to determine whether a flashcard system enhances preparation for the in-training examination in obstetrics and gynecology (ob-gyn) conducted by the Council on Resident Education in Obstetrics and Gynecology (CREOG).

Methods Ob-gyn residents at Duke University were included in this study. A total of 883 electronic flashcards were created and distributed. CREOG scores and flashcard usage statistics, generated internally by interacting with the electronic flashcard system, were collected after the 2019 exam. The primary outcome was study aid usage and satisfaction. The secondary outcome was the impact of flashcard usage on CREOG exam scores.

Results Of the 32 residents, 31 (97%) participated in this study. Eighteen (58%) residents used the study's flashcards with a median of 276 flashcards studied over a median of 3.7 h. All of the flashcard users found the study aid helpful, and all would recommend them to another ob-gyn resident. Using the flashcards to study for the 2019 CREOG exam appeared to correlate with improvement in scores from 2018 to 2019, but did not achieve statistical significance after adjusting for post-graduate year (beta coefficient = 10.5; 95% confidence interval = $-0.60, 21.7$; $p = 0.06$).

Discussion This flashcard resource was well received by ob-gyn residents for in-training examination preparation, though it was not significantly correlated with improvement in CREOG scores after adjusting for post-graduate year.

Keywords Flashcards · Retrieval-based practice · Spaced repetition · In-training examination

Introduction

Electronic flashcards have become an increasingly popular self-study resource for medical education [1–6]. This method of retrieval-based practice strengthens memory and subse-

✉ Shelun Tsai
stsai6326@gmail.com

1 Department of Obstetrics and Gynecology, Duke University Medical Center, Durham, NC, USA

2 Department of Psychiatry, University of North Carolina School of Medicine, Chapel Hill, NC, USA

3 Department of Biostatistics and Bioinformatics, Duke University Medical Center, Durham, NC, USA

4 Department of Psychiatry, Duke University Medical Center, Durham, NC, USA

quent information retention by employing the spacing effect, which states that retention improves with repeated exposure to information over time, and active recall, which stresses the importance of actively stimulating memory instead of passively absorbing information [7, 8]. Employment of these cognitive psychology principles aim to interrupt the natural process of forgetting, and instead improve short-term and long-term knowledge retention [9].

Electronic flashcards have gained popularity in medical schools predominantly as a self-study tool for factual knowledge retention [10, 11]. In particular, Anki, an open-source flashcard platform available on a website, laptop, or cell phone, has the added capability to automate and individualize graduated practice intervals based on user-perceived difficulty [12]. In a study by Deng et al., Anki usage was correlated with improved performance on the US Medical Licensing Examination (USMLE) Step 1 in a dose-dependent manner [11].

While retrieval practice in the form of electronic flash-cards is increasingly used in the medical school setting, few studies have adopted it for graduate medical education [13–15]. Among them, Larsen et al. evaluated repeated testing on retention of knowledge in pediatric and emergency medicine residents and showed that repeated testing improved long-term retention of information after 6 months [13]. Morin et al. developed a website platform to teach radiology residents through image-based flashcards [14], and Matos et al. demonstrated that testing internal medicine interns at spaced intervals improved quiz scores [15]. Despite the promising data supporting retrieval practice, no study has evaluated feasibility and efficacy of electronic flashcard usage for in-training examination (ITE) preparation for post-graduate learners.

In this study, we created an Anki flashcard resource directed at preparation for the annual obstetrics and gynecology (ob-gyn) ITE conducted by the Council on Resident Education in Obstetrics and Gynecology (CREOG). Performance on the CREOG exam predicts subsequent successful performance on the ob-gyn board certification examination, and thus, the CREOG ITE is used by residency and fellowship program directors to objectively assess resident clinical knowledge [16]. The primary aim of our study was to evaluate resident usage and satisfaction with this spaced repetition study resource. Our secondary aim was to evaluate whether flashcard usage correlated with higher CREOG scores.

