BRIEF

Optimizing the Spacing of Retrieval Practice to Improve Pharmacy Students’ Learning of Drug Names

James Terenyi, PharmD, a Heidi Anksorus, PharmD, a Adam M. Persky, PhD a,b

a UNC Eshelman School of Pharmacy, University of North Carolina at Chapel Hill, North Carolina
b Associate Editor, American Journal of Pharmaceutical Education, Arlington, Virginia

Submitted February 20, 2018; accepted July 13, 2018; published August 2019.

Objective. To optimize student pharmacists’ long-term retention of brand and generic drug names through spacing and modification of quiz format.

Methods. In three experiments, student pharmacists learned brand and generic names for the Top 200 medications taught in a self-paced course. The students completed two spaced quizzes on the course content, a final examination, and a measure of long-term retention administered several weeks after the course was completed. Experiment 1 examined the spacing of quiz 2 relative to quiz 1 and the final examination. Experiment 2 examined whether providing hints such as giving the first 3 letters of the paired associate impacted learning. Experiment 3 examined whether providing more elaborate context improved student retention.

Results. When quiz 2 was administered closer to quiz 1 but further from the final examination (expanding practice), students’ long-term knowledge retention was greatest. When the first three letters of the paired associate were provided, students’ long-term retention decreased. When more elaborate context was provided within the quizzes, students’ long-term retention was unaffected despite higher final examination performance.

Conclusion. Making retrieval of concepts easier reduces pharmacy students’ long-term knowledge retention, albeit only a small effect was seen in this study. In addition, providing more elaborate context in terms of drug class and indication during retrieval did not improve students’ learning.

Keywords: top 200, spacing, self-pace, brand, generic, testing effect

INTRODUCTION

Faculty members, preceptors, and trainees become frustrated when they spend time discussing a topic in class but later, the learner cannot remember the information. This raises the question of what instructors can do to improve long-term retention of learning. One learning method uses a system of spaced practice (practice over time) and regular retrieval (testing) of the information. 1 As such, it is important to consider methods to optimize spacing and retrieval within the classroom. Retrieval is a tool to promote long-term retention that focuses on a balance between retrieval difficulty and retrieval success. That is, the retrieval attempt must be difficult, but the learner must be successful in retrieving the desired information. In this study, we conducted three experiments using spaced quizzes to examine the balance between increasing retrieval success and modulating difficulty to promote long-term retention.

Memory formation involves several steps. The first step is encoding or the process of forming a mental representation of the information to be learned. Next, consolidation occurs to strengthen that mental representation so it is more likely to stay in longer-term storage in the brain. Storage of information is the next step. Storage ideally keeps learned information that is important to the individual. Finally, retrieval involves pulling that mental representation out of storage and back into working memory, and reprocessing the information to use in solving a current problem. This retrieval leads to reconsolidation of the mental representations that were previously formed, which strengthens memory storage. Thus, the act of retrieval is a very important part of the learning process because it causes reconsolidation and strengthens memory, and also improves accessibility. Accessibility means keeping the information so it can be retrieved and used when needed. Inaccessible memory may be
stored well but hard to retrieve, leading to a “tip-of-the-tongue” state. From an educational standpoint, helping students retrieve information on a regular basis can have important implications. Imagine a student is asked a question on rounds. They know the answer but cannot readily retrieve this information. Is that student helping the team? If a student cannot answer an examination question, is it because the information was never learned and stored or because it is just inaccessible at that moment in time? Thus, accessibility of learned information is important. In this study, we used regular quizzing to facilitate retrieval of brand and generic names of drugs. One way to optimize the impact of retrieval is through strategic spacing of retrieval attempts.

