Effect of uniform versus expanding retrieval practice on the recall of physiology information

John L. Dobson
Department of Health and Kinesiology, Georgia Southern University, Statesboro, Georgia

Submitted 23 August 2011; accepted in final form 24 October 2011

A LEARNING STRATEGY that has received a fair amount of attention in the recent educational literature is expanding retrieval practice. Expanding retrieval practice strategy is itself based on two highly robust empirical findings. The first of those findings recognizes that tests and assessments actually strengthen learning by prompting active processing via the recall of information (1, 4, 10, 11, 16). The second empirical finding indicates that, to a point, the greater the length of time between initial learning and later recall of a piece of information, the stronger it will be encoded in memory and the more likely it will be retained for a long period of time (1–6, 8, 10–14, 16). However, if too much time elapses between the initial learning and later recall, the information will be unsuccessfully retrieved and will likely be lost from memory. Therefore, if recall events are delayed either too long or too little, the strategy will not optimize learning and long-term retention. One practice that has been purported to optimize the delay between learning and recall is to use expanding schedules of retrieval events. With this technique, a recall assessment is given immediately after the initial learning, which enhances the chances of a successful retrieval, and then subsequent recall assessments occur during systematically increased intervals, which deepen the level of cognition and lead to better long-term retention.

The support for expanding versus uniform schedules of learning dates back to at least 1927, in a study conducted by Tsai (18). Although the strategies used by Tsai were similar to those used in most current pertinent research, what the author actually compared were relearning events, each of which included aided recall, that followed either expanding and uniform schedules. In contrast, spaced retrieval strategies, such as expanding and uniform retrieval practices, are administered without feedback to examine the unmediated effects of repeated retrieval (9). Subsequently, the first direct support for expanding retrieval practice was provided by Spitzer (17) in 1939, when he demonstrated that the technique enhanced retention in a group of sixth graders. Decades later, the strategy was found to be an effective means of enhancing retention in those with both cognitive impairments (15) and Alzheimer’s disease (3). However, none of those latter three studies actually compared expanding and uniform retrieval practices, whereas the landmark study by Landauer and Bjork in 1978 (12) did. Landauer and Bjork (12) demonstrated that the delayed recall of fictitious names was significantly better after an expanding schedule of retrieval compared with a uniform schedule of retrieval. Although a similar conclusion was reached by Cull in 1996 (6), the subsequent seven studies that compared expanding and uniform retrieval practices either failed to find a advantage of expanding retrieval (2, 4, 8, 11, 13) or found that uniform retrieval practice actually resulted in superior long-term recall (5, 10).

Very recently, Storm et al. (16) published a study in which they concluded that the extent of the benefit of expanding retrieval practice, and consequently the explanation for the inconsistency in the studies mentioned above, may be determined by how vulnerable the to-be-learned information was to forgetting. More specifically, almost all of the studies that found no advantage with expanding retrieval practice (2, 4, 8–11, 13) conducted their recall assessments over a short period of time (up to 30–60 min) and likely failed to induce enough forgetting to be able to observe a benefit from the strategy. In contrast, Storm et al. (16) were able to induce sufficient forgetting to observe a benefit by including an effective distraction task between the recall assessments.

Another way to induce forgetting, which may also be more representative of the manner in which we often study and revisit information, is to space the retrieval assessments out so that they occur over different days (16). Yet, to the best of the author’s knowledge, only two studies have compared expanding and uniform schedules of retrieval using recall assessments that were spread out over multiple days. One of those two
studies was conducted by the author (7) and did indeed find a significant advantage with expanding retrieval practice; however, the other study, conducted by Cull (5), failed to find that advantage. Other than the types of information that were studied in the two investigations [physiological concepts vs. verbal items for the Dobson (7) and Cull (5) studies, respectively], there were two important distinctions that may have helped explain the discrepancy in their findings. The first distinction was with the average spacing between the retrieval assessments, which is a variable that at least two pertinent studies (1, 14) have identified as having a significant impact on the success of expanding retrieval practice. The average spacing in the previous study by the author (7) was 1 day, whereas it was 2 days in the study by Cull (5). It is possible that the extent of the benefit of expanding retrieval practice varies with the length of the average spacing between assessments and that Cull observed no advantage with the practice because the average spacing used in that study was too great. Since every study that has found an advantage with expanding retrieval practice (6, 7, 12, 14–16) had an average spacing of no more than 1 day, it would be interesting to see if, in fact, that same effect could be demonstrated using a mean spacing of several days or more.