Methods

This project was granted exempt status by the Institutional Review Board of Duke University. The ob-gyn residents at Duke University were included in this study. An educational session was held on December 12, 2018, to obtain informed consent. Participants filled out an initial survey collecting demographic information. Personal Review of Learning in Obstetrics and Gynecology (PROLOG) questions, a free resource for self-study provided to ob-gyn residents by the American College of Obstetricians and Gynecologists (ACOG), were the basis for the study's flashcards. These PROLOG questions are only available in paperback book and portable document formats, and are a popular resource for studying for the ITE in ob-gyn. For this study, these questions were adapted to an electronic flashcard format using Anki software, and participants received access to a personal Anki account containing 883 flashcards based on PROLOG questions. During the educational session, participants were also instructed on how to access the flashcards on their laptop, website, and cell phone.

Each flashcard contained a question stem and multiple-choice answers. After revealing the correct answer, the user self-grades the difficulty of the question with choices

of "Again," "Hard," "Good," and "Easy." Next, the flash-card is automatically rescheduled for later review, with the interval dependent upon the difficulty evaluation and last review interval of the flashcard. For example, self-grading a flashcard as "Good" would increase the previous review interval by 250%, while self-grading as "Again" would reset the review interval of the flashcard. The "Hard" and "Easy" options would modulate the review interval to be shorter or longer, respectively. A flashcard user was defined as having greater than zero flashcard reviews of the study flashcards as determined by the internally generated Anki usage data.

At the 2019 CREOG exam on January 19th, a second survey was distributed evaluating study habits and experience with the flashcards. Usage and performance statistics generated internally by interacting with the flashcard system between the start of the study and the 2019 CREOG exam were collected. The 2019 and all previous CREOG scores standardized by all years were also obtained for each participating resident by the faculty author and blinded to all other remaining authors except for the statisticians. The CREOG exam is standardized nationally with a mean of 200 and a standard deviation of 20 points.

Continuous variables were summarized with median and interquartile range (IQR) and compared between users and non-users using Wilcoxon rank sum tests. Chi-square tests or Fisher's exact tests were used to compare between groups for categorical variables, contingent on cell counts. The difference between the 2018 and 2019 CREOG scores was calculated for each person, excluding first year residents. Linear regression models were then fit to examine the effect of (1) any flashcard usage on the 2019 CREOG score, (2) number of flashcards reviewed on the 2019 CREOG score, and (3) flashcard usage on the change in CREOG score between 2018 and 2019, both unadjusted and adjusting for post-graduate year (PGY). The distribution of residuals was checked for normality. As a sensitivity analysis, residents were relabeled as users if they self-reported any flashcard use on their study habit survey but had no flashcard review data. The flashcard usage models were then refit.

Results

Of the 32 residents, 31 (97%) participated in this study. Eighteen (58%) were flashcard users and thirteen (42%) were non-users (Table 1). These two groups represented similar demographics for age, sex, and race. There were no statistically significant differences in career plans or prior Anki experience. Flashcard users were predominantly second or third year residents, while flashcard non-users were predominantly first and fourth year residents ($p = 0.004$).

Among all participants, the study flashcards were the most popular study resource (68%). The second and third

Table 1 Participant demographics and study habits

	Flashcard non-user ($N=13$)	Flashcard user ($N=18$)	p value
Age (years), median (IQR)	30.0 (29.0–30.0)	29.0 (27.0–30.0)	0.18
Female	12 (92.3%)	16 (88.9%)	1.00
Race/Ethnicity			0.80
White	8 (61.5%)	12 (66.7%)	
Hispanic	0 (0.0%)	2 (11.1%)	
African American	2 (15.4%)	2 (11.1%)	
Asian	2 (15.4%)	1 (5.6%)	
Two or more races	1 (7.7%)	1 (5.6%)	
Post-graduate year			0.004
1	6 (46.2%)	2 (11.1%)	
2	0 (0.0%)	8 (44.4%)	
3	2 (15.4%)	6 (33.3%)	
4	5 (38.5%)	2 (11.1%)	
Prior Anki experience	5 (38.5%)	9 (50.0%)	0.52
Study resources			
Study flashcards	3 (23.1%)	18 (100.0%)	
Other Anki flashcards	0 (0.0%)	1 (5.6%)	
PROLOG books	4 (30.8%)	8 (44.4%)	
PROLOG games	0 (0.0%)	5 (27.8%)	
ACOG practice bulletins or committee opinions	5 (38.5%)	15 (83.3%)	
Formal didactic lecture materials	2 (15.4%)	10 (55.6%)	
Textbook(s)	2 (15.4%)	0 (0.0%)	
Other	2 (15.4%)	1 (5.6%)	
Average number of hours spent per day studying for the CREOG exam two weeks prior to flashcard distribution			0.75
<1 h	11 (84.6%)	16 (88.9%)	
1–2 h	2 (15.4%)	1 (5.6%)	
≥2 h	0 (0.0%)	1 (5.6%)	
Average number of hours spent studying for the CREOG exam two weeks before ITE			0.07
<1 h	10 (76.9%)	7 (38.9%)	
1–2 h	1 (7.7%)	8 (44.4%)	
≥2 h	2 (15.4%)	3 (16.7%)	

