

Teaching of Psychology

<http://top.sagepub.com/>

Retrieving Essential Material at the End of Lectures Improves Performance on Statistics Exams

Keith B. Lyle and Nicole A. Crawford
Teaching of Psychology 2011 38: 94
DOI: 10.1177/0098628311401587

The online version of this article can be found at:
<http://top.sagepub.com/content/38/2/94>

Published by:



<http://www.sagepublications.com>

On behalf of:



<http://www.sagepub.com/content/38/2/94>
Society for the Teaching of Psychology

Additional services and information for *Teaching of Psychology* can be found at:


Email Alerts: <http://top.sagepub.com/cgi/alerts>

Subscriptions: <http://top.sagepub.com/subscriptions>

Reprints: <http://www.sagepub.com/journalsReprints.nav>

Permissions: <http://www.sagepub.com/journalsPermissions.nav>

Retrieving Essential Material at the End of Lectures Improves Performance on Statistics Exams

Teaching of Psychology
38(2) 94-97
© The Author(s) 2011
Reprints and permission:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/0098628311401587
http://top.sagepub.com


Keith B. Lyle¹ and Nicole A. Crawford¹

Abstract

At the end of each lecture in a statistics for psychology course, students answered a small set of questions that required them to retrieve information from the same day's lecture. These exercises constituted retrieval practice for lecture material subsequently tested on four exams throughout the course. This technique is called the PUREMEM (pronounced "pure mem") procedure for Pure Memory or Practicing Unassisted Retrieval to Enhance Memory for Essential Material. Exam scores were significantly and substantially higher in a section of the course taught with the PUREMEM procedure than one taught without it. Students liked the procedure and believed it increased learning via several different mechanisms.

Keywords

learning, memory, retrieval, statistics, testing effect

Instructors spend many hours developing lectures, and students spend many hours attending them, but students' retention of lecture material is seemingly often poor. Here we present a technique for enhancing retention of essential lecture material by requiring students to retrieve the material from memory before leaving class. We were inspired by research showing that retrieving information from memory soon after acquisition increases long-term retention (Roediger & Karpicke, 2006). For example, Butler and Roediger (2007) showed that taking a short-answer test covering lecture material immediately after lectures increased long-term memory for the material by 31% relative to studying a summary of the material and by 135% relative to doing nothing with the material. Taking a test soon after acquiring information is a form of practicing retrieval in preparation for future tests and retrieval attempts. Butler and Roediger's results, like many similar ones, suggest that practicing retrieval may increase the retention of educationally relevant material in real classrooms (McDaniel, Roediger, & McDermott, 2007).

Research demonstrating the value of retrieval practice caused us, like others (e.g., Gates, 1917; Jones, 1923), to see tests as not only assessment devices but also potential learning aids for lecture material. Therefore, we incorporated retrieval practice into a course on statistics for psychology by administering brief tests requiring retrieval of essential lecture material at the end of each lecture. We dubbed the tests *PUREMEMs* (pronounced "pure mems"), which we told students they could think of as standing for either pure memory or, more formally, Practicing Unassisted Retrieval to Enhance Memory for Essential Material. We avoided the labels *tests* and *quizzes* because

these words primarily connote assessment rather than memory enhancement.

The PUREMEM procedure's explicit emphases on retrieval practice versus assessment and on memory for lecture material versus assigned readings differentiate it from other procedures involving frequent testing (cf. Landrum, 2007; Narloch, Garbin, & Turnage, 2006; Padilla-Walker, 2006). For example, in Leeming's (2002) exam-a-day procedure, students take an exam covering material from the preceding lecture and a textbook reading at the start of every class. The exams have a critical assessment function because scores on them are central to students' course grades.

Next we describe the PUREMEM procedure in detail and compare exam performance in two sections of a statistics course with and without PUREMEMs. We also report data on students' attitudes toward PUREMEMs and beliefs about how PUREMEMs enhanced learning. Theoretically, PUREMEMs enhance learning (specifically, retention) by inducing retrieval practice, but having retrieval practice at the end of lectures may help in additional ways, such as by causing students to attend more during lectures. We had students evaluate several mechanisms by which PUREMEMs conceivably may affect learning.

