Policy Insights from the
Behavioral and Brain Sciences
2024, Vol. 11(1) 59–66

DOI: 10.1177/23727322231218906
journals.sagepub.com/home/bbs
Sage

No Simple Solutions to Complex Problems: Cognitive Science Principles Can Guide but Not Prescribe Educational Decisions

Veronica X. Yan[1], Faria Sana[2], and Paulo F. Carvalho[3]

Abstract

Cognitive science of learning points to solutions for making use of existing study and instruction time more effectively and efficiently. However, solutions are not and cannot be one-size-fits-all. This paper outlines the danger of overreliance on specific strategies as one-size-fits-all recommendations and highlights instead the cognitive learning processes that facilitate meaningful and long-lasting learning. Three of the most commonly recommended strategies from cognitive science provide a starting point; understanding the underlying processes allows us to tailor these recommendations to implement at the right time, in the right way, for the right content, and for the right students. Recommendations regard teacher training, the funding and incentivizing of educational interventions, guidelines for the development of educational technologies, and policies that focus on using existing instructional time more wisely.

Keywords

science of learning, spacing, interleaving, retrieval practice, heterogeneity, educational data

Tweet

One size does not fit all. To empower instructional designers in adapting recommendations to their specific contexts, it is crucial to understand the cognitive processes that underlie learning, rather than getting caught up in ever-changing educational fads.

Key Points

- One size does not fit all: Examples of the most common recommendations from cognitive science of learning—spacing, interleaving, and retrieval—demonstrate how these do not always guarantee more learning.
- Instead of promoting specific strategies, underlying cognitive processes should guide decisions regarding how to sequence learning and when to revisit and practice content.
- Meaningful encoding and retention of knowledge requires attention to essential structures and features, as well as integration with prior knowledge; long-lasting learning requires spaced repetition.
- Action can be taken to incorporate the cognitive science of learning into teacher training and to create guidelines for the development of educational technologies.
- Large-scale educational interventions based on cognitive science-of-learning are more likely to succeed if there are incentives for schools and teachers to participate.
- Solutions to address learning losses should leverage cognitive learning processes to make more efficient use of existing instructional time, rather than adding instructional time.

How should teachers choose which recommendations matter most for their students and how should they tailor these general recommendations to fit their contexts? Every few years produce new educational buzzwords; every school administration touts a new educational paradigm—experiential learning, project-based learning, cooperative learning, flipped classrooms, blended learning, spiral curriculum, standards-based, brain-based, differentiated, student-centered, whole-child, and so on. These buzzwords and paradigms are not in and of themselves bad. Many educational trends are well-intentioned, often supported by empirical evidence (but not always, see the myth of individual learning styles, Furey, 2020). However, even empirically supported practices can fail if one does not understand their underlying

[1]Department of Educational Psychology, The University of Texas at Austin, Austin, TX, USA
[2]Centre for Social Science, Athabasca University, Athabasca, Canada
[3]Human-Computer Interaction Institute, Carnegie Mellon University, Pittsburgh, Pennsylvania, USA

Corresponding Author:
Veronica X. Yan, Department of Educational Psychology, The University of Texas at Austin, 1912 Speedway M/S D5800, Austin, TX 78712, USA.
Email: veronicayan@austin.utexas.edu

processes. General recommendations, by definition, cannot specify exactly how a teacher should implement them for their students in their subjects; selecting and combining them gets complicated (e.g., Koedinger et al., 2013).

Three of the most commonly recommended strategies from the cognitive science of learning are spacing, retrieval, and interleaving. None of these recommendations always guarantees more effective learning. This is not to say that insights from the science of learning are worthless. Rather, the key to understanding what will work for whom and when is to pay attention to underlying cognitive processes. Cognitive science of learning has revealed the processes that are important, and these processes can guide practical decisions regarding strategies to meet students' learning needs.

