

Learning Science and Individual Variation in How People Learn

Steady Practice Applied Science Series — SP-16

Steady Practice Research · 2026

Abstract

The learning styles framework — the claim that people are visual, auditory, or kinesthetic learners and that matching instruction to style improves outcomes — has no credible empirical support. Pashler et al. (2008) reviewed the evidence and found that well-controlled studies consistently fail to support the “meshing hypothesis.” What does exist is robust, replicable evidence for study methods that work regardless of modality preference: spaced practice ($d \approx 0.8$), retrieval practice, interleaving of problem types, and elaborative interrogation. Individual variation in learning effectiveness is real, but it is located in the wrong place in the popular framework. What actually varies across individuals is: working memory capacity and its relationship to cognitive load, metacognitive accuracy (the ability to judge whether you have actually learned something versus merely recognized it), prior knowledge and the way it determines optimal instruction complexity, and differential sensitivity to desirable difficulties. These dimensions can be measured and optimized through N=1 experimentation. This survey covers the evidence base for effective study methods, the mechanisms and dimensions of genuine individual variation, and three experiment protocols a reader can run to find their optimal learning approach.

Contents

1. The Learning Styles Myth	2
1.1 What the Framework Claims	2
1.2 The Pashler Review and the Meshing Hypothesis	2
1.3 Why the Myth Persists	3
1.4 What Is Real: Preference vs. Optimal Condition	3
2. What Actually Works: The Evidence Base	4
2.1 Spacing Effect (Distributed Practice)	4
2.2 Testing Effect / Retrieval Practice	5
2.3 Interleaving	5
2.4 Elaborative Interrogation	6
2.5 Desirable Difficulties	6
3. Individual Variation in Learning Effectiveness	7
3.1 Working Memory Capacity	7
3.2 Metacognitive Accuracy	8
3.3 Prior Knowledge and the Expertise Reversal Effect	8
3.4 Interleaving Response Heterogeneity	9

4. Metacognition and Calibration as Learnable Skills	9
5. Platform Implications	10
5.1 Spaced Repetition Systems	10
5.2 Retrieval Practice Integration	11
5.3 Normalizing Desirable Difficulties	11
5.4 Metacognitive Prompts	11
N=1 Experiment Protocols	12
Protocol 1: Study Method Comparison — Retrieval Practice vs. Re-Reading (8 Weeks)	12
Protocol 2: Spacing Interval Optimization (6 Weeks)	13
Protocol 3: Metacognitive Calibration Check (2 Weeks)	14
Individual Variation Summary	15
References	16

1. The Learning Styles Myth

1.1 What the Framework Claims

The VAK (visual-auditory-kinesthetic) model, and its various descendants (VARK, Honey-Mumford, Kolb’s learning styles inventory), make two related claims. The first is a classification claim: people have preferred modalities or styles through which they preferentially take in and process information. The second is the meshing hypothesis: that matching instructional presentation to a learner’s preferred style will produce better learning outcomes than mismatched instruction.

These are empirically distinct claims. The first is weak but not entirely wrong — people do have preferences for how information is presented. The second is where the framework collapses under scrutiny.

1.2 The Pashler Review and the Meshing Hypothesis

Pashler, McDaniel, Rohrer, and Bjork (2008) conducted a systematic review of the learning styles literature specifically to evaluate the meshing hypothesis. The logical structure of a valid test is clear: assign learners to instructional conditions that match or mismatch their assessed style; measure learning outcome; a genuine meshing effect requires an interaction such that visual learners outperform auditory learners under visual instruction, and auditory learners outperform visual learners under auditory instruction. The interaction, not the main effect, is the test.

Pashler et al. found that well-controlled studies meeting this standard consistently fail to find the predicted interaction. Some studies find main effects of instruction type (some content is better conveyed visually), but these apply regardless of the learner’s assessed style. The review’s conclusion is direct: “we have not found evidence that provides unambiguous support for the core claim of the learning-styles approach.”

Subsequent research has not revised this finding. A 2018 study by Rogowsky, Calhoun, and Tallal in the *Journal of Educational Research* replicated the null result. A survey of neuroscientists by Howard-Jones (2014) found that 93% considered learning styles a neuromyth.