The second major distinction between the studies by the author (7) and Cull (5) pertained to the length of the delay between the initial learning and the first retrieval assessment. Indeed, two studies by Karpicke and Roediger (10, 11) have proposed that the length of the delay preceding the first retrieval assessment is a more important determinant of the success of the strategy than are any differences in the pattern of delays between subsequent assessments (e.g., between expanding vs. uniform retrieval practices). So, it may be important that the previous study by the author (7) used delays of 0 (i.e., immediately) and 1 days for the expanding and uniform conditions, respectively, whereas Cull (5) used delays of 1 and 2 days for the expanding and uniform conditions, respectively. Perhaps the disagreement about the benefit of expanding retrieval practice between those two studies could be explained by the differences in the delays preceding their first assessments. It would therefore be interesting to compare the retention of information that was sufficiently vulnerable to forgetting after expanding and uniform retrieval practices that included both consistent and different delays between the initial learning and the first recall assessment.

In response to some of the unanswered questions summarized above, the first aim of this study was to extend the comparison of expanding and uniform retrieval practices further than it has been previously and such that the average spacing between the assessments was more than a couple of days. The second aim of the study was to compare expanding retrieval practices with different delays preceding the initial recall assessment as well as to compare expanding and uniform retrieval practices with the same delay preceding the initial assessment. The more specific purpose of this investigation was to compare the recall of physiology information throughout 30 days of the following four different retrieval practices: 1) expanding retrieval without a delay preceding the first assessment, 2) expanding retrieval with a 1-day delay preceding the first assessment, 3) uniform retrieval with a 1-day delay preceding the first assessment, and 4) uniform retrieval with a uniform (i.e., 1 wk) delay preceding each assessment, including the first assessment.

**METHODS**

**Students and course.** All experimental procedures were approved by the university’s Institutional Review Board. The students that participated in this investigation were from a spring 2011 APK 2105 Applied Human Physiology course at the University of Florida. The typical Applied Human Physiology student was either a second- or third-year student and was enrolled in a prehealthcare major (e.g., Pre-Medicine, Pre-Nursing, Pre-Pharmacy, Pre-Physical Therapy, Exercise Physiology, etc.).

**Experimental procedures.** Participants first completed a descriptive questionnaire in which they reported both their biology and chemistry/biochemistry course experience and their specific level of experience with the two experimental topics: immunology and reproductive physiology. Participants were matched according to 1) whether or not they had completed any introductory biology courses; 2) whether or not they had completed any introductory chemistry/biochemistry courses; 3) their self-reported knowledge of the reproductive system (i.e., extensive, some, or no knowledge); and 4) their self-reported knowledge of the immune system (i.e., extensive, some, or no knowledge). Participants from each cumulative matched group were then randomly assigned to one of the following five experimental groups: 1) uniform (days 1, 10, and 20); 2) uniform (days 8, 15, and 22); 3) expanding (days 1, 6, and 16); 4) expanding (days 2, 7, and 17); and 5) control.

Every participant completed all five sequential phases of the experiment over a 30-day period. The first phase required participants to read and then reread sets of sentences that summarized an overview of both the immune and reproductive systems. During the latter four phases, participants were repeatedly assessed, without feedback, to prompt them to repeatedly recall the immune and reproductive physiology information they had read. Participants performed all five of the experimental phases using the University of Florida’s e-learning course management system in Sakai. Briefly, with all five phases, participants first logged on to the e-learning course website and then selected a link that accessed the appropriate assignment. The individual elements that comprised each of the assignments were then automatically administered by the website, including the instructions for each phase, the to-be-learned physiological concepts in the first phase, and all of the assessments in the latter four phases. The e-learning system was also programmed to help ensure that the participants in each of the five experimental groups completed their assignments on the exact specified days (described below).