One Size Does Not Fit All: Example Recommendations from the Literature

Recommendation 1: Interleaved Practice

The interleaving effect is the finding that studying and practicing examples of different concepts in a mixed order can facilitate learning more than focusing on one concept at a time (Kang, 2016). This effect is robust across many types of to-be-learned content, ranging from visual categories (e.g., Yan & Sana, 2021a) to STEM concepts (e.g., Rohrer, 2012; Sana & Yan, 2022), and sports and motor skills (e.g., Hall et al., 1994). A recent meta-analysis by Brunmair and Richter (2019) found that interleaved practice, compared to blocked practice, improves performance on a final test by close to half a standard deviation (an average effect size of Hedges' $g = 0.42$, which is roughly equivalent to moving a student from the 50th percentile to the 66th percentile).

The same analyses also found that much of the variance in this meta-analysis could be attributed to heterogeneity—variability in the findings—instead of sampling error ($I^2 = 77\%$). In many instances, the interleaving effect is eliminated or even reversed. For example, blocking advantages occur when learning concepts are easy to tell apart but hard to integrate (e.g., classic physics versus quantum physics; Carvalho & Goldstone, 2014; Yan & Sana, 2021b), when verbalizable rules of a category are difficult to discover (Carvalho & Goldstone, 2015; Noh et al., 2016; Yan & Schuetze, 2022), when the test does not require knowing small differences (Carvalho & Goldstone, 2021), and when skills are complex and require attention to multiple components (e.g., learning different tennis or golf strokes; Guadagnoli et al., 1999; Hebert et al., 1996).

Recommendation 2: Spaced or Distributed Practice

Sustained learning requires that content must be revisited, at intervals. The *spacing effect* (also distributed practice or lag effect) is the finding that learning is better sustained when repetitions are separated (e.g., Dunlosky et al., 2013; Janiszewski et al., 2008). The benefit of spacing spans different learning materials (Carpenter et al., 2022) and various time scales (e.g., across minutes, days, and weeks; Cepeda et al., 2006). For example, students who were assigned to review a lecture 8 days after the lecture performed better on a test of both factual and applied learning 5 weeks later compared to students who reviewed the lecture only one day later. A recent meta-analysis (Latimier et al., 2021) found that spaced practice, compared to massed practice, improves performance on a final test by close to over half a standard deviation (an average effect size of Hedges' $g = 0.74$, which is roughly equivalent to moving a student from the 50th percentile to the 77th percentile). Similarly, an earlier meta-analysis (Janiszewski et al., 2003) found a moderate positive correlation between the length of spacing and performance on a final retention test, $r = .34$.

Importantly, a considerable amount of the variance in the more recent meta-analysis could be attributed to heterogeneity, or variability in the findings, instead of sampling error ($I^2 = 49\%$). Indeed, more spacing is not always better. For example, when repetitions are varied, the optimal spacing between repetitions is lower than when repetitions are exact (Appleton-Knapp et al., 2005). Optimal spacing depends on how far away the final test is from the last repetition and how far apart the repetitions are spread during study (e.g., Cepeda et al., 2008, 2009; Khajah et al., 2014; Lindsey et al., 2014; Toppino et al., 2018).

Recommendation 3: Retrieval Practice

Compared to repeated restudy, *retrieval practice* in the form of practice tests such as low-stakes testing improves later performance (e.g., Carpenter et al., 2022). This is a robust phenomenon that generalizes to different tasks (e.g., Dunlosky et al., 2013). A recent meta-analytic review of classroom studies (Yang et al., 2021) showed retrieval practice in the form of quizzing enhanced learning by half a standard deviation (Hedges' $g = 0.50$). A meta-analysis of mostly laboratory studies (Rowland, 2014) found a similar effect size (Hedges' $g = 0.50$).

Both meta-analyses also found that a considerable amount of the variance in this meta-analysis could be attributed to heterogeneity, or variability in the findings instead of sampling error ($I^2 = 88\%$ and $I^2 = 84\%$, respectively). Indeed, retrieval practice is unlikely to work for everyone all the time. The literature includes what may appear to be conflicting evidence. For example, worked examples, a form of restudy, can improve learning compared to further practice trials (e.g., van Gog & Sweller, 2015; Yeo & Fazio, 2019), especially when learning skills such as calculating the area of a geometrical shape compared to memorizing facts (Carvalho et al., 2022; Rachatasumrit et al., 2023). Similarly, when learners are not able to successfully retrieve the correct response, and when no feedback is provided, further retrieval practice has little, or even, a negative effect (Fiechter & Benjamin, 2018; Rowland, 2014).