1.3 Why the Myth Persists

The persistence of a debunked framework is itself a phenomenon worth understanding. Three mechanisms appear to drive it.

First, the framework *feels true*. People do have preferences for how they encounter new information — some prefer reading, some prefer video, some prefer hands-on exploration. Preferences are real. The error is inferring from a preference that the preferred mode is also the *most effective* mode for retention. These can be and often are different things.

Second, the framework offers identity. “I’m a visual learner” is a self-explanation that converts academic difficulty into a mismatch problem (“the teacher doesn’t present it my way”) rather than a skill or effort problem. This is motivationally appealing and socially safer than the alternatives.

Third, the framework is deeply embedded in teacher training programs. A survey by Dekker et al. (2012) found that the majority of teachers in the UK, Netherlands, Turkey, Greece, and China endorsed learning styles as a valid pedagogical tool, despite the lack of evidence.

1.4 What Is Real: Preference vs. Optimal Condition

The learning styles error is conflating two things: the condition a learner *prefers* and the condition under which the learner *retains information most effectively*. These need not align, and empirically they often do not.

A learner who prefers watching videos may retain information better when they read and

self-test. A learner who dislikes flashcards may still show larger retention gains from retrieval practice than from re-reading. Preferences are about subjective experience; learning outcomes are about objective performance on a retention test conducted after sufficient delay. The goal of studying is the latter, not the former.

This is the key point from which the rest of the survey proceeds: the individual variation that matters for learning is not in modality preference but in dimensions that are measurable, actionable, and partially trainable.

2. What Actually Works: The Evidence Base

The following study methods have the strongest and most replicated evidence base, evaluated across meta-analyses and controlled laboratory and classroom studies. Dunlosky et al. (2013) conducted the most comprehensive review of learning techniques, rating each on utility based on effect size, generalizability, and robustness across content types.

2.1 Spacing Effect (Distributed Practice)

Distributing practice over time — studying material across multiple sessions rather than in a single massed session — produces dramatically better long-term retention. Cepeda et al. (2006) conducted a meta-analysis of 254 studies involving nearly 14,000 participants and found an overall effect size of $d \approx 0.8$ for spaced over massed practice on delayed retention tests.

The mechanism is dual: each reactivation of a memory trace strengthens it, and the interval between study sessions forces greater retrieval effort as the memory fades, producing a stronger trace than effortless re-reading of still-fresh material. Retrieving a fading memory is harder than retrieving a fresh one — and the difficulty produces the benefit.

A practical implication follows: optimal spacing intervals increase as the target retention date is further away. If you need to remember something in one week, reviewing after one day is appropriate. If you need to retain it for a year, review intervals should expand to weeks and then months. This “expanding intervals” principle is the basis for spaced repetition system (SRS) algorithms.

The subjective experience of spaced practice is misleading: the material feels harder to retrieve at a spaced session than at a massed session, which creates the false impression that you are learning less effectively. You are not — you are learning more durably.

2.2 Testing Effect / Retrieval Practice

Roediger and Karpicke (2006) established that retrieving information from memory — the act of testing — is more effective for long-term retention than additional study of the same material. Karpicke and Blunt (2011) compared four conditions in a within-subjects design: study only, repeated study, concept mapping (an elaborative strategy), and retrieval practice. On a one-week delayed retention test, retrieval practice produced significantly better performance than all other conditions, including concept mapping.

The mechanism: retrieval practice strengthens the memory trace in ways that re-exposure to material does not. Each successful retrieval modifies the encoding, making future retrieval more likely and faster. Crucially, retrieval practice also surfaces knowledge gaps — the gaps become visible when the answer cannot be produced, whereas re-reading creates the illusion of knowing (the “fluency illusion”) because recognition is easier than recall.

This effect is robust across content domains (foreign vocabulary, science concepts, procedural knowledge), age groups, and test formats. It is one of the most replicated findings in cognitive psychology. Dunlosky et al. (2013) rate retrieval practice as “high utility” — the strongest category in their review.

2.3 Interleaving

Blocked practice involves completing all problems of one type before moving to the next type. Interleaved practice mixes different problem types together. Despite feeling less productive, interleaving produces better long-term learning and transfer.