The benefits of spaced practice versus massed practice (“cramming”) has been well established. Research into the ideal spacing of learning has focused on the impact of various practice schedules on long-term retention.\(^2\)\(^-\)\(^4\) Previously, we used testing as the form of practice to examine the impact of spacing of this type of practice on long-term retention of brand and generic drug names (Prior Study, Table 1).\(^5\) Spacing of practice or retrieval means the learner is retrieving or using information over time. This is in contrast to massing of practice or cramming where all the retrieval or practice occurs at in a short period of time (eg, the night before an examination).\(^2\)\(^-\)\(^7\) Spacing works because it allows forgetting to happen, followed by making the effort to retrieve that somewhat forgotten information.\(^6\)\(^-\)\(^7\) When successful, that effort leads to learning and has been termed a desirable difficulty.\(^8\) If the retrieval attempt is too easy or too hard to the point the learner is unsuccessful in retrieving the information, then spacing is not as effective. This balance between success and difficulty forms the basis for equal and expanding spacing intervals (Figure 1).\(^9\)\(^-\)\(^10\) Previously we showed that having a longer time period between the quiz immediately after study (quiz 1) and the second quiz (quiz 2), and a shorter time period between the second quiz and the final examination (ie, a contracting schedule of practice) was the superior mode compared to equal spacing (ie, the time between quiz 1 and quiz 2 is equal to the time between quiz 2 and the final examination), expanding spacing (time between quiz 1 and quiz 2 is less than time between quiz 2 and the final examination) or massed practice (quizzing immediately after study). In the contracting model, there is a larger space of time between the first and second retrieval attempt compared to the second and third retrieval attempt, ie, time is contracting.\(^5\) However, the contracting schedule is in opposition to other research that suggests equal or expanding

### Table 1. Summary of Three Experiments and Relevant Prior Research

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Condition</th>
<th>Definition</th>
<th>Long-term Retention</th>
<th>Findings on Long-Term Retention</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prior Study(^5)</td>
<td>Equal</td>
<td>Equidistant between quiz 1 and final</td>
<td>6 weeks after final</td>
<td>Moderate effect favoring equal and expanding (d=.4) and a large effect favoring contracting to study only (d=.76). Small effect of massing (d=.14) compared to study only.</td>
</tr>
<tr>
<td></td>
<td>Expand</td>
<td>Closer to quiz 1 than final</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Contract</td>
<td>Closer to final than quiz 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mass</td>
<td>Quiz after study only</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Study only</td>
<td>No quizzes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Experiment 1</td>
<td>Equal</td>
<td>Equidistant between quiz 1 and final</td>
<td>6 weeks after final</td>
<td>Small effect favoring expanding practice. Expand larger effect (d=.16) than contract but not equal.</td>
</tr>
<tr>
<td></td>
<td>Expand</td>
<td>Closer to quiz 1 than final</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Contract</td>
<td>Closer to final than quiz 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Experiment 2</td>
<td>Assisted</td>
<td>Bra__ - Generic Brand - Gen__</td>
<td>4 weeks after final</td>
<td>Small effect favoring unassisted learning. Assisted condition lower effect (d=.22) than control.</td>
</tr>
<tr>
<td>Experiment 3</td>
<td>Elaboration</td>
<td>(Generic) is a Class used for Indication (Brand) is a Class used for Indication</td>
<td>6 weeks after final</td>
<td>No effect. Elaboration had no effect compared to control.</td>
</tr>
</tbody>
</table>

Quiz 1 occurred immediately after study. Quiz 2 occurred at least two weeks after Quiz 1 and at least 2 weeks before the Final examination. Final examination and long-term retention students were presented with the brand name or generic name and had to generate the paired associate. For Experiment 2 and 3, Quiz 2 was in the expanding schedule – scheduled closer to Quiz 1 than the final examination.
practice may be optimal, especially if feedback is provided after the assessment.\(^9\) Regardless of the spacing schedule, the selection of spacing interval may depend on the desired retention interval.\(^10\) As such, we replicated our prior study to re-test the optimal scheduling for retrieval.

One reason our prior study was conflicting with the existing literature was the absolute time intervals between the quizzes and the spacing between the quizzes and final examination. In some instances, the second quiz was either within one week of the first quiz or within one week of the final examination. Because forgetting is an important part of the spacing effect, in the current study we allowed at least two weeks between the spaced quizzes or the last quiz and the final examination. We also removed two conditions: the study only control that involved no quizzing and the massed practice that only had one quiz administered immediately after studying. These were removed to increase the sampling within the spaced conditions in order to minimize any drug-specific effects.

Prior research indicates that spacing works because it forces students to overcome a desirable difficulty: they start to forget and have to expend effort to retrieve information. The difficulty of trying to remember makes learning stronger; thus, making learning less difficult (through attempts to facilitate easier retrieval) may actually have negative effects.\(^7\) This is the concept we explored in the second and third experiments. In experiment 2, we attempted to make retrieval easier by providing hints in the form of the first three letters of the drug name and hypothesized that learning should decrease because of the easier retrieval attempt.