IQR interquartile range

most used study resources were ACOG practice bulletins and committee opinions (65%) and the original PROLOG books (39%), respectively. Two-thirds of flashcard users reported that the study flashcards were their primary study resource. Flashcard users also reported using more study resources than non-users (median [IQR] of 3 [3, 4] vs. 1 [0–2]). Most study participants reported studying <1 h a day in the 2 weeks prior to flashcard distribution. Two weeks before the CREOG exam, flashcard users reported studying more frequently (>1 h a day in 61% of users vs. 23% of non-users), but this was not statistically significant ($p=0.07$).

Flashcard users studied a median of 276 flashcards (IQR 108–1146) over a median of 3.7 h (IQR 1.2–12.4). Cell phones were the most popular modality for utilizing the flashcards (Table 2). All users found the flashcards helpful.

and all would recommend them to another ob-gyn resident. Nearly three-quarters (72%) of flashcard users reported planning to continue use of the flashcards throughout the rest of the year. Furthermore, 56% of flashcard users reported that the study flashcards helped their CREOG test anxiety. When asked what the strengths of the study flashcards were, approximately 80% emphasized convenience and accessibility (Table 3). Suggestions for improvements included having more detailed answer explanations and providing earlier access to the study flashcards.

The difference in CREOG scores between users and non-users was not statistically significant in either the unadjusted model (beta coefficient = 11.6; 95% CI = − 4.01, 27.2; $p=0.14$) or after adjusting for PGY (beta coefficient = 2.76; 95% CI = − 13.4, 18.9; $p=0.73$)

Table 2 Flashcard users' self-reported preferences

	Flashcard users ($N = 18$)
Any device(s) used	
Anki website	0 (0.0%)
Laptop app	6 (33.3%)
Cell phone app	15 (83.3%)
Found the flashcards helpful	18 (100.0%)
Plan to use flashcards throughout the academic year	13 (72.2%)
Would recommend the flashcards to another ob-gyn resident	18 (100.0%)
Found the flashcards helpful for CREOG test anxiety	10 (55.6%)

(Table 4). Without adjusting for PGY, every 100 reviews was significantly associated with a 1.24 point increase in CREOG score (95% CI = 0.21, 2.27; $p = 0.02$) (Fig. 1). After adjusting for PGY, there was no longer an association between number of flashcards reviewed and the 2019 CREOG score (beta coefficient per 100 flashcards = 0.69; 95% CI = − 0.21, 1.60; $p = 0.13$). When considering the change in CREOG scores from 2018 to 2019 for PGY2-4 residents, users had a 14.6 point higher average change score compared to non-users (95% CI = 6.11, 23.1; $p = 0.002$) in the unadjusted model. After adjusting for PGY, users had a 10.5 point higher average change score compared to non-users (95% CI = − 0.60, 21.7), but this increase in score did not achieve statistical significance ($p = 0.06$).

We identified four participants that had zero reviews in their flashcard account, but selected responses that would indicate they used the study flashcards. This included exporting the cards to their personal account, reporting a usage device, and selecting which sub-decks they utilized. Relabeling these participants as users resulted in 22 users and 9 non-users. Similar conclusions to the primary analysis were drawn (data not shown).

