

Distributed Concept Reviews Improve Exam Performance

Rachael N. Blasiman¹

Teaching of Psychology
2017, Vol. 44(1) 46-50
© The Author(s) 2016
Reprints and permission:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/0098628316677646
journals.sagepub.com/home/top



Abstract

Distributed practice is a learning strategy in which studying is distributed, or spaced, across multiple study sessions. Another learning technique, interleaved practice, mixes material from multiple lectures. I designed this study to test the effect of distributed concept reviews of interleaved material on exam scores in an introductory psychology course. Students who received the concept review outperformed students who did not receive the review—a result driven by exam questions related to concepts presented in the review itself. In fact, the number of times a concept was presented in the review was directly related to the likelihood of a correct response on the exam. These results indicate that distributed, interleaved concept reviews are an effective method of improving student learning in broad introductory courses.

Keywords

distributed practice, interleaved practice

Introductory courses often include a broad array of topics, concepts, and theories as well as a great deal of discipline-specific terminology and content. As a result, students may feel overwhelmed at the amount of information they are expected to master for a single exam. Although recently covered topics may be fresh in their minds, they may have difficulty remembering topics covered earlier in the semester. To make matters worse, students often select ineffective study strategies to encode information from class lectures (Bjork, Dunlosky, & Kornell, 2013; Cepeda et al., 2009; Roediger & Pyc, 2012) or cram for exams at the last minute.

Many study strategies are available to students, with varying levels of effectiveness, but one promising strategy is interleaved practice. Interleaved practice is a study technique, in which material from multiple lectures or chapters is studied within the same session. This technique is different from blocked practice, in which a single type of problem or topic is studied, and from distributed practice, in which studying is spaced over time—although interleaved practice generally includes an element of spacing (Roediger & Pyc, 2012). The current study uses a teacher-led intervention to model interleaved practice and distributed practice with the aim of enhancing student performance on exams.

Distributed practice has been supported by a long history of research endeavors (see Cepeda, Pashler, Vul, Wixted, & Rohrer, 2006, for a review). Despite consistent evidence that distributed practice is far superior to massed practice (Son & Simon, 2012), it remains underused in education (Dempster, 1988; Gerbier & Toppino, 2015; Rohrer, 2015; Seabrook, Brown, & Solity, 2005).

Multiple researchers have commented on the deficits in the distributed practice literature; in particular, Rohrer (2015) identified three areas requiring further investigation. First, experiments need to take place over a longer time period, on the scale of weeks instead of hours or days (Bahrick & Hall, 2005; Gerbier & Toppino, 2015). In his recent review, Rohrer (2015) concluded that approximately a month or slightly longer is the ideal time period for distributed studying.

Second, many investigations of distributed practice lack ecological validity (Rohrer, 2015). Researchers need to conduct more studies in educational settings rather than in the laboratory (Dempster, 1988; Seabrook et al., 2005). Son and Simon (2012) think that implementing distributed practice in the classroom is challenging due to the amount of material that needs to be covered and the lack of time in which to cover it. In addition, people often spontaneously choose blocked practice over distributed practice (Kornell & Bjork, 2008) and changing natural inclinations can take time and effort.

Third, the material used in the experiments needs to be more complex (Rohrer, 2015; Son & Simon, 2012). Although the benefits of distributed practice have been examined using vocabulary (Sobel, Cepeda, & Kapler, 2011), math (Budé, Imbos, van de Wiel, & Berger, 2011), and foreign language (Bird,

¹ Department of Psychological Sciences, Kent State University at Salem, Salem, OH, USA

Corresponding Author:

Rachael N. Blasiman, Department of Psychological Sciences, Kent State University Salem, 2491 State Route 45 South, Salem, OH 44460, USA.
Email: rvolokho@kent.edu

2010), few studies have used classroom stimuli and even fewer have used the more abstract concepts often covered in college-level courses.