In experiment 3, we attempted to make the retrieval more difficult by adding more context to the cue.\(^11\) Experiment 3 also served a second purpose. As instructors we want to increase long-term retention and ensure learning is in the context of pharmacy practice. As such, in this final experiment, we provided elaborative or contextual cues to help reinforce the learning of brand and generic drug names. Anecdotally, students that work in a community or retail setting may learn brand and generics product names more efficiently than students trying to learn these with flash cards in isolation of clinical context. This would be consistent with the literature on context, distinctiveness, and elaboration.\(^12\) For example, Craik and Tulving found that when surrounding sentences were more elaborate, the more-detailed sentence frame produced better recall.\(^13\)

The purpose of this series of experiments was to answer three questions. The first question addressed the issue regarding the best spacing strategy to optimize long-term retention. The second question investigated whether retrieval difficulty was an important mediator in long-term retention effects of spaced practice in the classroom. The third question attempted to optimize the self-study of brand and generic drug names by adding elaborative context to the retrieval cues.

**METHODS**

The same students participated in experiments 1 and 2, with experiment 1 being conducted in the fall semester and experiment 2 in the spring semester. Participants were students enrolled in the medical terminology course sequence. A different cohort of students who were enrolled in the same fall semester medical terminology course experiment 1 participated in experiment 3.

For experiment 1, students were assigned 100 drug names to learn during the fall semester. The drugs were separated into 10 lists with each list based roughly on the body systems corresponding to the content of the concurrently taught pathophysiology course. An additional 100 drugs were covered in the second semester course, when experiment 2 was conducted. At the beginning of the semester, students were given a matched list of brand and generic drug names that they would be responsible for memorizing, with some generic products having up to three brand names. Students were instructed to study two drugs lists (20 drugs) for each weekly quiz and were required to look up the class and indications for each drug. Access to weekly online quizzes was opened each Friday morning and closed on (were due by) Sunday night. Students were given 15-25 minutes (depending on number of drugs on the quiz) and one attempt to complete the quiz on a learning management system (Sakai, Apereo Foundation) at their own convenience. Quizzes were fill-in-the-blank, and drug names had to be spelled correctly for students to receive credit. After the first five weeks of the course, all 100 drugs were studied and quizzed once. For the following five weeks, students completed weekly quizzes on unannounced drug lists according to the
Experiment 1 examined three study conditions with three drug lists randomly assigned to each condition (Table 1). In the “equal” practice condition, students were quizzed on the drug list immediately after they studied them for the first time. The second quiz took place equidistant between the original study time and the final examination. In the “expanding” practice condition, students were, again, quizzed on the drug list immediately after study. The second quiz occurred at a period closer to the original study period compared to the final examination; there was at least two weeks between the initial quiz and the second quiz. In the “contracting” practice condition, students were quizzed on the drug list immediately after study and again closer to the final examination than the original study period; again, there were at least two weeks separating the second quiz and the final examination. Students were balanced across the conditions to minimize list-dependent effects. For example, for list 1, half of the students were in the contracting condition, while the other half were in the expanding condition. In this design, students were aware of the content on the first quiz but not of that on future quizzes, which allowed us to focus on retrieval effects rather than re-study effects.

The primary outcome was performance on the final examination, which included 50 questions on brand and generic drugs, with drugs taken from one list for each study condition. The design used two examination formats in order to assess all 200 drugs. Half the students received lists 1, 3, 5, 7, and 9 and half received lists 2, 4, 6, 8, and 10. The final examination was similar to the online quizzes (fill-in-the-blank, with correct spelling required for credit) but completed in the classroom. There were 15 questions per study condition. The secondary outcome was student performance on a measure of their long-term knowledge retention. This content of this assessment was the same as that of the final examination but was completed six weeks later during the first day of class of the sequential course in the spring semester. Both examinations were completed in ExamSoft (ExamSoft Worldwide, Inc., Dallas, TX).

Experiment 2 was similar to experiment 1 except all students received the expanding spaced conditions only. Within the expanded spacing, students were exposed to two conditions: unassisted or assisted retrieval. In the unassisted condition, students received the expanding practice schedule as in experiment 1 (eg, ____ - Advil). In the assisted condition, as a hint, students were given the first three letters for half the drugs (eg, ibu____ - Advil). All drug lists were spaced, but we attempted to maximize the number of lists on the expanding schedule (for practical reasons we could not have all lists on this schedule). Conditions were balanced across participants. The final examination consisted of 50 of the 100 drugs, with each condition covering 15 drugs. The long-term retention assessment was conducted four weeks after the final examinations and occurred during the capstone experience.

