

Retrieval Practice and Durable Learning

A focused review of experimental and meta-analytic evidence

Spaces Literature Review Agent · 7 August 2026

Abstract

This review asks whether actively retrieving studied material produces more durable learning than rereading. Four focal publications were selected to span controlled experiments, a comparison with elaborative study, and meta-analysis. The evidence converges on a qualified conclusion: retrieval practice generally improves delayed retention relative to restudy, even when restudy creates greater immediate fluency or confidence. The size and transfer of the benefit depend on retention interval, retrieval success, feedback, practice format, and the match between practice and final assessment. The evidence supports low-stakes retrieval after initial comprehension, followed by correction and later repetition; it does not support replacing instruction with constant testing. The review distinguishes replicated findings from boundary conditions and identifies open questions about complex performance, unsuccessful retrieval, and classroom implementation.

1 Research Question and Scope

The primary question is: *When does retrieval practice outperform rereading for delayed learning?* The search frame combined “retrieval practice” or “testing effect” with retention, classroom, transfer, and meta-analysis. This concise review is a structured synthesis of four foundational publications, not a new systematic review.

Evidence status: Claims are tied to the reported designs and outcomes of the cited publications. No effect size is generalized beyond the population, material, or outcome for which it was reported.

1.1 Selection logic

The corpus was chosen to test four propositions: whether delay changes the relative advantage of test and study; whether retrieval itself is the critical event; whether retrieval can outperform an active elaborative strategy; and whether the effect survives aggregation across studies.

2 Conceptual Model

Rereading increases exposure and can raise immediate familiarity. Retrieval requires reconstruction from memory. Let delayed performance be represented by

$$P_d = f(R, F, \Delta, A, K), \quad (1)$$

where R is retrieval effort, F is feedback quality, Δ is the delay before later practice or assessment, A is alignment between practice and assessment, and K

is prior knowledge. The literature does not imply that any increase in R improves P_d ; unsuccessful retrieval without correction may contribute little or introduce error.

3 Evidence Synthesis

3.1 Delay can reverse the winner

Roediger and Karpicke (2006) compared repeated study with one or three free-recall tests after participants read prose passages. Repeated study improved recall on a five-minute test. On tests two days or one week later, prior testing produced substantially greater retention, even though repeated study increased learners’ confidence. The study directly separates short-term performance from durable access.

3.2 Retrieval is not another exposure

Karpicke and Roediger (2008) varied study and test schedules during learning. Repeated retrieval after successful recall was critical for one-week retention, while repeated study after successful recall added little. This result strengthens the causal claim that retrieval changes later accessibility rather than merely adding another encounter with the material.

3.3 An active comparison

Karpicke and Blunt (2011) compared retrieval practice with elaborative studying through concept mapping. Reported outcomes favored retrieval

Table 1: Methodological comparison of the focal literature.

Study	Design	Outcome horizon	Main contribution and boundary
Roediger & Karpicke (2006)	Two controlled experiments with prose passages	5 minutes, 2 days, 1 week	Demonstrates a delay-dependent reversal between restudy and test; laboratory prose limits direct classroom generalization.
Karpicke & Roediger (2008)	Controlled study/test schedules	One-week retention	Isolates repeated retrieval as the critical learning event; materials and procedure remain laboratory constrained.
Karpicke & Blunt (2011)	Retrieval versus concept mapping	Factual and inference learning	Tests transfer beyond verbatim recall; one comparison does not settle every use of elaborative mapping.
Adesope et al. (2017)	Meta-analysis	Multiple populations and outcomes	Establishes a robust average advantage and identifies moderators; pooled effects do not prescribe a single classroom routine.

practice for later factual and inference questions. The comparison matters because it tests retrieval against a visibly generative strategy, not passive rereading alone. It should not be interpreted as evidence that concept mapping has no useful role.

3.4 Aggregation and moderators

Adesope, Trevisan, and Sundararajan (2017) synthesized practice-testing research and found a positive average effect relative to restudy and other non-testing conditions. Effects varied with practice-test features, feedback, participants, outcomes, and study methods. The synthesis supports robustness while ruling out a one-size-fits-all implementation.

4 Consensus and Controversy

Four conclusions are well supported. First, practice testing commonly improves delayed retention relative to restudy. Second, immediate ease and confidence are unreliable indicators of later access. Third, retrieval is more consequential than repeated exposure after material has been learned. Fourth, feedback, timing, test format, and assessment alignment moderate the result.

The literature leaves several live disputes. Transfer to complex, open-ended performance is less settled than retention of studied material. The value of unsuccessful retrieval depends on what follows it. Classroom effects depend on whether practice remains low stakes, whether feedback repairs errors, and whether implementation adds learning opportunities rather than merely adding assessment.

5 Practical Implications

A conservative evidence-aligned sequence is:

1. establish first comprehension through explanation and guided practice;
2. close notes and retrieve three to five central ideas from memory;
3. provide concise corrective feedback immediately after the attempt;
4. repeat retrieval after a meaningful delay with reordered prompts; and
5. include one application question before graded assessment.

The design principle is to use retrieval as a learning event, not a surveillance event. Low-stakes conditions protect the attempt while corrective feedback protects accuracy.

6 Open Questions

Future work should test transfer to complex performance, age- and subject-specific retrieval formats, the timing of feedback after failed retrieval, and the point at which frequent quizzing becomes counterproductive. Evidence is also needed on accessibility, anxiety, and differential effects for learners with weaker prior knowledge.

7 Conclusion

For durable learning, some rereading time should be replaced with effortful retrieval. The strongest recommendation is deliberately modest: retrieve after comprehension, correct errors, and return after a delay. The evidence does not justify constant testing or the elimination of elaborative study.

References

- Adesope, O. O., Trevisan, D. A., & Sundararajan, N. (2017). Rethinking the use of tests: A meta-analysis of practice testing. *Review of Educational Research*, 87(3), 659–701. doi:10.3102/0034654316689306.
- Karpicke, J. D., & Blunt, J. R. (2011). Retrieval practice produces more learning than elaborative studying with concept mapping. *Science*, 331(6018), 772–775. doi:10.1126/science.1199327.
- Karpicke, J. D., & Roediger, H. L., III. (2008). The critical importance of retrieval for learning. *Science*, 319(5865), 966–968. doi:10.1126/science.1152408.
- Roediger, H. L., III, & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science*, 17(3), 249–255. doi:10.1111/j.1467-9280.2006.01693.x.