Kornell and Bjork (2008) showed this effect for category learning: participants who studied paintings in an interleaved (mixed by artist) schedule performed better on a later artist-attribution test than participants who studied in blocked (all paintings by one artist together) schedules, despite rating the blocked schedule as more effective during study.

The mechanism: blocked practice allows learners to apply a single strategy repeatedly to

similar problems, which produces fast within-session performance but does not require genuine discrimination between problem types. Interleaving forces the learner to identify which type of problem they are facing and select the appropriate strategy on each trial — the kind of discrimination that transfer requires.

The effect has been replicated in mathematics (Rohrer & Taylor, 2007), motor learning (Shea & Morgan, 1979), and music practice. It is particularly robust for learners who rely on surface-feature pattern matching, a common novice strategy (Taylor & Rohrer, 2010).

2.4 Elaborative Interrogation

Asking “why” and “how” questions during study — generating explanations that connect new material to prior knowledge — produces better retention than passive reading or note-taking. The mechanism is that elaboration creates more retrieval pathways: content encoded with multiple associative connections is more likely to be successfully retrieved than isolated facts.

Dunlosky et al. (2013) rate elaborative interrogation as “moderate utility.” The caveat: the benefit depends on having sufficient prior knowledge to generate meaningful explanations. A complete novice cannot elaborate usefully on material they cannot yet contextualize. The technique is most effective once a foundation exists.

2.5 Desirable Difficulties

Robert Bjork (1994) introduced the concept of desirable difficulties: conditions that slow initial learning but produce more durable, transferable knowledge. Testing, spacing, and interleaving are all desirable difficulties — they impair short-term performance on practice tasks while improving long-term retention.

The key insight is that the subjective experience of learning (how fluent and productive a study session feels) is *inversely* correlated with actual retention under desirable difficulty conditions. Studying with flashcards is harder than re-reading highlighted notes. Interleaved practice feels messy and slow. Spaced sessions feel like starting over. In each case, the difficulty is the learning mechanism, not an obstacle to it.

This has a practical implication: learners who use self-reported “productivity” or “how well a session went” as a signal for their study methods will systematically choose the least effective

methods. Objective outcome measurement — quiz scores at delay — is the only valid signal.

3. Individual Variation in Learning Effectiveness

The population-level evidence above establishes that certain study methods work better on average. But individual variation is substantial and matters for practice. The relevant dimensions of variation are not modality preference — they are working memory capacity, metacognitive accuracy, prior knowledge level, and sensitivity to desirable difficulties.

3.1 Working Memory Capacity

Baddeley and Hitch (1974), subsequently revised by Baddeley (1992, 2000), proposed a multi-component working memory model consisting of a central executive, a phonological loop (verbal-acoustic information), and a visuospatial sketchpad (visual and spatial information). Working memory capacity — the amount of information that can be held and manipulated simultaneously — varies approximately three-fold across adults and predicts learning speed, reading comprehension, mathematical problem-solving, and the ability to process complex instruction.

Cognitive load theory (Sweller, 1988) formalizes the implication: instructional materials that exceed a learner’s working memory capacity produce worse learning, not better, because cognitive resources are exhausted by managing information load rather than encoding it. Sweller distinguishes intrinsic cognitive load (complexity inherent to the material), extraneous cognitive load (complexity added by poor instructional design), and germane cognitive load (processing that contributes to schema formation).

The individual implication: the same instructional material imposes different cognitive loads on different learners depending on their working memory capacity and their existing schemas for the domain. A worked example that efficiently teaches a novice produces the expertise reversal effect in an advanced learner who now finds the worked example’s guidance intrusive and who would learn better without it (Kalyuga et al., 2003). Optimal instruction complexity is not universal — it is determined by the individual’s current working memory capacity and knowledge state.

3.2 Metacognitive Accuracy

Metacognition refers to monitoring and control of one’s own cognitive processes — specifically, the ability to judge whether one is learning (monitoring) and to adjust study behavior accordingly (control). Both components show substantial individual variation, and both can be improved.

Kruger and Dunning (1999) documented a systematic failure of metacognitive monitoring: people who perform poorly on tasks also tend to overestimate their performance, because the skills needed to evaluate competence overlap with the skills needed to demonstrate it. This is not limited to low performers: high performers tend to underestimate the difficulty others find in tasks they find easy.