A Better Starting Point: Cognitive Processes Rather than Learning Strategies

Solutions are unlikely to be simple or easy recipes to follow. Rather than focusing on learning strategies, the focus should be on cognitive learning *processes*.

Process 1: Attention to Essential Structures and Features

Why would interleaving be generally an effective learning strategy but blocking be more effective in some situations? One conceptualization is that the sequence of practice changes what learners attend to, encode, and remember. That is, what a learner will learn from an interleaved sequence differs from what a learner would learn from a blocked sequence of practice. When examples from different concepts are juxtaposed (i.e., interleaving), attention is drawn to the critical features that distinguish between them; when examples from the same concept are juxtaposed, attention is drawn to the critical features (i.e., blocking) that tie a concept together (Brunmair & Richter, 2019; Carvalho & Goldstone, 2017; Yan & Schuetze, 2022). This general proposal is consistent with the findings described before that interleaved practice improves learning of hard-to-discriminate concepts, but blocked practice improves learning of easy-to-discriminate concepts. Thus, there are no good or bad sequences but instead whether the sequence emphasizes what needs to be learned. If the learning sequence draws attention to aspects of the concepts that are challenging to learn, learning is improved. However, if the learning sequence does not draw attention to critical aspects that are hard or challenging to learn, learning is not improved.

Process 2: Spaced Repetition

Why would greater spacing be more effective in some situations, but less effective in others? One conceptualization is that when repetitions are spaced apart, revisiting the content serves as a reminder of the previous repetition, increasing the likelihood of retrieving the associated details (Benjamin & Tullis, 2010). Essentially, each repetition event activates reminding and subsequent retrieval processes of the initial or previous repetition, which in turn strengthens memory. Thus, introducing some amount of spacing between repetitions improves learning outcomes by increasing forgetting, and in turn, reducing the salience of the content enough for reminding and retrieval processes to engage (Bjork & Bjork, 2006). However, too much spacing between repetitions may not improve retention if the learners fail to make the connection between the new repetition and the previous one, particularly if too much time has passed, and the new repetition appears significantly different (Appleton-Knapp et al., 2005). In essence, content should be revisited when retrieval of prior learning is challenging, but still possible —not too soon, yet not too late. In addition to repetition salience and concept difficulty, the optimal spacing will depend on how long the retention interval is between the last presentation and the final test (Cepeda et al., 2008): Longer retention intervals, on average, require longer spacing and more repetitions because the memory needs to be strengthened enough to last for a longer period of time.

Process 3: Successful, Effortful Retrieval from Long-Term Memory

Why would quizzing be more effective than restudying or worked examples in some situations, but not in others? Quizzing works when it engages retrieval; retrieval practice improves learning by slowing forgetting (Bjork, 1975; Rachatasumrit et al., 2023). For this to take place one needs to have initially encoded the correct information; thus retrieval practice is particularly effective when it elicits successful, effortful retrievals (Pyc & Rawson, 2009). Similarly, if the learning task requires one to integrate information before memorization, retrieval practice will not be as effective as extended studying (Carvalho et al., 2022; Yeo & Fazio, 2019). Moreover, retrieval of knowledge should be practiced at the right level of learning. For example, students who practice recalling concept-term definitions will strengthen their memories of the definitions, but this type of practice will not help them when a final test consists of higher level application questions (Agarwal, 2019). Finally, if retrieval is too easy—perhaps because the learner *just* studied the answers, perhaps because prior knowledge means that the content itself is too easy for the learner—then there will also be little benefit to quizzing (Minear et al., 2018; Pyc & Rawson, 2009).

