Spaced Repetition for UI Expertise Development

Hannah Kostuk
hannah.kostuk@usask.ca
University of Saskatchewan
Saskatoon, Saskatchewan, Canada

Carl Gutwin
gutwin@usask.ca
University of Saskatchewan
Saskatoon, Saskatchewan, Canada

CCS Concepts

• **Human-centered computing → Keyboards**; **Graphical user interfaces**; **Empirical studies in HCI**; • **Applied computing →** *Interactive learning environments*; *Psychology*.

Keywords

Human-Computer Interactions, Keyboard Shortcuts, Spaced Repetition, Spacing Effects

ACM Reference Format:
Hannah Kostuk and Carl Gutwin. 2026. Spaced Repetition for UI Expertise Development. In *Proceedings of Poster at Graphics Interface 2026 (GI '26).* ACM, New York, NY, USA, 3 pages.

1 Introduction

In Graphical User Interfaces (GUIs), novice users are supported by the visual nature of the interface, which allows them to quickly discover which tools are available to complete their task. To achieve a low upfront learning cost for users, GUIs sacrifice selection efficiency for command discoverability.

Expert-level commands allow a user to complete a task in a manner more efficient than a novice user. Expert-level commands often employ non-graphically based User Interfaces (UIs), such as keyboard shortcuts or gesture-based inputs. These commands require practice to use effectively and lack discoverability, but provides efficiency that novice-level selections do not by removing the physical and temporal costs of navigating through menus and icons with the on-screen cursor. This creates a barrier which prevents novice users from transitioning to expert-level command selection methods within an application, especially when coupled with users' preference to maintain the novice-level selection methods they have practiced since they first began using the application [4, 5].

Various methods (i.e., video guides, FAQ pages, startup tutorials, designated documentation pages) have been used to provide help to users learning an application. However, these methods do not necessarily teach expert-level techniques, and users also tend to avoid help systems until they cannot proceed without assistance [14]. Consequentially, these methods have not been effective for transitioning users from novice- to expert-level selection techniques.

Spaced repetition is a learning method which spaces out information recalls over time. In contrast, massed learning refers to performing information recalls all in one session. Spaced repetition has been used successfully for declarative knowledge (i.e., knowledge that requires conscious, intentional thought to recall) [6, 13, 17] and procedural knowledge (i.e., knowledge that allows a person to perform actions automatically, without conscious thought) [1, 10, 17].

We evaluated the use of spaced repetition for supporting the transition from novice- to expert-level command selections through an empirical study. This study focused on expertise development with keyboard shortcuts, due to their popularity across many computer applications. Results from this study suggest that spaced repetition, assisted by an adaptive scheduling algorithm, can aid users' transition from novice- to expert-level commands. Although differences between each condition in objective *test* trial performance measures were non-significant, each condition required different amounts of practice to achieve that performance, where those in the *Spaced Adaptive* condition required the fewest *learning* trials on average.

Our research makes two main contributions. First, we identify the need for teaching methods within software to support users' transition to expert-level command use, given the limitations of help methods currently employed. Second, we provide empirical evidence through a user study to support the use of spaced repetition as a viable option for learning expert-level command selections. Overall, spaced repetition is a method that requires little time from users to practice commands in comparison to massed practice, and is easy for users to understand and for application designers to implement.

2 Related Work

2.1 The Testing Effect

The testing effect is an established psychological phenomenon which posits that simply testing a person's memory will strengthen their recall for the items tested. The testing effect yields better learning than simply restudying - that is, study methods that use active recall strategies, such as self-testing, have consistently better outcomes for learning than methods of studying which employ passive or inactive strategies, such as re-reading content, or copying notes from an example [2, 15, 16, 18].

2.2 Current Applications of Spaced Repetition

Spaced repetition has been used successfully for different kinds of knowledge. Empirical evidence in support of spaced repetition has been found in studies examining rotary pursuit tasks [10], typewriter skills [1], and video game performance [8, 9]. Medical students who used the spaced repetition flashcard software Anki to study scored significantly higher across all four standardized exams analyzed in one study [6]. Similar spaced repetition flashcard software has also been used successfully as a study tool for statistics students [13] and criminal law students [17]. Given this prior research, it is reasonable to expect that spaced repetition may

Poster at Graphics Interface 2026 (GI '26), Waterloo, ON, Canada

be a viable option for acquiring and developing the motor skills and procedural memories necessary to learn keyboard shortcuts for a piece of software.

2.3 The Interstudy and Retention Gaps

In spaced repetition studies, the interstudy gap refers to the amount of time between consecutive study sessions, and the retention gap is the amount of time between the last repetition and the final test.