For learning specifically, the failure mode is the fluency illusion: re-reading familiar material feels productive because the words are recognized easily, but recognition is not retrieval. Kornell and Bjork (2009) showed that students consistently prefer less effective study methods (re-reading, highlighting) over more effective ones (self-testing, retrieval practice) precisely because the less effective methods feel more productive. The gap between subjective learning confidence and actual performance on a retention test is individually variable and constitutes a measurable diagnostic.

Calibration can be improved. The most effective intervention is regular testing: producing an answer exposes its absence in a way that recognition cannot. Learners who test themselves regularly develop more accurate models of what they know and do not know, which improves their ability to allocate study time to material that needs it most.

3.3 Prior Knowledge and the Expertise Reversal Effect

Expert-novice differences in learning are among the most robustly established findings in cognitive psychology (Ericsson, Chase, & Faloon, 1980). Experts do not simply know more — they have reorganized knowledge into larger, more integrated schemas that function as single units in working memory. A chess expert “sees” positions as configurations of meaningful patterns; a novice sees individual piece locations.

The instructional implication is the expertise reversal effect (Kalyuga et al., 2003): instructional approaches that work for novices become ineffective or counterproductive for more advanced

learners. Specifically:

- **Novices** benefit from worked examples, explicit guidance, and reduced problem-solving demands — because unguided problem-solving overwhelms working memory before schemas exist to chunk information.
- **Experts** benefit from problem-solving, reduced guidance, and higher-level abstraction — because detailed guidance introduces extraneous load by explaining things the expert already knows.

The practical implication is that optimal instruction is a moving target for each learner. An approach that worked six months ago may be suboptimal now. Routine re-assessment of knowledge level — not just preference assessment once — is necessary for ongoing optimization.

3.4 Interleaving Response Heterogeneity

Interleaving benefits are not uniform across learners. Taylor and Rohrer (2010) found that the interleaving advantage is larger for learners who rely on surface-feature pattern matching — identifying a problem type by superficial cues and applying a memorized procedure — than for learners who already discriminate conceptually between problem types.

Novices tend to match on surface features; they benefit more from interleaving's forcing function. Advanced learners who already discriminate conceptually benefit less from interleaving but still show no harm from it. The effect size of interleaving in a given learner depends on where the learner currently is in their understanding.

This is a case where the population average effect (interleaving works) conceals individual heterogeneity that can be measured. A learner who finds interleaving provides no benefit relative to blocked practice is likely already discriminating conceptually and may need a different challenge — more transfer problems, novel applications, or increased material complexity.

4. Metacognition and Calibration as Learnable Skills

Metacognitive monitoring and control are not fixed traits. They can be trained. The evidence for this is most direct in the retrieval practice literature: learners who regularly test themselves

develop better calibration — their confidence ratings more accurately predict their actual performance — than learners who re-read (Roediger & Karpicke, 2006).

The mechanism is direct experience with the gap. A learner who is confident they know something and then fails to produce the answer on a test receives unambiguous disconfirmation. A learner who re-reads the same material does not encounter this disconfirmation — the material remains familiar, confidence remains high, and the gap is never exposed.

Two additional techniques improve metacognitive calibration:

Prospective confidence ratings: Before a practice test, write a prediction: “I expect to score X% on this quiz.” After the test, compare predicted and actual. Tracking this gap over time reveals systematic biases — consistent overestimation of recall for material that has only been re-read, for example — and provides information for allocating future study time.

The teach-to-learn effect: Explaining material to someone else reveals the limits of one’s own understanding in a way that private review does not. The preparation for explanation forces identification of gaps (the student will notice if an explanation is circular or incomplete), and the act of generating a coherent explanation strengthens encoding through elaboration. Koh, Lee, and Lim (2018) found teaching significantly improved retention in the teacher relative to studying alone.

The N=1 implication is direct: if your confidence-performance gap is large and consistent, you have a metacognitive calibration problem. The fix is not more study time — it is more testing relative to reviewing, and tracking the gap until it closes.

5. Platform Implications

5.1 Spaced Repetition Systems

Spaced repetition systems (SRS) — software implementations of expanding-interval review schedules — are the most direct translation of spacing effect research into a learning tool. The SuperMemo algorithm (Wozniak, 1990) and its open-source descendant used in Anki calculate the optimal next review date for each item based on past retrieval performance. Items retrieved confidently are scheduled further out; items retrieved with difficulty are scheduled sooner.