One the first day of the experiment, all participants began by reading a total of 17 short paragraphs that summarized numerous key immune and reproductive physiology concepts. Examples of these paragraphs are shown in Table 1. Participants were instructed to carefully concentrate on, and try to retain, the things they read. Then, to better insure that the participants had sufficiently internalized the experimental concepts, every participant immediately read through a PowerPoint slideshow that presented many of the same immune and reproductive physiology concepts. More specifically, 30 key sentences were taken from the 17 experimental paragraphs (Table 2), and each sentence was presented on a single PowerPoint slide. Thus, the goal of the first phase of the experiment was for participants to study the experimental concepts in roughly the same manner in which they normally internalized information in APK 2105. That is, the students were generally expected to prepare for lectures by reading ahead in the textbook, and they were then exposed to much of that same information again during a subsequent class meeting and PowerPoint lecture.

During the latter four phases of the experiment, participants were repeatedly prompted to recall some of the physiology concepts they had read in the first phase. All 4 recall phases consisted of a
The immune system includes leukocytes, lymphoid tissues, and the molecules they produce. The functions of the immune system include clearing foreign material from the body and bringing about long-term immunity to infectious agents. There are five major types of leukocytes: basophils, eosinophils, macrophages, monocytes, and neutrophils.

The Female Reproductive System: Part I

The female reproductive system includes the external genitalia, ovaries, and reproductive tract (uterus, uterine tubes, and vagina). The menstrual cycle lasts about 28 days, includes cyclical changes in ovarian and pituitary hormone secretions, and includes menstruation. Ova develop from germ cells whose number is fixed at birth, and they fully mature only after being fertilized.

20-question assessment that was automatically administered by the e-learning course management system. The assessment consisted of 6 complex multiple-choice questions and 14 fill-in-the-blank questions (Table 3). As previously stated, participants received no feedback as to how many assessment questions they answered correctly or incorrectly. Importantly, participants were also assured that they would receive complete APK 2105 course credit for each assessment if they simply carefully read and then answered every assessment question to the best of their ability. Since their course credit did not depend on the number of questions they answered correctly or incorrectly, participants were therefore instructed not to consult any additional resources (e.g., textbooks, fellow students, or the internet) when answering the questions.

The precise administration date of each of the first three recall assessments depended on the experimental group, but every participant completed the fourth assessment (i.e., the posttest) on the 30th day of the experiment. Those assigned to follow uniform schedules of retrieval had a uniform amount of time (e.g., either 7 or 10 days) between their four assessments, whereas those assigned to follow expanding schedules of retrieval had roughly 5, 10, and 14 days between their four assessments, respectively. Furthermore, approximately half of the participants that followed both the uniform and expanding schedules of retrieval completed their first recall assessment immediately after rereading the physiology concepts, whereas the other half of both groups completed the first assessment after some delay. More specifically, those in the uniform (days 1, 10, and 20) group completed their first three assessments on the 1st, 10th, and 20th days of the experiment (i.e., immediately after rereading the physiology concepts and then again exactly 9 and 19 days later). Those in the uniform (days 8, 15, and 22) group completed their first three assessments on the 8th, 15th, and 22nd days of the experiment (i.e., 7, 14, and 21 days after rereading the physiology concepts). Those in the expanding (days 1, 6, and 16) group completed their first three assessments on the 1st, 6th, and 16th days of the experiment (i.e., immediately after rereading the physiology concepts and then again exactly 5 and 15 days later). Those in the expanding (days 2, 7, and 17) group completed their first three assessments on the 2nd, 7th, and 17th days of the experiment (i.e., 1, 6, and 16 days after rereading the physiology concepts). Finally, those in the control group did not complete any of the first three recall assessments and, instead, completed three distraction assessments on a uniform schedule. The distraction assessments consisted of questions about digestive physiology.