Research on interleaved practice seems to suffer from many of the same limitations as distributed practice. As several researchers have noted (Roediger & Pyc, 2012; Rohrer, Dedrick, & Stershic, 2014), many of the investigations of interleaved practice have been lab based rather than classroom based. No studies have reported using this technique in a college classroom. So far, the two studies done in the classroom have used seventh graders as participants (Rohrer, Dedrick, & Burgess, 2014). Much of the research on the effectiveness of interleaved practice has been conducted in mathematics, with such problems as fractions and geometry (Mayfield & Chase, 2002; Rau, Alevin, & Rummel, 2013; Rohrer & Taylor, 2007; Taylor & Rohrer, 2010). Others have investigated the benefits of interleaved practice in identifying artists by their paintings (Kang & Pashler, 2012; Kornell & Bjork, 2008), bird classification (Wahlheim, Dunlosky, & Jacoby, 2011), and medical diagnostics (Hatala, Brooks, & Norman, 2003). However, no study to date has investigated the effect of interleaved practice on relatively abstract concepts, particularly at the college level.

In this article, I describe an easy-to-implement intervention that incorporates distributed and interleaved practice into an introductory psychology course. Instead of attempting to shape students' often ineffective study habits outside of the classroom, this distributed concept review (DCR) takes advantage of effective techniques to improve their learning during each class. To create the intervention, I first identified several key concepts from each prepared lecture. At the start of each class, I reviewed key concepts from the previous lecture. In addition, randomly selected concepts from earlier lectures were included in this preclass review, creating an interleaved study session. In this way, many concepts were presented multiple times over the course of a unit, and these concepts were spaced out over time. Thus, the intervention used both distributed and interleaved practice.

I developed two hypotheses regarding this study. First, students who received the concept review intervention would outperform students who did not receive the intervention, especially on questions directly related to the concepts reviewed. Second, the number of times a concept appeared in the reviews would be related to performance, so that increased concept dosage would result in enhanced performance on the exam.

Method

Participants

All participants were students enrolled at a regional campus of a public university in a relatively rural area. Demographic information was not collected as part of this study, but previous research and instructor observation indicate that the student population is predominantly female, mostly Caucasian, and

strongly nontraditional (i.e., either over 25 years or having children in the home).

Procedure

Four introductory psychology classes over the course of three semesters were sampled for this study. All students who participated in this study signed an informed consent document to allow their exams to be used in analyses. All students agreed to participate. No other information besides exams was collected, and students' names were not associated with their data once coding was complete. I taught all four sections of the course and used all the same lecture materials and exams for each of the four classes. In addition, all classes received the same exam review in the lecture period just before an exam. Both exams consisted of 50 multiple choice questions. Only students who completed the course were included in analyses, even if they completed with a failing grade.

Two of the four classes received the concept review intervention and two classes did not. The control classes did not receive any special instruction beyond my normal level of assistance and the preexam reviews already in place. Importantly, the experimental classes did not receive any special instruction for Exam 1 to allow for a baseline comparison. After Exam 1, I provided the experimental classes with concept reviews at the start of every lecture. Each concept review contained ideas from the previous lecture in addition to randomly chosen concepts from earlier lectures. The pattern was an interleaved design; concepts from several lectures and chapters were mingled together with concepts from the previous lecture. Using this system, many concepts were presented twice, with several concepts presented 3 times. As the instructor, I determined the order of concepts and the frequency at which they were presented but used a simple randomization scheme in which concepts were assigned letters of the alphabet. The letters were randomly selected for each review.

Each concept review took between 5 and 10 min at the start of every class. First, I presented the list of concepts on a PowerPoint slide. I read the first concept aloud and prompted students to describe the concept. Based on the students' responses, I provided feedback that either confirmed a correct answer or altered an incorrect answer. If no response was given after 5 s, I provided an answer or explanation. Because of frequent in-class assignments and quizzes, I estimate attendance in both control and experimental classes at 90% or better, indicating that nearly all students received the concept reviews.