The evidence for SRS effectiveness is strongest for discrete fact retrieval: vocabulary acquisition, medical terminology, historical dates, and procedural rules. Kornell (2009) showed that SRS outperformed self-paced flashcard study on delayed retention tests. The advantage over conventional study methods for long-term retention is consistent.

5.2 Retrieval Practice Integration

Conventional learning platform design sequences review after content delivery: read the material, then optionally take a quiz. Evidence-aligned design reverses this: quiz before reviewing, to surface gaps, then study to fill them. The pre-test effect (Richland, Kornell, & Kao, 2009) shows that attempting to retrieve information even before seeing it improves subsequent encoding — failure itself is an encoding event when followed by correct information.

Platform prompts can operationalize this: “Before you review today’s material, try to recall the three key points from last session” creates a retrieval attempt before any content is delivered.

5.3 Normalizing Desirable Difficulties

The largest practical obstacle to effective study is the inverse correlation between subjective effort and subjective productivity. Learners who feel a session is not going well will abandon it. Platform design should explicitly communicate that harder retrieval, longer pauses, and reduced fluency are signals of effective practice, not failure.

Progress indicators that track quiz performance over time — rather than session subjective ratings — provide the objective signal that reveals learning is occurring even when it does not feel that way.

5.4 Metacognitive Prompts

Routine metacognitive prompts have small but real effects on calibration. “Before reviewing this material, predict your score on the upcoming quiz” forces a calibration event. “How confident are you that you could explain this to someone with no background?” creates a different demand than “do you feel like you understand this?” — the production demand is more diagnostic than the recognition demand.

N=1 Experiment Protocols

The following protocols are designed to generate individual-level evidence about learning effectiveness. Each follows the standard N=1 structure: single participant, alternating conditions or measurement across time, objective outcome measurement, and a pre-specified decision criterion. They are designed to be run on real learning content the participant already needs to acquire — not artificial laboratory material.

Protocol 1: Study Method Comparison — Retrieval Practice vs. Re-Reading (8 Weeks)

Objective: Determine whether retrieval practice or re-reading produces better retention for your learning content.

Method: Select a domain where you have ongoing learning material — a language, a professional knowledge area, a structured course, or a book sequence. Divide the material into four two-week blocks of roughly equal length and difficulty.

- Weeks 1–2: Study using **re-reading and review** — read the material, make notes, highlight, re-read. Do not quiz yourself.
- Weeks 3–4: Study using **retrieval practice** — after an initial read, close the material and attempt to recall the key points from memory. Use flashcards, written recall, or self-quizzing. Spend approximately the same total time as in weeks 1–2.
- Weeks 5–6: Return to re-reading.
- Weeks 7–8: Return to retrieval practice.

At the end of each two-week block, take a standardized quiz on the material covered in that block. The quiz should be written in advance (before the block begins) so it cannot be tailored to what you happened to review. Use the same quiz format (short answer, not multiple choice) across all blocks. Score each quiz out of 100.

Duration: 8 weeks.

Measurement: Quiz score at end of each two-week block. Compare retrieval practice scores (average of weeks 3–4 and 7–8 quizzes) against re-reading scores (average of weeks 1–2 and 5–6 quizzes).

Decision criterion: If retrieval practice average exceeds re-reading average by 10 or more points, adopt retrieval practice as your primary study method for this content type. If scores are within 10 points, the methods are approximately equivalent for your current material and you can choose on other grounds (e.g., time efficiency).

Note: Most participants will show a retrieval practice advantage. A smaller group — often those with very high prior knowledge in the domain — will show smaller differences because they already retrieve fluently. The experiment will reveal which case applies to you.

Protocol 2: Spacing Interval Optimization (6 Weeks)

Objective: Determine the optimal review interval for new material you want to retain for several months.

Method: Identify a discrete body of material to memorize — foreign language vocabulary, professional definitions, factual content. Divide the items into three groups of equal size and comparable difficulty. Assign each group to one of three spacing conditions:

- **Group A:** Review the next day after initial study.
- **Group B:** Review after 3 days.
- **Group C:** Review after 7 days.