Discussion

This study evaluated resident usage and satisfaction with a new electronic flashcard self-study resource. Overall the flashcards were well received by participants, with all flashcard users finding them helpful and would recommend it to another ob-gyn resident. In particular, the flashcard users liked the convenience and accessibility of the study flashcards, which allowed them to incorporate studying between day-to-day activities. It was also the most popular study resource used by the residents for CREOG exam preparation. Using the flashcards to study for the 2019 CREOG exam appeared to correlate with improvement in scores from 2018 to 2019, but fell short of statistical significance after adjusting for PGY likely due to our limited sample size. Not surprisingly, PGY 2 and 3 residents were more likely to use the flashcards, as the ITE during these post-graduate years is often used as an objective measure of clinical knowledge by both residency and fellowship programs.

A strength of this flashcard resource is its availability on mobile devices, which allows for convenient access during short breaks in addition to more formal study sessions. The ease of accessibility is of particular importance given the busy schedules and clinical responsibilities of everyday resident life. All the platforms are free except for the Apple

Table 3 Narrative comments from post-exam survey

What were the strengths of the study flashcards?	Examples
Convenience and accessibility (×16)	"Quick access to questions on the phone between clinical duties, while waiting in line at grocery store, or being lazy on the couch at home"
Saves time (×3)	"Easier to access and therefore quicker to move through questions"
Short interval studying (×3)	"Easy to do only a few questions when I only have a few minutes"
Repetition (×2)	"Repetition is valuable"
What would you like to see changed or improved in future iterations of this project?	Examples
More detailed answer explanations (×3)	"Full length explanations on the cards as an optional follow-up to the abbreviated explanations"
Earlier access (×3)	"Would like for this to be available at the start of the year"
Expand content covered (×2)	"Include more complex, reflective questions"

Table 4 Association between flashcard use and CREOG scores standardized across all years

|	Mean (SD)	Unadjusted		Adjusted[b]	
		Beta coefficient (95% CI)	p value	Beta coefficient (95% CI)	p value
2019 CREOG score			0.14		0.73
Flashcard user	[a]	11.6 (−4.01, 27.2)		2.76 (−13.4, 18.9)	
Flashcard non-user	[a]	Reference		Reference	
Change in CREOG score from 2018 to 2019			0.002		0.06
Flashcard user	8.4 (9.1)	14.6 (6.11, 23.1)		10.5 (−0.60, 21.7)	
Flashcard non-user	−6.1 (8.8)	Reference		Reference	

[a]Mean scores withheld to protect institutional privacy

[b]Adjusted for post-graduate year

iOS cell phone application, which is $24.99. The Anki platform also offers several areas for customization, including changing the order of flashcard reviews, determining the gradation of review intervals, and marking cards for easy identification later [12]. The automatic and algorithmic nature of determining flashcard review intervals by user-perceived difficulty is also a unique feature of this flashcard platform that allows for optimization of the spacing effect to strengthen memory encoding without burdening the user. Another beneficial feature of the spaced repetition flashcard system was the multiple-choice question format. This not only mimics exam situations but also supports knowledge encoding through factual knowledge recall, analysis of clinical scenarios, and information synthesis.

These features differentiate this study from prior studies [13–15]. Specifically our study aid is designed for self-study through a more convenient mobile application and does not rely on additional prompting from programs and educators. Perhaps the most similar implementation was Morin et al.'s [14] creation of their own spaced education flashcard website in radiology. However, they were ultimately limited by the difficulty with website maintenance and funding. In contrast, we were able to harness a readily available, affordable platform, and apply it in a novel way for graduate medical education. Our chosen platform also allows for more nuanced gradations with determining flashcard intervals to personalize learning.