Since each participant carried out all phases of the experiment over the internet and while unsupervised, one last measure was included to safeguard against using data from any participant that failed to correctly follow the required procedures. On the posttest, participants were asked to indicate if they either 1) failed to carefully read all of the instructions and all of the immune and reproductive physiology concepts in first phase of the experiment or 2) consulted any additional resources (e.g., textbooks, fellow students, or the internet) when answering any of the assessment questions. To encourage honest responses, participants were assured that there would be no penalty if they answered “yes” to either of the above questions. However, those participants that did answer “yes” to either of those questions were excluded from the study.

Data analysis. Statistical differences in the assessment scores between experimental groups and over the course of the four assessments were made using four sets of ANOVAs with Bonferroni post hoc tests. The first set of ANOVAs compared the five experimental groups (uniform (days 1, 10, and 20), uniform (days 8, 15, and 22), expanding (days 1, 6, and 16), expanding (days 2, 7, and 17), and control) on each of the four assessments (assessments 1–3 and the posttest). Note that the control group was only included in the analysis of the posttest because that group did not complete the first three assessments. The second set of ANOVAs compared the two retrieval conditions (the two expanding retrieval groups vs. the two combined uniform retrieval groups) on each of the four assessments. The third and fourth sets of analyses were both repeated-measures ANOVAs. One that compared the four experimental groups and two retrieval conditions, respectively, across the four assessments. Although there was a great deal of variability between the scores within each experimental group, ANOVAs were used because Levene’s tests had confirmed that the error variances were equal across the groups. Statistical significance was set at $P < 0.05$, but some trends up to $P \leq 0.10$ are discussed. Data are expressed as means $\pm SE$.

RESULTS

A total of 288 students was enrolled in the spring 2011 APK 2105 Applied Human Physiology course. Of those students, 250 (87%) correctly completed all six phases of the investiga-

---

**Table 1. Two of the seventeen paragraphs used in the first phase of the experiment**

The immune system includes leukocytes, lymphoid tissues, and the molecules they produce. There are five major types of leukocytes: basophils, eosinophils, macrophages, monocytes, and neutrophils. The menstrual cycle lasts about 28 days, includes cyclical changes in ovarian and pituitary hormone secretions, and includes menstruation. Ova develop from germ cells whose number is fixed at birth, and they fully mature only after being fertilized.

---

**Table 2. Four of the thirty sentences used in the second phase of the experiment**

The immune system includes leukocytes, lymphoid tissues, and the molecules they produce. There are five major types of leukocytes: basophils, eosinophils, macrophages, monocytes, and neutrophils. The menstrual cycle lasts about 28 days, includes cyclical changes in ovarian and pituitary hormone secretions, and includes menstruation. Ova develop from germ cells whose number is fixed at birth, and they fully mature only after being fertilized.

---

**Table 3. Selected questions from the recall assessment**

1. Leukocytes develop to maturity in the central lymphoid tissues, which include bone marrow and the __________.
2. B cells and T cells provide for all of the following features of immune responses except:
   - A. Specificity
   - B. Diversity
   - C. Memory
   - D. Tolerance
   - E. All of the above are true.
3. Which of the following is/are true concerning the menstrual cycle?
   - A. It typically lasts about 28 days.
   - B. It is marked by cyclic changes in ovarian hormones.
   - C. It is marked by cyclic changes in pituitary hormones.
   - D. All of the above are true.
   - E. Answers A and B are true.
4. Most cells of the body contain ____ set(s) of chromosomes, but gametes (sperm cells and ova) contain ____ set(s) of chromosomes.
significantly higher (F and 2.56, respectively (Fig. 1). Those in the uniform (days 1, 6, and 16) group were 48.99, 56.6 (days 1, 6, and 15, days 2, immediately after learning the physiology concepts, days 1, 6, and 15) and expanding (days 2, 7, and 17) groups. There was also a nearly significant statistical difference (P = 0.06) between those in the expanding (days 1, 6, and 16) and expanding (days 2, 7, and 17) groups on this assessment.