At the end of each semester, I collected all scan sheets from the two exams given in each class and entered the data into SPSS (Version 21). I entered a score of 1 for each question answered correctly and 0 for each incorrect answer, computed a total score for each exam, then transformed this score into a percentage. I reviewed the final exam to determine which questions matched concepts from the intervention and marked questions with the number of times the concept was presented in order to test the effect of concept dosage.

Table 1. Comparison of Course Performance between Groups (in Percentage Points).

Measure	Control Group		Experimental Group		Cohen's <i>d</i>
	Mean	Standard Deviation	Mean	Standard Deviation	
Exam 1	79.73	12.38	81.13	13.92	.07
Exam 2	80.11	11.34	86.66	11.68	.56
Final grade	82.60	14.95	86.45	12.73	.27

Results

Descriptives

A total of 98 students participated in this study, with 51 students in the control group and 47 students in the experimental group.

Overall Exam Grades

Table 1 presents means, standard deviations, and effect sizes for both the control and experimental group. I used a two-way mixed analysis of variance to test the between-subject effect of control and experiment group and the within-subject effect of exam. There was a significant main effect of group, $F(1, 93) = 5.68, p = .02$, for the two exams as well as a significant interaction between group and exam, $F(1, 93) = 10.39, p = .01$. As expected, the difference between groups for Exam 1 was not significant, a result supported by near 0 Cohen's *d* effect size value. Thus, the two groups were similar in ability at the start of the semester. Most important, the group that received the experimental intervention scored significantly higher on Exam 2, $t(92) = 2.76, p = .01$. Further, Cohen's effect size value ($d = .56$) suggests a moderate significance. However, final grades in the course for the two groups were similar, $t(96) = 1.36, p = .18$. Final grades in the course included 10 quizzes, in-class assignments, a research paper, and the two exams, with the exams accounting for nearly half of the points in the class.

Effect of Dosage

To evaluate whether the concepts included in the intervention were driving the higher test scores seen on Exam 2, rather than the mere presence of the intervention, individual questions were examined. The 50 questions on Exam 2 were divided into four categories: questions that tested (a) concepts not part of the intervention (25 questions), (b) concepts present in the review once (10 questions), (c) concepts presented twice (10 questions), and (d) concepts presented 3 times (5 questions). A total was created for each question type and converted to a percentage score. Table 2 presents means, standard deviations, and effect sizes for these four question types divided into control and experimental groups.

I used a two-way mixed analysis of variance to compare the control and experimental groups on the different levels of

Table 2. Individual Analysis of Exam Questions Based on Dosage (in % Correct).

Dosage	Control Group		Experimental Group		Cohen's <i>d</i>
	Mean (% Correct)	Standard Deviation	Mean (% Correct)	Standard Deviation	
Concepts not in review	76.79	12.82	77.18	12.04	.03
Concepts presented 1x	70.86	15.57	77.57	11.57	.51
Concepts presented 2x	71.75	16.58	78.58	15.75	.43
Concepts presented 3x	62.67	31.67	81.00	26.00	.66

dosage. The effect of dosage was significant, $F(4, 98) = 21.24, p < .001$, as was the interaction between dosage and group, $F(4, 98) = 3.65, p = .01$. In terms of concepts not in the review, no difference was observed between the control and experimental group. However, a significant difference between groups emerged for concepts presented once in the concept review, $t(100) = 2.50, p = .02$; twice, $t(100) = 2.14, p = .04$; and 3 times, $t(100) = 3.26, p = .01$. Further, Cohen's effect size values suggest a moderate significance, with the largest effect size for concepts presented 3 times. The difference between groups was most pronounced for concepts presented 3 times, indicating that the number of times a concept is presented directly influences student performance.