Fig. 1 Scatterplot of 2019 CREOG score standardized across all years vs. number of flashcards reviewed. The mean CREOG score of the study sample is represented by the 0 on the *y*-axis. Without adjusting for PGY, every 100 reviews was significantly associated with a 1.24 point increase in CREOG score (95% CI = 0.21,2.27; *p* = 0.02). However, there was no longer an association between number of flashcards reviewed and the 2019 CREOG score after adjusting for post-graduate year. Absolute CREOG scores were withheld to protect institutional privacy
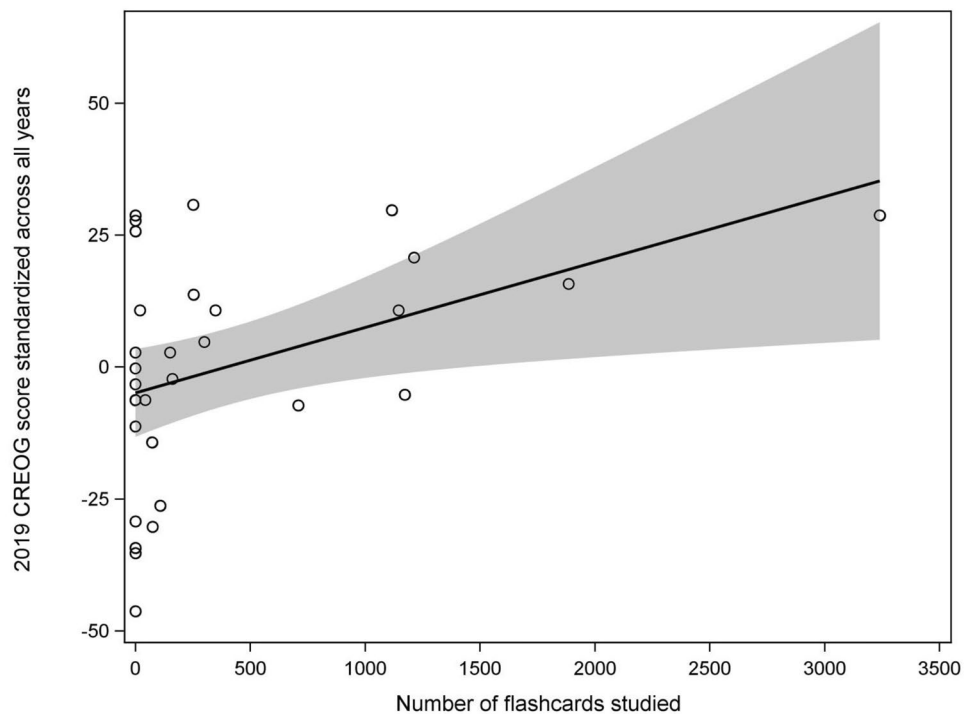
In addition to platform features for users, our study takes advantage of platform features conducive to quantitative research. In Deng et al.'s [11] study, Anki flashcard use was correlated with higher USMLE Step 1 scores in medical students. However, they retrospectively collected self-reported flashcard usage. In contrast, this study uniquely collected objective flashcard usage data generated when the user interacts with flashcard interface, which avoids the recall bias inherent in self-reported usage. Interestingly, we identified four participants with zero reviews in their flashcard account who reported using the study's flashcards for studying. We hypothesize that this may be due to failure of data synchronization after interaction with the flashcard application, exporting flashcards to their personal Anki accounts (for which data were not available), or use of the flashcards in a group setting on a different participant's account. Additional sensitivity analyses with these non-users moved to the user group yielded similar conclusions to the primary analysis.

One of the study limitations is the sample size and inclusion of only one residency program, which may not be representative of the general population of ob-gyn residents. It is also difficult to control for all the potential confounding variables as different residents are constantly exposed to different learning environments and teaching initiatives. There may also be participant bias as flashcard users were also more likely to study more for the CREOG examination. We also recognize that electronic flashcards may not be the most effective self-study tool for every single learner and that residents likely self-select study methods that they have found success with in the past. However, it is promising that only 9 out of 18 participants in the flashcard user group had prior Anki experience, suggesting that several participants used Anki for the first time and reported satisfaction with this study method.