Those in the expanding (days 1, 6, and 16) and expanding (days 2, 7, and 17) groups completed assessment 3 15 and 16 days, respectively, after learning the physiology concepts, whereas those in the uniform (days 1, 10, and 20) and uniform (days 8, 15, and 22) groups completed it 19 and 21 days, respectively, after learning the concepts. Mean assessment 3 scores for the expanding (days 1, 6, and 16), expanding (days 2, 7, and 17), uniform (days 1, 10, and 20), and uniform (days 8, 15, and 22) groups were 46.90 ± 2.81, 38.22 ± 2.60, 35.43 ± 2.02, and 33.19 ± 1.90, respectively. Those in the expanding (days 1, 6, and 16) group scored significantly higher (F = 6.28, P = 0.000, partial η² = 0.09) on assessment 3 than did those in both the uniform (days 1, 10, and 20) and uniform (days 8, 15, and 22) groups. There was also a nearly significant statistical difference (P = 0.06) between those in the expanding (days 1, 6, and 16) and expanding (days 2, 7, and 17) groups on this assessment.

The combined means for those two groups that followed uniform schedules of retrieval were 40.54 ± 1.80, 33.95 ± 1.44, and 34.26 ± 1.38 for assessments 1, 2, and 3, respectively (Fig. 2). The combined means for those two groups that followed expanding schedules of retrieval were 45.48 ± 2.10, 41.57 ± 1.78, and 42.30 ± 1.95 for assessments 1, 2, and 3, respectively. There were no significant differences in scores between the expanding and uniform retrieval conditions on assessment 1 (F = 3.21, P = 0.08), but those in the expanding

Fig. 1. Relationship between the experimental groups and assessment scores throughout the 30 days of the experiment (n = 250). Scores are means ± SE. There were significant differences between the uniform (days 1, 10, and 20) group and both the uniform (days 8, 15, and 22) and expanding (days 2, 7, and 17) groups, and there were significant differences between the expanding (days 1, 6, and 16) group and the uniform (days 8, 15, and 22) and expanding (days 2, 7, and 17) groups on assessment 1 (ANOVA; F = 19.14, P = 0.000). There were significant differences between the expanding (days 1, 6, and 16) group and both the uniform (days 1, 10, and 20) and uniform (days 8, 15, and 22) groups on the latter three assessments (ANOVA, assessment 2: F = 6.15, P ≤ 0.001; assessment 3: F = 6.28, P = 0.000; posttest: F = 6.52, P = 0.000). There were significant differences between the control group and all four of the retrieval groups on the posttest (ANOVA, F = 21.57, P ≤ 0.000).

Fig. 2. Relationship between the two retrieval conditions and assessment scores (n = 196). Scores are means ± SE. *There were significant differences between the uniform and expanding retrieval conditions on the latter three assessments (ANOVA, assessment 2: F = 11.03, P = 0.001; assessment 3: F = 11.32, P = 0.001; posttest: F = 14.09, P = 0.01). The difference between the uniform and expanding retrieval conditions on assessment 1 just missed statistical significance (ANOVA, F = 3.21, P = 0.07).
condition did score significantly higher on assessment 2 ($F = 11.03, P = 0.001$, partial $\eta^2 = 0.05$) and assessment 3 ($F = 11.32, P = 0.001$, partial $\eta^2 = 0.06$).

Posttest. All five experimental groups completed the posttest 29 days after they had learned the physiological concepts. Mean posttest scores for the uniform (days 1, 10, and 20), uniform (days 8, 15, and 22), expanding (days 1, 6, and 16), expanding (days 2, 7, and 17), and control groups were $36.15 \pm 1.97, 32.31 \pm 1.87, 45.80 \pm 2.56, 39.71 \pm 2.48,$ and $21.26 \pm 1.40$, respectively. There were no significant differences in posttest scores between the two groups that followed uniform schedules of retrieval ($P = 1.00$), nor were there significant differences in scores between the two groups that followed expanding schedules of retrieval ($P = 0.34$). Those in the expanding (days 1, 6, and 16) group did significantly outscore those in both uniform retrieval groups ($F = 6.52, P = 0.000$) on the posttest. Also, the posttest scores of those in the expanding (days 2, 7, and 17) group were nearly statistically greater ($P = 0.09$) than those in the uniform (days 8, 15, and 22) group, but they were statistically equivalent ($P = 1.00$) to those in the uniform (days 1, 10, and 20) group. The partial $\eta^2$ indicated that $10\%$ of the between-subjects variance was accounted for by the type of retrieval condition ($\eta^2_p = 0.1$). Finally, all four retrieval groups had significantly higher scores on the posttest than did those in the control group ($P = 0.000$).