Some studies have observed the proportionality effect, where measured performance was positively correlated with the length of the interstudy gap when the retention gap was long relative to the interstudy gap [3, 7]. Evidence has also been found showing that interstudy gaps which grow after each repetition facilitated learning declarative knowledge better than interstudy gaps which maintained the same length [12], although contradictory evidence has been found in support of the opposite [15]. Some researchers posit that the relative spacing of repetitions - that is, whether the interstudy gaps grow, contract, or remain consistent between each repetition - did not yield any significant differences in long-term recall performance [11]. In fact, the absolute length of each interstudy gap may be what facilitates learning, rather than the relative spacing of repetitions [15]. Longer interstudy gaps seem to be beneficial regardless of the relative spacing, which may explain the contradictory messaging on optimal interstudy and retention gaps in the research body.

2.4 Scheduling Algorithms

Scheduling algorithms prescribe an optimal time to re-study a piece of information, where an optimal schedule maximizes knowledge retention while minimizing the temporal cost. One such algorithm, the Free Spaced Repetition Scheduler (FSRS), is a recently developed scheduling algorithm which defines the optimal interstudy gap as one that corresponds to the probability of recalling the information on a flashcard [19]. For example, to recall a card with 90% certainty, FSRS will schedule the card for review when it predicts the information on the card to have dropped to a 90% recall probability.

3 Methods

3.1 Procedure

To test the effectiveness of spaced repetition for learning hotkeys, we carried out a study comparing the use of non-adaptive and adaptive massed practice against adaptive spaced repetition. Adaptive conditions used the FSRS scheduling algorithm to queue the cards based on the participants performance.

3.1.1 Introductory trials. Each of the 13 participants first completed 10 *introductory* trials, where they were shown a flashcard with the name of the tool to select, the tool icon, and the keyboard shortcut used to select the tool. Once the correct keyboard shortcut was selected, the system automatically moved to the next trial.

3.1.2 Learning trials. After the *introductory trials*, participants began the *learning* trials, where they were shown the same flashcards again without the correct keyboard shortcuts displayed on them. Participants had one attempt to use the correct keyboard shortcut to select the tool shown on the card, after which they were shown the correct keyboard shortcut. They could then press 'Enter' key to progress to the next trial.

In the *Massed Adaptive* condition (N=5), participants completed trials until they selected at least 9 of the 10 tools correctly twice. Participants in the *Spaced Adaptive* condition (N=4) had a similar *learning* trial experience, except their trials were spread over several days. *Massed Non-Adaptive* condition participants (N=4) completed 100 learning trials. Each tool was presented once per 10 trials, in random order.

3.1.3 Test trials. Participants then completed *test* trials, where they saw each of the 10 tools once. The flashcard format was unchanged from the *learning* trials. To prevent deduction, they were not shown the correct keyboard shortcut after their selection. *Spaced Adaptive* participants performed *test* trials on the same day as their last *learning* trials. *Massed Adaptive* and *Massed Non-Adaptive* participants performed all components of the experiment in one day.

4 Results

All participants in the *Massed Non-Adaptive* condition completed exactly 100 *learning* trials. Participants in *Massed Adaptive* and *Spaced Adaptive* conditions completed 71.6 ([CI: 50.6 - 92.6]) and 47.8 ([CI: 36.4 - 59.1]) *learning* trials on average, respectively.

While significant differences in error rates were found during *learning* trials between the *Massed Non-Adaptive* (16.6% of selections were erroneous on average [CI: 12.9% - 20.3%]) and *Spaced Adaptive* conditions (63.4% of selections were erroneous on average [CI: 56.5% - 70.4%]), no significant differences in error rates were found between the conditions in the *test* trials. This means that participants in the *Spaced Adaptive* condition were able to achieve a similar level of performance as the *Massed Non-Adaptive* and *Massed Adaptive* conditions in the *test* trials, with the least amount of *learning* trials required to achieve that performance level.

Although the differences between each condition in *test* trial error rates were non-significant, the massed conditions may have had an advantage due to the proportionality effect - if the interstudy gaps are short relative to the retention gap, performance is positively correlated with the interstudy gap length [3, 7]. In this experiment, *test* trials were presented immediately after the completion of the last *learning* trials, making the interstudy gap long relative to the retention gap in the *Spaced Adaptive* condition. This may explain the lower average error rates in the two massed conditions, compared to the *Spaced Adaptive* condition. Further, participants in the *Massed Non-Adaptive* had the most *learning* trials on average, which may have given them more experience with certain selections than those in the adaptive conditions.

Based on these results, we recommend that designers implement short, adaptive sessions that are available daily within their application for users to practice their least-remembered keyboard shortcuts. This would allow users to achieve the same level of performance as they would with massed practice regardless of adaptivity, while minimizing the time invested in studying.