In experiment 3, we again used only the expanding study condition. In this case experiment, unlike experiments 1 and 2, students were given a list of all 200 drug names at the beginning of the fall semester. Only the first six lists were included in the study because of the practical issues involved in spacing 200 drugs within a semester, ie, it is logistically challenging to have expanding practice for 200 drugs within a single semester without significantly increasing students’ study burden. Students were exposed to one of two conditions in this study: control or elaboration. In the control condition, students received the expanding practice schedule as described in experiment 1 and the unassisted condition in experiment 2. In the elaboration group, students received additional information in the form of the drug’s class and indication on the quizzes and examinations (eg, “Prozac, generic name _____, is a selective serotonin reuptake inhibitor used to treat depression.”). In experiment 3, we added another control condition. This additional control included drugs that did not have spaced quizzes (ie, just quizzed once immediately after study).

Paired t tests were used for experiment 1 and 3 because of the within-subjects design. A one-sided t test was performed for experiment 2 because of the between-group design and the hypothesis that intervention would decrease retention. Bonferroni-Holm correction was used for multiple comparisons. Cohen d effect size with 95% confidence interval was constructed between each condition. Significance was set at p < .05. This study was deemed exempt by the university’s institutional review board.

RESULTS

In experiment 1, we attempted to replicate prior findings and differentiate the superiority of a specific spacing condition. On delayed recall of drug knowledge, there were significant differences between conditions, with the expanding schedule resulting in the highest performance compared to the contracting schedule (d = .16) but not the equal schedule (Table 2). There were significant changes over time from the final examination to the six-week assessment, in order by effect size: contracting (d = 2.2), equal (d = 2.1), and expanding (d = 2.0). This suggests the expanding practice schedule may result in...
the highest retention, albeit the effect is small. As such, we used the expanding condition in experiments 2 and 3.

Experiment 2 focused on reducing the retrieval difficulty by giving the first three letters of the name as a hint. By reducing retrieval difficulty, we hypothesized this would reduce retention. Giving hints significantly reduced performance compared to the control on measures of long-term retention (d = 2.22) (p = .05 for one-sided) (Table 3). There were no differences on the final examination. Over time, performance decreased in the control (d = 2.8) and assisted condition (d = 3.0). As hypothesized, providing hints reduced retention, although the effect was small.

Experiment 3 focused on providing a more elaborate retrieval cue to help strengthen the retrieval process. The more elaborate cue did not result in better long-term retention. However, there were small differences in final examination performance. Under spaced conditions, the

Table 3. Summary of Experiment 2: Final Examination (Final) and Long-term Retention (4-weeks post) Performance when the First Three Letters were Provided as Hint

<table>
<thead>
<tr>
<th>Condition</th>
<th>Final Examination (%)</th>
<th>4-Weeks Posttest (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assisted (n=60)</td>
<td>94 (10)</td>
<td>40 (23)(^a)</td>
</tr>
<tr>
<td>Control (n=53)</td>
<td>95 (9)</td>
<td>45 (23)</td>
</tr>
</tbody>
</table>

Control: No hints provided during quizzing with quizzes in the expanding schedule
Assisted: The first three letters were provided during quizzes with quizzes in the expanding schedule

\(^a\) p < .05 vs control

more elaborate cues performed better than the control (d = .14, Table 4). In addition, the spaced conditions resulted in better performance on measures of long-term retention than the non-spaced conditions (standard: d = .43; elaborated: d = .54). Finally, without spacing, the students who received the elaborated cues performed worse on the final examination (d = .13) and on long-term retention (d = .26), albeit with small effects. The three experiments are summarized in Table 1.

DISCUSSION

The purpose of the study was to help elucidate an optimal retrieval regimen for learning. In this case, we applied the strategy to pharmacy students learning brand and generic drug names in a self-paced course. We found that the expanding schedule for retrieval had a small but significant benefit over other schedules, and that providing hints or more robust cues to students during retrieval did not significantly improve knowledge retention.