When the two uniform groups were combined and the two expanding groups were combined, the resulting means on the posttest were $34.10 \pm 1.36$ for the uniform condition and $42.57 \pm 1.80$ for the expanding condition. Those scores were significantly different ($F = 14.09, P = 0.000$), and the partial $\eta^2$ was 0.07.

Repeated-measures effects. In terms of within-subjects effects, there was a significant decline in scores from assessment 1 to the posttest ($F = 22.02, P = 0.000$), and the associated partial $\eta^2$ was 0.10. There was no significant repeated-measures interaction between the two uniform retrieval groups ($P = 0.24$), nor was there a significant repeated-measures interaction between the expanding (days 2, 7, and 17) group and either the uniform (days 1, 10, and 20) group ($P = 1.00$) or the uniform (days 8, 15, and 22) group ($P = 0.45$). However, there was a significant within-subjects interaction between the expanding (days 1, 6, and 16) group and the expanding (days 2, 7, and 17) group ($P = 0.003$), uniform (days 1, 10, and 20) group ($P = 0.01$), and uniform (days 8, 15, and 22) group ($P = 0.000$). The percentage of variance accounted for by that interaction was $16\%$ ($\eta^2_p = 0.16$). Finally, there was no significant repeated-measures interaction between the combined uniform and combined expanding retrieval conditions ($F = 1.75, P = 0.17$).

**DISCUSSION**

The purpose of this study was to compare the retention of physiology information throughout 30 days of the following four different retrieval practices: 1) expanding retrieval without a delay preceding the first assessment, 2) expanding retrieval with a 1-day delay preceding the first assessment, 3) uniform retrieval with a 1-day delay preceding the first assessment, and 4) uniform retrieval with a uniform (i.e., 1 wk) delay preceding each assessment. The major novel conclusion was that expanding retrieval practice resulted in a significantly greater recall of information than did uniform retrieval practice. There was also a strong tendency for longer delays between the initial learning and the first recall assessment to lead to poorer retention, and so the most successful retention strategy was expanding retrieval practice with no delay preceding the first assessment.

Beginning with the first recall assessment, it was not surprising that performance on that assessment was at least somewhat inversely related to the length of the delay that preceded it. More specifically, the uniform (days 1, 10, and 20) and expanding (days 1, 6, and 16) groups, which both completed their first recall assessment immediately after learning the physiology concepts, performed significantly better than those in the expanding (days 2, 7, and 17) and uniform (days 8, 15, and 22) groups on assessment 1. Surprisingly, the expanding (days 2, 7, and 17) and uniform (days 8, 15, and 22) groups had statistically equivalent scores on assessment 1, despite having completed it after a 1- versus 7-day, respectively, delay (i.e., 6 days apart from one another). Thus, there was a clear advantage to completing the first recall assessment <1 day (i.e., immediately) after the initial learning, but there was no further disadvantage when that delay was increased from at least 1 day up to 7 days. It is also interesting to note that the relative amount of forgetting experienced by the aforementioned two sets of groups reversed before assessment 2. That is, despite having comparable sets of delays between assessments 1 and 2, those in both the expanding (days 2, 7, and 17) and uniform (days 8, 15, and 22) groups experienced no meaningful forgetting during that delay, whereas those in the expanding (days 1, 6, and 16) and uniform (days 1, 10, and 20) groups experienced an ~18% and 29% reduction in scores, respectively, from assessment 1 to assessment 2. Nevertheless, the results indicate that the expanding (days 1, 6, and 16) group still successfully recalled significantly more information than did those in the two uniform groups and very nearly significantly ($P = 0.06$) more than those in the expanding (days 1, 6, and 16) and uniform (days 1, 10, and 20) groups experienced an ~18% and 29% reduction in scores, respectively, from assessment 1 to assessment 2. Those same relationships and each group’s mean assessment scores remained particularly consistent throughout both assessment 3 and the posttest. The only exception to that consistent pattern was that the expanding (days 1, 6, and 16) and expanding (days 2, 7, and 17) groups reached even greater statistical similarity by the time they completed the posttest.

