GRADEDRAFT · SAMPLE LIBRARY

The Effect of Structured Retrieval Practice on Undergraduate Exam Performance

A complete, original research paper example written for gradedraft.com/research-paper/research-paper-examples/ and annotated at every structural move. Read the paper straight through for the argument; read the notes for the reasoning behind each choice.

Abstract

Retrieval practice — recalling information from memory rather than re-reading it — is well established in laboratory research as a durable learning strategy, but its effect in live undergraduate courses is less consistently measured. This study compares exam performance between two sections of an introductory psychology course ($N = 84$): one using weekly structured retrieval quizzes, one using weekly re-reading assignments of equal time cost. Students in the retrieval-practice section scored an average of 6.8 points higher on a cumulative final exam ($t(82) = 3.14$, $p = .002$). The effect held after controlling for prior GPA. Results support assigning low-stakes retrieval quizzes as a replacement for, not an addition to, standard reading assignments.

> **Note 1:** The abstract states the method, the sample size, the direction and size of the effect, and the statistical result — in that order — inside 150–250 words. A reader should be able to stop here and still know what the paper found.

Introduction

Undergraduate instructors routinely assign reading as the primary tool for exam preparation, yet cognitive psychology has shown for decades that re-reading produces weaker long-term retention than actively recalling material (Roediger & Karpicke, 2006). Despite this evidence, retrieval-based assignments remain rare in general-education courses, where reading quizzes are used to enforce compliance rather than to build memory.

This study tests whether replacing standard reading assignments with structured retrieval quizzes — of equal time cost to students — improves exam performance in a live undergraduate course.

We hypothesize that students assigned weekly retrieval quizzes will outperform students assigned equivalent re-reading tasks on a cumulative final exam, independent of prior academic achievement.

> **Note 2:** The bolded sentence is the thesis. It is testable, it names both conditions being compared, and it states the expected direction of the effect — the three requirements an introduction has to meet before the Method section can be trusted.

Literature Review

Retrieval practice research divides into two lines. Laboratory studies (Roediger & Karpicke, 2006; Karpicke & Blunt, 2011) consistently show that testing material — even without feedback — produces stronger retention than restudying it, an effect that survives delays of a week or more. Classroom studies are fewer and more mixed: McDaniel et al. (2007) found gains from weekly quizzing in a college course, while Batsell et al. (2017) found gains only for factual, not conceptual, exam items.

No study identified compares retrieval quizzing against a *time-matched* reading control in a live course; most classroom studies compare quizzing against no intervention at all, which confounds the effect of retrieval with the effect of simply doing more coursework. This study closes that gap by matching both conditions for time spent.

> **Note 3:** A literature review does not summarize each source in its own paragraph — it groups sources by finding, states where they agree, and names the specific gap the current paper fills. That gap sentence is what separates a literature review from an annotated bibliography.

Method

Participants. Eighty-four students enrolled in two sections of an introductory psychology course at the same institution, taught by the same instructor, using identical lecture slides and exam items.

Design. A quasi-experimental between-subjects design. Section assignment (not condition assignment) was determined by course registration, a limitation addressed below.

Materials. The retrieval-practice section completed a 10-question, no-feedback recall quiz each week, covering that week's material, taking approximately 12 minutes. The control section completed a re-reading assignment of the same material, timed at 12 minutes via a reading-log self-report.

Procedure. Both sections received identical lectures and identical exam questions. The only manipulated variable was the weekly assignment type, delivered over 10 weeks before a shared cumulative final exam.

> **Note 4:** A Method section is written so another researcher could repeat the study from this paragraph alone. Naming the exact quiz length, the exact time cost, and how time was controlled for the reading condition is what makes this section replicable rather than descriptive.

Results

The retrieval-practice section ($n = 41$) scored an average of 84.2 on the cumulative final exam ($SD = 7.9$); the control section ($n = 43$) scored an average of 77.4 ($SD = 8.6$). An independent-samples t-test found this difference significant, $t(82) = 3.14$, $p = .002$, and the effect held ($p = .006$) after covarying for prior semester GPA.

Table 1 *Final exam scores by section*

Condition	*n*	Mean	*SD*
Retrieval quiz	41	84.2	7.9
Re-reading	43	77.4	8.6

Figure 1 (described): a bar chart comparing mean scores across the two conditions, with error bars at one standard deviation, shows non-overlapping confidence intervals at the 95% level.

> **Note 5:** Results are reported without interpretation — no claim about *why* the effect occurred appears here. That argument is reserved for Discussion. Mixing the two is the single most common structural error in a student research paper.

Discussion

These results are consistent with laboratory findings that retrieval strengthens memory more than restudy, extended here to a live, graded, undergraduate course with a time-matched control — the comparison the existing classroom literature lacks.

Limitations. Section assignment was not randomized; students self-selected into course sections at registration, so a systematic difference between the two groups (motivation, schedule flexibility) cannot be fully ruled out, though the GPA covariate addresses the most likely confound. The single-institution, single-course design also limits generalization to other subjects and class sizes.

Future work should randomize condition within a single section — for example, by quiz topic rather than by student — to remove the self-selection concern entirely.

> **Note 6:** Naming the exact confound the design could not rule out, and naming the specific fix, is what an admitted limitation looks like. A paper that skips this paragraph reads as unaware of its own weakest point, which costs more credibility than the limitation itself.

References

Batsell, W. R., Perry, J. L., Hanley, E., & Hostetter, A. B. (2017). Ecological validity of the testing effect: The use of daily quizzes in introductory psychology. *Teaching of Psychology, 44*(1), 18–23.

Karpicke, J. D., & Blunt, J. R. (2011). Retrieval practice produces more learning than elaborative studying with concept mapping. *Science, 331*(6018), 772–775. https://doi.org/10.1126/science.1199327

McDaniel, M. A., Anderson, J. L., Derbish, M. H., & Morrisette, N. (2007). Testing the testing effect in the classroom. *European Journal of Cognitive Psychology, 19*(4-5), 494–513.

Roediger, H. L., & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science, 17*(3), 249–255.

> **Note 7:** Every reference uses a hanging indent — the first line sits flush left, every line after it indents half an inch. In APA 7th edition, that indent, not bolding or underlining, is what signals a reference entry.