Discussion

The results of this study strongly support the hypotheses. Most important, initiation of the DCR using an interleaved schedule of practice led to a significant advantage on the final exam. Students in the experimental condition performed 8% better on the final exam than students in the control condition. This advantage was specifically driven by more correct answers on questions that tested information from the concept reviews, indicating that simply the presence of a regular review was not the cause of improvement. Second, these results provide some evidence that the number of times a concept is reviewed is related to exam performance. Although students in the experimental group performed better on all questions related to the concept reviews, concepts that were covered most often were more often answered correctly. This outcome indicates that review dosage has a direct and measureable effect on student learning and later recall.

The results of this study also indicate that the effects of distributed, interleaved practice are maintained over time. Dunlosky, Rawson, Marsh, Nathan, and Willingham (2013) commented that one limitation of research in this area is that many studies test recall after an interval of 2 weeks or less.

Besides the current study, only one other interleaving study has examined criterion performance over a long interval. Rohrer, Dedrick, and Stershic (2014) had seventh graders use interleaved practice while learning to solve math problems. They implemented both immediate and delayed testing (1 and 30 days) and found that interleaved practice was better than blocked practice at both times but that the benefit of interleaved practice was greatest at the 30 day mark. In the current study, the delay between the DCRs and the final exam varied between 1 and 5 weeks.

It is interesting to note that prior research on the effect of interleaved practice on mathematics has shown much greater grade improvements than in the current study. The difference in test scores between interleaved and blocked practice at the one day delay was 16%, but a 38% difference was observed at day 30 in the study conducted by Rohrer, Dedrick, and Stershic (2014). Rohrer and Taylor (2007) found that interleaved practice tripled test scores after a 1-week delay. One possible reason for this difference is the restriction of range seen in a college-level population although the students in this study were enrolled in an open admissions university and were from an academic population that was fairly diverse. A second possible reason for this difference may lie in the nature of the material itself: the difference between the expectations and cognitive demands in mathematics and the expectations and cognitive demands in psychology.

Given the success of this intervention, the next step involves discovering its active ingredients, which may provide insight into improving its efficacy. For instance, the DCR may highlight important to-be-learned concepts by virtue of their inclusion in the review session. In this way, students may believe these concepts are the most important and may subsequently prioritize them for restudy. Another reason why the intervention may have worked is that it reduced illusions of learning. Illusions of learning occur when a learner overestimates how well they have learned the material (Koriat, 1997; Koriat & Bjork, 2005). Students who struggled with recalling information during the concept reviews may have gained a more realistic picture of their own knowledge base, which may have encouraged them to study more for the final exam. Bahrick and Hall (2005) asserted that distributed practice leads to better metacognitive monitoring, that is, students realize they have forgotten earlier material and subsequently use better strategies to learn what they had forgotten.

Although DCR shows promise for improving student achievement, some cautions are required. First, simply reviewing concepts is not the same as practicing concepts. An instructor must require active participation from students, rather than a passive presentation of information, in order for a review to be effective (Carvalho & Goldstone, 2015). Second, the benefits of interleaved practice may not be readily apparent during the practice session itself. Indeed, interleaved testing can impair immediate practice performance (Taylor & Rohrer, 2010). Rohrer and Pashler (2010) attribute this effect to the greater cognitive demands associated with interleaved practice. Their hypothesis is supported by functional magnetic resonance

imaging (fMRI) studies, which show greater frontoparietal activity during interleaved practice compared to block practice (Lin et al., 2011). Lin et al. (2013) attribute this activity to the formation of stronger memory traces and enhanced cooperation between brain regions. Third, although this intervention did improve student performance, it is impossible to separate the unique contribution of distributed practice and interleaved practice. Future work may attempt to separate these two aspects to determine whether the distributed or interleaved aspect of the DCR is driving the effect or if aspects contribute equally.