Nonetheless, this preliminary study represents an innovative flashcard resource that participants found valuable for their self-study. Additional longitudinal investigation is recommended, as perhaps the greatest potential benefit of this self-directed learning tool is long-term retention of information beyond one examination with potential application to licensing examinations and clinical practice. It may also be beneficial to incorporate electronic flashcard resources into formal graduate medical education curriculum through enhancement of didactic sessions as well as facilitation of knowledge retention of ACOG practice bulletins.

Conclusion

In summary, implementation of an electronic flashcard resource for ITE preparation was not only successful within an ob-gyn residency program but also very well received by ob-gyn residents, the majority of whom were highly satis-

fied and planned for continued use throughout the academic year. Despite the association with improved CREOG performance from 2018 to 2019 in the unadjusted model, use of the flashcard resource did not meet statistical significance after adjusting for PGY.

Author Contribution Shelun Tsai: PGY-4 Resident. Co-conceived the idea behind the project; planned, designed, and executed the project; co-wrote the manuscript.

Michael Sun: PGY-4 Fellow. Co-conceived the idea behind the project; planned and designed the project; co-wrote the manuscript.

Melinda Asbury: Assistant professor of psychiatry. Assisted with project design; edited the manuscript.

Jeremy Weber: Biostatistician. Analyzed data; edited the manuscript.

Tracy Truong: Biostatistician. Analyzed data; edited the manuscript.

Elizabeth Deans: Assistant professor of obstetrics and gynecology. Assisted with project design; facilitated coordination of the project with resident education curriculum; edited the manuscript.

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

Declarations

Disclaimer The contents of this project are solely the responsibility of the authors and do not necessarily represent the official views of NCATS or NIH.

Ethics Approval This project was granted exempt status by the Institutional Review Board of Duke University.

Consent to Participate Informed consent was obtained from all participants.

Consent for Publication No identifying details were included in this manuscript.

Conflict of Interest The authors declare no competing interests.

References

1. 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.
2. Kornell N, Bjork RA. The promise and perils of self-regulated study. Psychon Bull Rev. 2007;14(2):219–24.
3. Kornell N, Son LK. Learners' choices and beliefs about self-testing. Memory. 2009;17(5):493–501.

4. Sleight DA, Mavis BE. Study skills and academic performance among second-year medical students in problem-based learning. Med Educ Online. 2006;11(1):1–6.

5. Allen EB, Walls RT, Reilly FD. Effects of interactive instructional techniques in a web-based peripheral nervous system component for human anatomy. Med Teach. 2008;30(1):40–7.

6. Bottiroli S, Dunlosky J, Guerini K, Cavallini E, Hertzog C. Does task affordance moderate age-related deficits in strategy production? Aging Neuropsychol Cogn. 2010;17(5):591–602.

7. Kang SHK. Spaced repetition promotes efficient and effective learning: policy implications for instruction. Policy Insights from Behav Brain Sci. 2016;3(1):12–9.

8. Carpenter SK, Cepeda NJ, Rohrer D, Kang SHK, Pashler H. Using spacing to enhance diverse forms of learning: review of recent research and implications for instruction. Educ Psychol Rev. 2012;24(3):369–78.

9. Ebbinghaus H. Memory: A contribution to experimental psychology. Mem A Contrib to Exp Psychol. 2013;20(4):155–6.

10. Ronner L, Linkowski L. Online forums and the "Step 1 climate": perspectives From a medical student reddit user. Acad Med. 2020;95(9):1329–31.

11. 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.

12. Anki. https://apps.ankiweb.net. Accessed March 1 2020.

13. 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.

14. Morin CE, Hostetter JM, Jeudy J, Kim WG, McCabe JA, Merrow AC, et al. Spaced radiology: encouraging durable memory using spaced testing in pediatric radiology. Pediatr Radiol. 2019;49(8):990–9.

15. Matos J, Petri CR, Mukamal KJ, Vanka A. Spaced education in medical residents: an electronic intervention to improve competency and retention of medical knowledge. PLoS ONE. 2017;12(7):1–8.

16. Lingenfelter BM, Jiang X, Schnatz PF, O'Sullivan DM, Minassian SS, Forstein DA. CREOG in-training examination results: contemporary use to predict ABOG written examination outcomes. J Grad Med Educ. 2016;8(3):353–7.