In this study, the benefits of elaborative processing were observed when unassisted retrieval was used compared with when assisted retrieval was used. These results point to the potential benefit of unassisted retrieval because of its greater difficulty (ie, greater elaborative processing) without a negative impact on short-term (final examination) scores. Therefore, while students may dislike practice testing because of the mental effort it requires, making the task easier for them may impair their learning. In addition, adding cues that help frame clinical relevance did not aid the testing. This lack of improvement may have been caused by the enhanced difficulty imposed by the extra cues or because the cues were not effective in providing better context. Further work is needed to find ways to improve on the spacing and testing effect. In addition, further studies should be performed to examine the impact of how elaboration may help reinforce the material used for elaboration. For instance, in the case of experiment 3, the following is an example of the kind of the cue that includes drug class and indication: “Prozac, generic name __________ , is a selective serotonin reuptake inhibitor used to treat depression.”

There was some variability in students’ long-term retention of information within the same course (experiment 1 and experiment 3). In addition, results from experiment 1 also differed from our prior research. Differences in outcomes from this study may be a result of changes within the semester. From our prior study to experiment 1, an additional course was added to that semester. This new course was a laboratory-based course with extensive pre-class learning. In experiment 3, the new course was still present but the amount of work outside of class was reduced. The outside influences could have
impacted the long-term retention. Other outside influences are possible, but admission criteria and number of students working were relatively consistent across cohorts. Also, there were changes in the self-paced courses (eg, changes in holidays, course duration, other assessments) that resulted in changes in some of the spacing from the previous study. Researchers who have considered spacing schedules have hypothesized that the optimal schedule is not fixed, but rather a function of the desired retention interval. It is unclear to what extent small changes in these schedules impact students’ performance on measures of long-term retention. Previous research used experiments with these schedules (ie, contracting, equal, expanding) but their spacing schedule was days between practices as opposed to weeks which was used in the current research. Scaling these schedules proportionately may not be an effective representation of the original experimental conditions. Additionally, providing feedback to students can eliminate some of the negative consequences of retrieval difficulty that results in retrieval failure. Finally, the benefits of spacing and the testing effect are important considerations in curriculum design and should not be overlooked as they provide an additional tool to maximize retention.

Our findings have several instructional implications. The first, having multiple opportunities to retrieve information improves students’ retention. Thus, instructors should deliberately plan testing and feedback sessions to review material. It may be more important to provide students with multiple retrieval attempts than to plan the way the retrieval attempts, ie, quizzes, are spaced out. However, the equal and expanding retrieval schedules may provide longer retention intervals.

The second implication is that making retrieval easier in an attempt to reduce student dislike for testing may actually hinder learning. Finally, a minimalist approach to testing may be as effective as a more elaborate approach.

CONCLUSION
Spacing of quizzes can improve long-term retention of learning in pharmacy students. Changing the ease of retrieval by adding cues may actually reduce the effectiveness of the instructional strategy, and adding more detailed cues does not improve knowledge retention. Repeated testing is an important part of learning, and instructors need to space out practice so that retrieval attempts are challenging but feasible.

REFERENCES

Table 4. Summary of Experiment 3: Final Examination (Final) and Long-term Retention (6-weeks post) Performance when more Elaborated Cues were Provided

<table>
<thead>
<tr>
<th>Spacing</th>
<th>Condition</th>
<th>Final Examination</th>
<th>6-Weeks Posttest</th>
<th>Effect Size (95% CI) for Forgetting</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spaced (n=129)</td>
<td>Standard</td>
<td>91 (13)</td>
<td>61 (26)*</td>
<td>1.5 [1.2-1.7]</td>
</tr>
<tr>
<td></td>
<td>Elaborated</td>
<td>92 (13)a,b</td>
<td>58 (28)*</td>
<td>1.6 [1.3-1.8]</td>
</tr>
<tr>
<td>Non-spaced (n=129)</td>
<td>Standard</td>
<td>89 (15)</td>
<td>50 (25)</td>
<td>1.9 [1.6-2.2]</td>
</tr>
<tr>
<td></td>
<td>Elaborated</td>
<td>87 (19)b</td>
<td>43 (28)c</td>
<td>1.8 [1.5-2.1]</td>
</tr>
</tbody>
</table>

Standard: No hints provided during quizzing or testing
Elaborated: More elaborative cues were provided during quizzes and testing
Spaced: Quizzes were in the expanding practice schedule
Non-spaced: Quizzes that were not spaced within the study
Data presented as percentage correct and standard deviation
a p<.001 vs Non-spaced
b p<.05 vs Standard within the condition
c p<.001 vs Standard within the condition