¹University of Louisville, Louisville, KY

Corresponding Author:

Keith B. Lyle, University of Louisville, Department of Psychology, 2301 S. 3rd St., Louisville, KY 40292
Email: keith.lyle@louisville.edu

Method

Participants

The first author taught two sections of an undergraduate course on statistics for psychology in consecutive years. Only the second section included PUREMEMs. The first section served as a comparison. There were initially 77 students in the PUREMEM section and 78 in the comparison section, with one withdrawal in the former (1.3%) and five in the latter (6.4%), a nonsignificant difference, $\chi^2(1) = 1.52, p = .22$. Students in the two sections were indistinguishable on demographic and academic variables collected anonymously on a university-mandated course evaluation form. Across the two sections, 70% were 20 to 24 years old, 75% were female, 19% were minority or international students, 48% had at least a 3.0 GPA, 73% had accumulated at least 51 credit hours, and 41% had employment or other extracurricular commitments for 21 to 40 hours per week (for tests of between-section differences, all $ps > .15$).

Procedure

Class format. The courses had two 75-minute lectures per week. The style and content of lectures were the same in the PUREMEM and comparison sections, as were the syllabus and grade scale. Students took four noncumulative, multiple-choice exams covering the lecture material; exams were evenly distributed throughout the semester. Exams in the two sections covered the same material, were of comparable length, contained four-alternative questions, and consisted of a mixture of computational, noncomputational, and application problems. Between sections, there were slight differences in exams restricted to superficial elements of the problems, such as specific numbers in computational problems or details of scenarios in application problems.

PUREMEM procedure. Administration of PUREMEMs occurred after each lecture during the final 5 to 10 minutes of class. Before the PUREMEM, students put away notes and took out paper on which to write answers. Then several questions—most often four, but ranging from two to six—were projected from a PowerPoint slide (see Table 1 for examples). The questions probed material from the same day’s lecture only, specifically material the instructor deemed most essential. For most questions, the correct answer consisted of one or two words, a number, or a simple figure. We did not use multiple-choice questions because they may not enhance long-term retention as much as do more open-ended questions (McDaniel et al., 2007). Students submitted their answers before leaving class. We projected the correct answers from a PowerPoint slide at the start of the next class and posted them on a course website. Students could review the answers and ask questions before the new lecture. If there were no questions, the feedback phase did not consume any class time.

Although assessment was not the primary purpose of PUREMEMs, the second author scored them, and they had

Table 1. Example PUREMEMs Related to Each of the Four Exams

Exam	PUREMEM
1	<p>All distributions of variables show two things: The values the variables could possibly have and what else about the values?</p> <p>Frequency tables can be made to organize data from which type of variable(s)?</p> <p>Histograms can be made to organize data from which type of variable(s)?</p> <p>Without worrying about any details, draw the basic shape of a unimodal, symmetrical distribution.</p>
2	<p>The _____ is the most commonly given representative value for a numeric variable.</p> <p>What is the measure of central tendency for a categorical variable?</p> <p>If you measure the exact same behavior in two different samples and find that scores are more spread out in Sample B than in Sample A, then it must be that both the _____ and the _____ are larger in Sample B than in Sample A.</p> <p>What do we call the number that gives the average amount by which scores in a sample differ from the mean score?</p>
3	<p>Suppose you see the following: $t(60) = 2.40, p < .05$. What are the degrees of freedom and what is the t score?</p> <p>How many scores do you have from each subject when you do a t test for independent means?</p> <p>Suppose the difference between your group means is +4 and the SD of the comparison distribution is 2. Change the difference between the means into a t score.</p> <p>A Type I error occurs when we _____ the null hypothesis even though the null hypothesis is _____.</p>
4	<p>r values can be any number from _____ to _____.</p> <p>Which correlation is stronger: One that has an r value of $-.60$ or one that has an r value of $+.47$?</p> <p>Would you predict that number of classes attended in a semester is <i>positively</i> or <i>negatively</i> correlated with final grades?</p> <p>Suppose the correlation between height and weight is $+.50$. What percentage of the variability in weight is explained by differences in height?</p>

a small weight in the grading scheme to motivate students to take them seriously. Students’ mean score for all PUREMEMs counted for 8% of their overall course grade. There were 21 PUREMEMs in total, so individual PUREMEMs counted for only .38% of one’s overall grade. Because individual scores had little impact on grades, we did not allow students to drop or make up any low or missing scores, thereby minimizing administrative load. Students received a zero for missed PUREMEMs.