Although this study requires replication, these initial results are encouraging. The intervention described in this study is easy to implement in any lecture class. It does take some time and thought to prepare the concept reviews in advance, but the benefit to students is a strong motivator. Although there were no differences in overall course evaluations between groups, informal verbal inquiries yielded a positive response from students in the experimental condition. Students felt that the concept reviews helped to clarify the most important ideas from the previous lecture and reminded them of topics that had been covered earlier in the semester. In addition to these benefits, this intervention is entirely cost free, which is surely a boon to educators. Future work in this area should examine the effectiveness of this technique in other disciplines, in online classes, and in classes of various sizes. Another suggestion might be to include students in the process of creating the concept reviews to give them a level of ownership and to foster critical thinking. Implementing a DCR using real-time technology or clickers is also a viable option. Based on the results of this study, DCRs using interleaved material are a promising teaching strategy that can improve student performance in content-heavy college-level courses.

Acknowledgments

I am grateful to John Dunlosky and three anonymous reviewers for their suggestions and comments on this article.

Declaration of Conflicting Interests

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author received no financial support for the research, authorship, and/or publication of this article.

References

- Bahrick, H. P., & Hall, L. K. (2005). The importance of retrieval failures to long-term retention: A metacognitive explanation of the spacing effect. *Journal of Memory and Language*, *52*, 566–577. doi:10.1016/j.jml.2005.01.012
- Bird, S. (2010). Effects of distributed practice on the acquisition of second language English syntax. *Applied Psycholinguistics*, *31*, 635–650. doi:10.1017/S0142716410000172
- Bjork, R. A., Dunlosky, J., & Kornell, N. (2013). Self-regulated learning: Beliefs, techniques, and illusions. *Annual Review of Psychology*, *64*, 417–444. doi:10.1146/annurev-psych-113011-143823