Each group receives exactly one review session under its assigned interval. After the review session, no further study of that group occurs.

At 4 weeks after the review session for each group (i.e., at different calendar dates), take a recall test on that group's items. Record the number of items correctly recalled from memory (not recognition from multiple choice).

Duration: 6 weeks total (initial study in week 1, staggered reviews in weeks 1–2, follow-up tests in weeks 4–6).

Measurement: Percent of items recalled at 4-week follow-up for each spacing condition.

Decision criterion: The spacing interval associated with the highest 4-week recall rate is your optimal spacing interval for this type of material at this retention horizon. Use this interval as the initial review spacing in your spaced repetition system or study schedule.

Note: Population averages suggest the 3–7 day interval outperforms next-day review for 4-week retention. But if you have very high prior familiarity with the material, the pattern may differ. Run the experiment on your actual content rather than assuming the population result applies.

Protocol 3: Metacognitive Calibration Check (2 Weeks)

Objective: Measure your confidence-performance gap to determine whether overconfidence is causing you to under-study material you have not actually retained.

Method: For each study session over 14 consecutive days, before reviewing any material, complete the following:

1. Write a confidence prediction: “I believe I currently know ____% of the material I will be tested on today.”
2. Review the material using your normal method.
3. Take a short practice test (5–10 questions) on the material reviewed. The test should cover the same material you reviewed, but use recall (write the answer) not recognition (select from options).
4. Record your actual score as a percentage.
5. Calculate the gap: Predicted % minus Actual %.

Track the daily gap. At the end of 14 days, compute your average gap and note whether it is consistently positive (overconfidence), consistently negative (underconfidence), or variable.

Duration: 14 days.

Measurement: Daily confidence-performance gap (predicted score minus actual score). Mean gap across all sessions. Standard deviation of the gap (a measure of calibration stability).

Decision criterion:

- **Mean gap > +20%:** You are systematically overconfident. Your subjective sense of learning is a poor guide to actual retention. Shift study time toward topics where you score lowest on tests, not toward topics you feel uncertain about (which are already getting your attention). Increase retrieval practice frequency relative to review time.
- **Mean gap within $\pm 10\%$:** You are well-calibrated. Your time allocation is likely appropriate. Maintain current approach and re-assess quarterly.
- **Mean gap < -10%:** You are systematically underconfident. You likely over-study material you already know. This is a less common pattern; it suggests you can safely reduce time on reviewed material without retention loss.

A high standard deviation (gap varies widely across sessions) suggests that calibration is content-dependent — you are well-calibrated on some material and poorly calibrated on other material. Identify which content types produce the largest overconfidence and prioritize testing over reviewing those specifically.

Individual Variation Summary

The learning styles framework failed because it searched for individual variation in the wrong dimension: modality preference (visual, auditory, kinesthetic) predicts how a learner *wants* to study but not how effectively they retain material under different study conditions. The meshing hypothesis — that learning improves when instruction matches style — has no credible empirical support.

The individual dimensions of variation that actually matter are:

1. **Working memory capacity:** Determines how much complexity can be processed simultaneously and thus the optimal level of instructional difficulty. Not fixed over the lifespan but constrained in the short term. Measure indirectly through performance on novel, complex instruction: if cognitive overload symptoms appear (loss of thread, need to re-read frequently, inability to integrate concepts), the material exceeds current working memory + schema capacity.
2. **Metacognitive accuracy:** The gap between confidence and actual recall performance. Variable across individuals and across content types within individuals. Measurable

through Protocol 3. Improvable through retrieval practice and test-based feedback.

3. **Prior knowledge level:** Determines the effectiveness of guidance-heavy vs. problem-solving-heavy instruction. The expertise reversal effect means that yesterday’s optimal approach may be suboptimal today as knowledge grows. Requires ongoing reassessment rather than a single learning style assessment at one point in time.
4. **Sensitivity to desirable difficulties:** The degree to which a learner benefits from spacing, testing, and interleaving relative to massed review. Variable: novices in a domain show the largest interleaving benefits; learners with higher working memory capacity show more efficient responses to spaced practice. Not directly measurable through self-report — measurable only through outcome-based experiments like Protocol 1.