At the end of the semester, students in the PUREMEM section anonymously completed an evaluation form concerning PUREMEMs. Table 2 shows the questions on the form. Sixty-five students (86%) completed the evaluation.

Table 2. Mean Student Evaluations of PUREMEMs (with SDs)

Question	Rating
At the present time, how would you describe your overall feelings about the use of PUREMEMs in this course? ^a	5.7 (1.4)
To what extent do you think PUREMEMs helped you in this course specifically by . . . ^b	
Giving you a chance to practice answering questions that were similar to those on problem sets and exams?	6.0 (1.3)
Helping you identify some of the important topics that were going to be covered on problem sets and exams?	6.0 (1.2)
Causing you to come to lecture more than you otherwise would have?	5.9 (1.5)
Causing you to pay more attention during lecture than you otherwise would have?	5.8 (1.5)
Giving you a better sense of how much you had learned in each lecture than you otherwise would have had?	5.8 (1.3)

a. The response scale ranged from 1 (*very negative*) to 7 (*very positive*).

b. The response scale for the five questions below ranged from 1 (*very little*) to 7 (*a lot*).

Relationship between PUREMEM and exam questions. PUREMEM questions targeted essential lecture material and, logically, exams included questions that targeted the same material. However, PUREMEM questions did not appear verbatim on exams. As stated, PUREMEM involved short answer questions, whereas exams had multiple-choice questions. PUREMEM questions were often the basis for the stems of exam questions, but although the conceptual focus was the same in those cases, the exact wording or numbers differed. For example, the exam question targeting the same material as the third example PUREMEM question in Table 1 was “Which of the following is a way to show the distribution of a numeric variable? A) A histogram, B) A bar graph, C) Both (A) and (B), D) Neither (A) nor (B).”

Results

Retrieval practice can increase the long-term retention of information but presumably only if retrieval attempts are successful (e.g., Kang, McDermott, & Roediger, 2007). To determine whether students could retrieve lecture material immediately after its presentation, we calculated each student’s mean PUREMEM score, excluding scores of 0 for missed PUREMEMs because there was no retrieval attempt on those exercises. The class mean was 86% ($SD = 10\%$). Thus, retrieval attempts on PUREMEMs were usually successful.

To analyze exam performance, because proportions often violate the homogeneity and normality assumptions of inferential statistical tests, we applied the arcsine transformation to proportion correct on exams. Significance test statistics present analyses using the transformed variable as the dependent measure. Descriptive statistics present raw proportions to retain interpretability. We submitted proportion correct on the four exams to a 2 (section) \times 4 (exam) mixed-design ANOVA in which the first factor was between participants and the second was within. We included exam scores from only those students

who took all four exams ($ns = 74$ and 70 in the PUREMEM and comparison sections, respectively). Overall, exam scores were significantly higher in the PUREMEM section ($M = .86$, $SD = .09$) than in the comparison section ($M = .78$, $SD = .14$), $F(1, 142) = 11.23$, $MSE = .42$, $p = .001$, $\eta_p^2 = .073$. There was also a significant main effect of exam, $F(3, 426) = 8.65$, $MSE = .066$, $p < .001$, $\eta_p^2 = .057$, and, more importantly, a significant interaction, $F(1, 142) = 13.07$, $MSE = .066$, $p < .001$, $\eta_p^2 = .084$. The interaction was such that the mean score on Exams 1, 2, and 4 was .08 to .13 higher in the PUREMEM section. The PUREMEM advantage was significant for each of these exams individually, smallest $t(142) = 3.16$, $p = .002$, Cohen’s $d = 0.55$. However, mean scores were identical in the two sections on Exam 3. Inspection of the means revealed that students in the comparison section scored markedly higher on Exam 3 than on any of the other exams, suggesting that students can sometimes achieve good retention of lecture material without end-of-lecture retrieval practice. Nevertheless, PUREMEMs substantially improved performance on the majority of exams. Correspondingly, far fewer students earned mean exam scores lower than 70% in the PUREMEM section (5.4%) than in the comparison section (27.1%).