- Budé, L., Imbos, T., van de Wiel, M. W., & Berger, M. P. (2011). The effect of distributed practice on students' conceptual understanding of statistics. *Higher Education, 62*, 69–79. doi:10.1007/s10734-010-9366-y
- Carvalho, P. F., & Goldstone, R. L. (2015). The benefits of interleaved and blocked study: Different tasks benefit from different schedules of study. *Psychonomic Bulletin Review, 22*, 281–288. doi:10.3758/s13423-014-0676-4
- Cepeda, N. J., Mozer, M. C., Coburn, N., Rohrer, D., Wixted, J. T., & Pashler, H. (2009). Optimizing distributed practice: Theoretical analysis and practical implications. *Experimental Psychology, 56*, 236–246. doi:10.1027/1618-3169.56.4.236
- Cepeda, N. J., Pashler, H., Vul, E., Wixted, J. T., & Rohrer, D. (2006). Distributed practice in verbal recall tasks: A review and quantitative synthesis. *Psychological Bulletin, 132*, 354–380. doi:10.1037/0033-2909.132.3.354
- Dempster, F. N. (1988). The spacing effect: A case study in the failure to apply the results of psychological research. *American Psychologist, 43*, 627–634. doi:10.1037//0003-066X.43.8.627
- Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving students' learning with effective learning techniques: Promising directions from cognitive and educational psychology. *Psychological Science in the Public Interest, 14*, 4–58. doi:10.1177/1529100612453266
- Gerbier, E., & Toppino, T. C. (2015). The effect of distributed practice: Neuroscience, cognition, and education. *Trends in Neuroscience and Education, 4*, 49–59. doi:10.1016/j.tine.2015.01.001
- Hatala, R. M., Brooks, L. R., & Norman, G. R. (2003). Practice makes perfect: The critical role of mixed practice in the acquisition of ECG interpretation skills. *Advances in Health Sciences Education, 8*, 17–26. doi:10.1023/A:1022687404380
- Kang, S. H. K., & Pashler, H. (2012). Learning painting styles: Spacing is advantageous when it promotes discriminative contrast. *Applied Cognitive Psychology, 26*, 97–103. doi:10.1002/acp.1801
- Koriat, A. (1997). Monitoring one's own knowledge during study: A cue-utilization approach to judgments of learning. *Journal of Experimental Psychology: General, 126*, 349–370. doi:10.1037/0096-3445.126.4.349
- Koriat, A., & Bjork, R. A. (2005). Illusions of competence in monitoring one's knowledge during study. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*, 187–194. doi:10.1037/0278-7393.31.2.187
- Kornell, N., & Bjork, R. A. (2008). Learning concepts and categories: Is spacing the “enemy of induction”? *Psychological Science, 19*, 585–592. doi:10.1111/j.1467-9280.2008.02127.x
- Lin, C.-H., Chiang, M.-C., Knowlton, B. J., Iacoboni, M., Udompholkul, P., & Wu, A. D. (2013). Interleaved practice enhances skill learning and the functional connectivity of fronto-parietal networks. *Human Brain Mapping, 34*, 1542–1558. doi:10.1002/hbm.22009
- Lin, C.-H., Knowlton, B. J., Chiang, M.-C., Iacoboni, M., Udompholkul, P., & Wu, A. D. (2011). Brain-behavior correlates of optimizing learning through interleaved practice. *Neuroimage, 56*, 1758–1772. doi:10.1016/j.neuroimage.2011.02.066
- Mayfield, K. H., & Chase, P. N. (2002). The effects of cumulative practice on mathematics problem solving. *Journal of Applied Behavior Analysis, 35*, 105–123. doi:10.1901/jaba.2002.35-105
- Rau, M. A., Aleven, V., & Rummel, N. (2013). Interleaved practice in multi-dimensional learning tasks: Which dimension should we interleave? *Learning and Instruction, 23*, 98–114. doi:10.1016/j.learninstruc.2012.07.003
- Roediger, H. L., & Pyc, M. A. (2012). Inexpensive techniques to improve education: Applying cognitive psychology to enhance educational practice. *Journal of Applied Research in Memory & Cognition, 1*, 242–248. doi:10.1016/j.jarmac.2012.09.002
- Rohrer, D. (2015). Student instruction should be distributed over long time periods. *Education Psychology Review, 1–9*. doi:10.1007/s10648-015-9332-4
- Rohrer, D., Dedrick, R. F., & Burgess, K. (2014). The benefit of interleaved mathematics practice is not limited to superficially similar kinds of problems. *Psychonomic Bulletin & Review, 21*, 1323–1330. doi:10.3758/s13423-014-0588-3
- Rohrer, D., Dedrick, R. F., & Stershic, S. (2014). Interleaved practice improves mathematics learning. *Journal of Educational Psychology, 107*, 900–908. doi:10.1037/edu0000001
- Rohrer, D., & Pashler, H. (2010). Recent research on human learning challenges conventional instructional strategies. *Educational Researcher, 39*, 406–412. doi:10.3102/0013189x10374770
- Rohrer, D., & Taylor, K. (2007). The shuffling of mathematics practice problems boosts learning. *Instructional Science, 35*, 481–498. doi:10.1007/s11251-007-9015-8
- Seabrook, R., Brown, G. D. A., & Solity, J. E. (2005). Distributed and massed practice: From laboratory to classroom. *Applied Cognitive Psychology, 19*, 107–122. doi:10.1002/acp.1066
- Sobel, H. S., Cepeda, N. J., & Kapler, I. V. (2011). Spacing effects in real-world classroom vocabulary learning. *Applied Cognitive Psychology, 25*, 763–767. doi:10.1002/acp.1747
- Son, L. K., & Simon, D. A. (2012). Distributed learning: Data, metacognition, and educational implications. *Educational Psychology Review, 24*, 379–399. doi:10.1007/s10648-012-9206-y
- Taylor, K., & Rohrer, D. (2010). The effect of interleaving practice. *Applied Cognitive Psychology, 24*, 837–848. doi:10.1002/acp.1598
- Wahlheim, C. N., Dunlosky, J., & Jacoby, L. L. (2011). Spacing enhances the learning of natural concepts: An investigation of mechanisms, metacognition, and aging. *Memory & Cognition, 39*, 750–763. doi:10.3758/s13421-010-0063-y