These dimensions are measurable, partially trainable, and domain-specific. A learner who is well-calibrated in one domain may be overconfident in another. A learner whose optimal spacing interval for language vocabulary is 7 days may retain mathematical procedures better at 3-day intervals. The appropriate response is domain-specific experimentation, not a single “learning style” assessment applied universally.

The population evidence reviewed in this survey — spacing, retrieval practice, interleaving, desirable difficulties — applies broadly. The individual question is not whether these methods work but *how much* each works for you in your current domain at your current knowledge level. The protocols above are designed to answer that question with data, not self-report.

References

- Baddeley, A.D. (1992). Working memory. *Science*, 255(5044), 556–559.
- Baddeley, A.D. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences*, 4(11), 417–423.
- Bjork, R.A. (1994). Memory and metamemory considerations in the training of human beings. In J. Metcalfe & A. Shimamura (Eds.), *Metacognition: Knowing about Knowing* (pp. 185–205). MIT Press.
- Brown, P.C., Roediger, H.L., & McDaniel, M.A. (2014). *Make It Stick*. Harvard University

Press.

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(3), 354–380.

Dekker, S., Lee, N.C., Howard-Jones, P., & Jolles, J. (2012). Neuromyths in education: Prevalence and predictors of misconceptions among teachers. *Frontiers in Psychology*, 3, 429.

Dunlosky, J., Rawson, K.A., Marsh, E.J., Nathan, M.J., & Willingham, D.T. (2013). Improving students' learning with effective learning techniques. *Psychological Science in the Public Interest*, 14(1), 4–58.

Ericsson, K.A., Chase, W.G., & Faloon, S. (1980). Acquisition of a memory skill. *Science*, 208(4448), 1181–1182.

Howard-Jones, P.A. (2014). Neuroscience and education: Myths and messages. *Nature Reviews Neuroscience*, 15(12), 817–824.

Kalyuga, S., Ayres, P., Chandler, P., & Sweller, J. (2003). The expertise reversal effect. *Educational Psychologist*, 38(1), 23–31.

Karpicke, J.D., & Blunt, J.R. (2011). Retrieval practice produces more learning than elaborative studying with concept mapping. *Science*, 331(6018), 772–775.

Koh, A.W.L., Lee, S.C., & Lim, S.W.H. (2018). The learning benefits of teaching: A retrieval practice hypothesis. *Applied Cognitive Psychology*, 32(3), 401–410.

Kornell, N. (2009). Optimising learning using flashcards: Spacing is more effective than cramming. *Applied Cognitive Psychology*, 23(9), 1297–1317.

Kornell, N., & Bjork, R.A. (2008). Learning concepts and categories: Is spacing the “enemy of induction”? *Psychological Science*, 19(9), 901–906.

Kornell, N., & Bjork, R.A. (2009). A stability bias in human memory: Overestimating remembering and underestimating learning. *Journal of Experimental Psychology: General*, 138(4), 449–468.

Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology*, 77(6), 1121–1134.

- Pashler, H., McDaniel, M., Rohrer, D., & Bjork, R. (2008). Learning styles: Concepts and evidence. *Psychological Science in the Public Interest*, 9(3), 103–119.
- Richland, L.E., Kornell, N., & Kao, L.S. (2009). The pretesting effect: Do unsuccessful retrieval attempts enhance learning? *Journal of Experimental Psychology: Applied*, 15(3), 243–257.
- Roediger, H.L., & Karpicke, J.D. (2006). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science*, 1(3), 181–210.
- Rogowsky, B.A., Calhoun, B.M., & Tallal, P. (2018). Matching learning style to instructional method: Effects on comprehension. *Journal of Educational Research*, 111(1), 1–8.
- Rohrer, D., & Taylor, K. (2007). The shuffling of mathematics practice problems improves learning. *Instructional Science*, 35(6), 481–498.
- Shea, J.B., & Morgan, R.L. (1979). Contextual interference effects on the acquisition, retention, and transfer of a motor skill. *Journal of Experimental Psychology: Human Learning and Memory*, 5(2), 179–187.
- Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, 12(2), 257–285.
- Taylor, K., & Rohrer, D. (2010). The effects of interleaved practice. *Applied Cognitive Psychology*, 24(6), 837–848.
- Wozniak, P.A. (1990). Optimization of learning. Master's thesis, University of Technology, Poznan.