When the relationships described above are considered as a whole, two important trends emerge that both also reinforce conclusions very recently reached by Maddox et al. (14). The first was that all meaningful forgetting of the experimental material occurred by the time each group reached their second recall assessment. Therefore, performance on a second retrieval assessment can be a strong predictor of both performance on later assessments and longer-term retention (14). The second trend was an inverse relationship between the length of time preceding the first two recall assessments and the extent of long-term retention. Maddox et al. (14) was the first to report this inverse relationship, and they found it to be particularly strong with older adults. The results of the present investigation strengthen their finding and add that the inverse relationship remains when retrieval events and retention are examined over periods of time that are far longer than a few hours (i.e., days and weeks).

With specific regard for the posttest scores and retention 30 days after learning had occurred, it is first important to observe that all four retrieval groups performed twice as well as did the
control group. It is also important to note that there were no statistical differences between the two uniform retrieval groups on the posttest, nor were there any between the two expanding retrieval groups. However, those in the expanding retrieval condition did outscore those in the uniform retrieval condition on the latter three assessments, including the posttest. That difference in performance occurred despite the fact that the retention interval between assessment 3 and the posttest was twice as great in the two expanding retrieval schedules compared with the two uniform retrieval schedules (i.e., the uniform group retention intervals were 8 and 10 days and the expanding group intervals were 18 and 19 days). Therefore, the results of this investigation agree with previous studies that found an advantage with expanding retrieval practice (6, 7, 12, 14–16). According to Storm et al. (16), the common element in all of those studies is that they all induced sufficient forgetting to facilitate a benefit from expanding retrieval practice, whereas the many studies that failed to find that advantage (2, 4, 8, 10, 11, 13) did not induce sufficient forgetting. Indeed, almost all pertinent studies have used recall assessments that were conducted over a total of 30–60 min, and that is a period of time that may, by itself, be insufficient to facilitate and observe actual differences between expanding and uniform retrieval practices. Although most of the studies that did find an advantage with expanding retrieval practice also used protocols that were equally as short (6, 12, 14–16), those studies were able to induce sufficient forgetting by including an effective distraction task between the recall assessments. In this investigation, forgetting was instead induced by spreading the retrieval assessments out over relatively long periods of time (>24 h), and, once again, expanding retrieval practice resulted in superior long-term retention.

It is worth drawing special attention to recent pertinent study by Karpicke and Bauernschmidt (9). Like the similar studies mentioned immediately above, Karpicke and Bauernschmidt (9) examined retention after expanding and uniform schedules of retrieval; however, only those pieces of information that were initially recallable were included in their retrieval comparison. That is, if a subject was able to initially recall an item, then he or she was thereafter required to retrieve it after either an expanding or uniform schedule, but if a subject failed to initially recall an item, then it was dropped from further practice. Within that specific context, Karpicke and Bauernschmidt (9) found no differences between expanding or uniform schedules of retrieval. Those results are compelling and should be carefully considered by both educators and future investigators of this topic; however, it is also important to acknowledge that the learning/retrieval strategies used in that study clearly differ from those used in most spaced retrieval research, including the present investigation. Also, and perhaps more importantly, the retrieval assessments used by Karpicke and Bauernschmidt were spaced no more than minutes apart and, once again, may therefore not have induced sufficient interassessment forgetting to facilitate and observe an advantage with expanding retrieval practice.

Before the present study, only two other pertinent investigations have used retrieval assessments that were spread out over the course of days; one was a previous similar study by the author (7) that did find an advantage with expanding retrieval practice, whereas the other study by Cull (5) did not. One of the motivations for conducting the present investigation was to better understand why those latter two studies produced conflicting results. As was mentioned above, two of the major methodological differences between those two studies were the average spacing between retrieval assessments and the delay preceding their first retrieval assessments. In terms of the average spacing between assessments, at least two previous investigations (1, 14) have proposed that this variable may have a significant impact on the success of expanding retrieval practice. While that may indeed be the case, it is very unlikely to explain the aforementioned discrepancy between studies because a relatively long-term advantage of expanding retrieval practice has now been demonstrated using assessment spacing averages that were both less than used by Cull and more than that used by Cull. More specifically, the average spacing used in the previous study by the author (7), the study by Cull (5), and the present study was 1 day, 2 days, and over 7 days, respectively.