References

[1] A. D. Baddely and D. J. A. Longman. 1978. The Influence of Length and Frequency of Training Session on the Rate of Learning to Type. *Ergonomics* 21, 8 (1978), 627–635. arXiv:https://doi.org/10.1080/00140137808931764 doi:10.1080/00140137808931764

[2] Robert L Bangert-Drowns, James A Kulik, and Chen-Lin C Kulik. 1991. Effects of frequent classroom testing. *The journal of educational research* 85, 2 (1991), 89–99.

[3] NJ Cepeda, N Coburn, D Rohrer, JT Wixted, MC Mozer, and H Pashler. 2009. Optimizing distributed practice: theoretical analysis and practical implications. *Experimental Psychology* 56, 4 (2009), 236–246. doi:10.1027/1618-3169.56.4.236

[4] Andy Cockburn, Carl Gutwin, Joey Scarr, and Sylvain Malacria. 2014. Supporting novice to expert transitions in user interfaces. *ACM Computing Surveys (CSUR)* 47, 2 (2014), 1–36.

[5] Wai-Tat Fu and Wayne D Gray. 2004. Resolving the paradox of the active user: Stable suboptimal performance in interactive tasks. *Cognitive science* 28, 6 (2004), 901–935.

[6] Michael Gilbert, Timothy Frommeyer, Garrett Brittain, Nickolas Stewart, Todd Turner, Adrienne Stolfi, and Dean Parmelee. 2023. A Cohort Study Assessing the Impact of Anki as a Spaced Repetition Tool on Academic Performance in Medical School. *Medical Science Educator* 33 (07 2023). doi:10.1007/s40670-023-01826-8

[7] A. M. Glenberg and T. S. Lehmann. 1980. Spacing repetitions over 1 week. *Memory & Cognition* 8, 6 (1980), 528–538. doi:doi/10.3758/BF03213772

[8] Colby Johanson, Carl Gutwin, Jason T. Bowey, and Regan L. Mandryk. 2019. Press Pause when you Play: Comparing Spaced Practice Intervals for Skill Development in Games. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play* (Barcelona, Spain) *(CHI PLAY '19)*. Association for Computing Machinery, New York, NY, USA, 169–184. doi:10.1145/3311350.3347195

[9] Colby Johanson, Brandon Piller, Carl Gutwin, and Regan L Mandryk. 2023. If at first you don't succeed: Helping players make progress in games with breaks and checkpoints. *Proceedings of the ACM on Human-Computer Interaction* 7, CHI PLAY (2023), 342–368.

[10] Lyle E. Bourne Jr and E. James Archer. 1956. Time Continuously on Target as a Function of Distribution of Practice. *Journal of Experimental Psychology* 51, 1 (1956), 25. doi:10.1037/h0039887

[11] Jeffrey Karpicke and Althea Bauernschmidt. 2011. Spaced Retrieval: Absolute Spacing Enhances Learning Regardless of Relative Spacing. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 37 (05 2011), 1250–1257. doi:10.1037/a0023436

[12] Thomas K Landauer. 1978. Optimum rehearsal patterns and name learning. *Practical aspects of memory* (1978).

[13] Jaclyn K. Maass, Philip I. Pavlik, and Henry Hua. 2015. How Spacing and Variable Retrieval Practice Affect the Learning of Statistics Concepts. In *Artificial Intelligence in Education*, Cristina Conati, Neil Heffernan, Antonija Mitrovic, and M. Felisa Verdejo (Eds.). Springer International Publishing, Cham, 247–256.

[14] Neil Randall and Isabel Pedersen. 1998. Who exactly is trying to help us? The ethos of help systems in popular computer applications. In *Proceedings of the 16th annual international conference on Computer documentation*. 63–69.

[15] Henry L Roediger III and Jeffrey D Karpicke. 2006. The power of testing memory: Basic research and implications for educational practice. *Perspectives on psychological science* 1, 3 (2006), 181–210.

[16] Henry L Roediger III and Jeffrey D Karpicke. 2006. Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological science* 17, 3 (2006), 249–255.

[17] Mary Jane Samonte, Yeshua Miguel Abrenica, Juan Miguel Caparas, and Denise Nicole Marcelo. 2025. Anapolo: A Web-Based Spaced Repetition E-Learning Platform for Enhanced Long-Term Memory Retention. In *Proceeding of the 2024 8th International Conference on Education and E-Learning (ICEEL '24)*. Association for Computing Machinery, New York, NY, USA, 91–97. doi:10.1145/3719487.3719520

[18] Chunliang Yang, Liang Luo, Miguel A Vadillo, Rongjun Yu, and David R Shanks. 2021. Testing (quizzing) boosts classroom learning: A systematic and meta-analytic review. *Psychological bulletin* 147, 4 (2021), 399.

[19] Jarrett Ye. 2026. ABC of FSRS. https://github.com/open-spaced-repetition/fsrs4anki/wiki/ABC-of-FSRS Accessed: 2025-12-09.

Received 6 May 2026; accepted 18 May 2026