Controlling for number of PUREMEMs taken, which indexes lecture attendance, the partial correlation between mean PUREMEM scores (excluding missed PUREMEMs) and mean exam scores was significantly positive, $r(71) = .66$, $p < .001$.

As shown in Table 2, students’ attitudes toward PUREMEMs were very positive, and students believed that PUREMEMs were of considerable help in all five ways queried.

Discussion

Adding a brief retrieval exercise for essential lecture material at the end of every lecture in a statistics course significantly and substantially increased exam scores. Students liked the

retrieval practice and believed it was helpful. Specifically, students indicated that PUREMEMs helped them identify important topics, monitor their learning, increase their attendance, and pay more attention to lectures. Although we obtained these benefits in a statistics course, we see no reason not to expect that PUREMEMs would produce similar benefits in any course involving informationally rich lectures.

The PUREMEM procedure is potent and simple to implement, but concerns may exist about the amount of time necessary to administer and score PUREMEMs. Leeming (2002) addressed the same concerns regarding the exam-a-day procedure mentioned in the introduction, and our response is similar. First, it was possible to compensate for time spent in class on PUREMEMs by delivering more focused lectures. Although the 5 to 10 minutes devoted to PUREMEMs could be spent presenting additional information, we see no point in presenting information if the likelihood of retention is low. Second, because PUREMEMs were brief and it was not necessary to provide written feedback, scoring took only about 30 minutes.

We conclude by observing that retrieval practice is seemingly a highly flexible means of enhancing memory for lecture material. We used one particular method of implementing retrieval practice, but we can imagine numerous variations. Aspects of the procedure that different instructors might alter to better suit the needs of their own teaching styles or class formats include, for example, the number of questions, whether questioning occurs spaced throughout lectures or exclusively at the end, and the weight given to retrieval practice exercises in the grading scheme. Regardless of the details, we believe instructors should consider harnessing the power of retrieval practice to enhance memory for the lectures they so painstakingly develop.

Acknowledgment

We thank James M. Edlin for assistance with preparing this article.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interests with respect to the authorship and/or publication of this article.

Financial Disclosure/Funding

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

References

- Butler, A. C., & Roediger, H. L., III. (2007). Testing improves long-term retention in a simulated classroom setting. *European Journal of Cognitive Psychology, 19*, 514-527.
- Gates, A. I. (1917). Recitation as a factor in memorizing. *Archives of Psychology, 6*, No. 40.
- Jones, H. E. (1923). The effects of examination on the performance of learning. *Archives of Psychology, 10*, 1-70.
- Kang, S. H. K., McDermott, K. B., & Roediger, H. L., III. (2007). Test format and corrective feedback modify the effect of testing on long-term retention. *European Journal of Cognitive Psychology, 19*, 528-558.
- Landrum, R. E. (2007). Introductory psychology student performance: Weekly quizzes followed by a cumulative final exam. *Teaching of Psychology, 34*, 177-180.
- Leeming, F. C. (2002). The exam-a-day procedure improves performance in psychology classes. *Teaching of Psychology, 29*, 210-212.
- McDaniel, M. A., Roediger, H. L., III, & McDermott, K. B. (2007). Generalizing test-enhanced learning from the laboratory to the classroom. *Psychonomic Bulletin & Review, 14*, 200-206.
- Narloch, R., Garbin, C. P., & Turnage, K. D. (2006). Benefits of pre-lecture quizzes. *Teaching of Psychology, 33*, 109-112.
- Padilla-Walker, L. M. (2006). The impact of daily extra credit quizzes on exam performance. *Teaching of Psychology, 33*, 236-239.
- Roediger, H. L., III, & Karpicke, J. D. (2006). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science, 1*, 181-210.

Bios

Keith B. Lyle is an Assistant Professor of Psychological and Brain Sciences at the University of Louisville.

Nicole A. Crawford is a PhD candidate in the Department of Psychological and Brain Sciences at the University of Louisville.