The second methodological difference that could help explain the discrepancy between Dobson (7) and Cull (5) was with the delay preceding the first recall assessment. Again, those delays were 0 (i.e., immediately) and 1 days for the expanding and uniform retrieval conditions, respectively, in the previous study by the author (7) and 1 and 2 days for the expanding and uniform conditions, respectively, in the study by Cull (5). Could that extra 24-h delay used by Cull explain why that investigation found no advantage with expanding retrieval practice? Given the statistical similarity between the expanding (days 1, 6, and 16) and expanding (days 2, 7, and 17) groups, the similarity between the expanding (days 2, 7, and 17) and uniform (days 1, 19, and 20) groups and yet the differences between the expanding (days 1, 6, and 16) and uniform (days 1, 19, 20) groups throughout the latter assessments in the present study, the answer would seem to be yes. That is, the results of this study indicate that the advantage of expanding retrieval practice was counterbalanced or reduced when the delay preceding the first recall assessment was ≥24 h. Therefore, to optimize the effectiveness of expanding retrieval practice, the first recall assessment should be administered immediately after learning has occurred. The results of this study, along with those previously reported both by the author (7) and by Maddox et al. (14), therefore disagree with the conclusions of Karpicke and Roediger (10, 11) that an increase in the delay preceding the first recall assessment enhances retention.

The most meaningful limitation to this study was the manner in which the participants completed the experimental readings and assessments, which was all unsupervised via computers and the internet. However, the experimental procedures were simple and very straightforward, and those procedures were clearly and concisely explained to the participants. Furthermore, as was described above, the only data used in the analysis were those taken from participants that indicated they were confident they had performed all phases of the experiment to the best of their ability and exactly as they were required to do. Nevertheless, it would certainly strengthen the results of this investigation if comparable results are found with a very similar study in which participants are directly observed as they complete each phase of the experiment. Future studies that compare expanding and uniform retrieval practices should also focus on 1) using variable numbers of retrieval assessments; 2) comparing different groups (e.g., young, adult, and elderly);
and J) examining retention over even longer periods of time. Future pertinent studies should also expand on the very recent work of Karpicke and Bauernschmidt (9) by using (only) initially recallable information when comparing the effects of multiple days of expanding versus uniform schedules of retrieval on long-term retention.

In conclusion, the results of this study further support those of Storm et al. (16) showing that expanding retrieval practice may facilitate greater retention than does uniform retrieval practice when the to-be-learned information is vulnerable to forgetting. This investigation, along with the previous study conducted by the author (7), are the only two that have provided evidence of that advantage using retrieval assessments that were conducted over the course of multiple days. The present study advances the findings of the previous study by demonstrating that the benefit of expanding retrieval practice can remain throughout at least a 1-mo period, particularly when the first recall assessment is administered immediately after learning. Educators wishing to use retrieval practice should consider the potential advantages of using expanding schedules; however, it is critical to acknowledge that further study is needed to clarify both the conditions under which expanding retrieval may be advantageous to students and the extent of that potential benefit. At this time, the more general and more reliable pertinent recommendation to educators is to administer the first recall assessment immediately after learning has occurred and then the second recall assessment soon thereafter.

DISCLOSURES
No conflicts of interest, financial or otherwise, are declared by the author(s).

AUTHOR CONTRIBUTIONS
J.L.D., conception and design of the research; J.L.D., performed the experiments; J.L.D., analyzed the data; J.L.D., interpreted the results of the experiments; J.L.D., prepared the figures; J.L.D., drafted the manuscript; J.L.D., edited and revised the manuscript; J.L.D., approved the final version of the manuscript.

REFERENCES