Example Decision Process and Considerations for Implementation

The benefits of different cognitive interventions are heterogeneous and do not guarantee better outcomes for everyone all the time. This next section provides an example of how an instructional designer might decisions based the cognitive processes described in the previous section and data they can collect about which of the three recommendations to use and when (see the emerging Learning Engineering discipline; Barrett et al., 2022). These are intended as examples of how an instructional designer may make decisions about sequencing, spacing, and retrieval; these do not present an exhaustive list of influencing factors or an exhaustive list of decision options.

Consider the Students

What prior knowledge and skills do the students bring in with them? One way to understand what prior knowledge students already hold is to have initial low-stakes assessments. Understanding how much students already know allows tailoring instruction for students with low and high knowledge. If students are still attempting to identify critical features relevant to a topic, then worked examples emphasizing deep structural features are more beneficial for their learning than retrieval practice. Engaging in problem-solving and retrieval practice may not be effective when foundational understanding is lacking, and nothing relevant can be retrieved (Rachatasumrit et al., 2023). On the other hand, if students have some relevant prior knowledge, then appropriately designed problems that encourage them to draw upon that knowledge can be useful (Kapur & Bielaczyc, 2012). Moreover, if students have already had some prior experience with the to-be-learned concepts, then they will benefit more from activities that allow them to retrieve that knowledge and practice applying it to problems.

Consider the Content

What if students are learning different types of content? Low-stakes formative assessments and class activities can show which materials students are having difficulty grasping as a coherent concept and which they are having difficulty telling apart. For example, when learning about feedback loops, students will often struggle to grasp how examples from various systems can be classified as positive feedback loops. For example, fruit ripening, childbirth, and blood-clotting are all examples of positive feedback loops even though they look very different on the surface (Goldstone et al., 2010). This difficulty may be evident in formative assessments, where students have difficulty consistently identifying positive feedback loops, often making systematic errors in their identification. In this situation, assessments or activities that encourage blocked practice will result in better learning of feedback loops.

Are there sets of concepts that students tend to confuse? For example, students might mix fraction addition and fraction multiplication, systematically applying the procedural rules for fraction addition to fraction multiplication or vice versa. This will show up in formative assessments as predictable errors. In this situation, activities that encourage the interleaved practice of fraction addition and fraction multiplication can help learners understand the difference and learn which procedure is relevant to which problem type (Patel et al., 2016). Understanding what *aspects* of the content students finding confusing is also important. For example, fraction problems can also be graphically represented in multiple ways (e.g., circle diagrams, number lines, sets); but interleaving graphical representations (which are easy to distinguish) is not as effective as interleaving task types (Rau et al., 2013).

Depending on the content, it may make sense for a learning sequence to include some amount of blocking to support initial competency for each concept and some amount of interleaving to support differentiating between similar concepts (Mielicki & Wiley, 2022; Porter & Saemi, 2010; Yan et al., 2017). The optimal time to move from blocked to interleaved practice, however, depends on what each learner has been able to understand.

Consider the Practice Opportunities for Sustained Learning

For those concepts that need to be maintained, what are the opportunities students have for practicing them and what do they look like? Consider what is the broader goal of learning —memorization and fluency in foundational skills or facts or higher level application of knowledge? Both types of learning are important—having fluency in foundational facts and skills (e.g., historical events and dates) frees up learners' working memories so that they can focus on more complex tasks; meaningful learning requires that learners can also apply knowledge and skills to new scenarios (e.g., understanding the broader social and contextual factors influencing events). The learning goal and the learner's stage will determine the optimal type of retrieval practice opportunity to provide: The practice opportunities might be relatively simple at first to support memorization, but then move toward more complex and varied activities requiring learners to apply their knowledge to different scenarios. Note that practice opportunities should be no- or low-stakes: Punishing students for not being perfect on these tests directly undermines the principle that students should be given practice opportunities that require effortful engagement (and hence, imperfect performance).

When should retrieval practice opportunities be provided? Consider how quickly the content is forgotten—this is related to both how well-learned the content is in the first place and how much retrieval practice with the material they have already had. Retrieval practice to sustain learning is most effective when learners have to engage in some effort to recall the content. Hence, the first retrieval practice opportunity may occur relatively close to initial instruction to provide more supports (e.g., hints or scaffolding) so there is a high rate of retrieval success; subsequent retrieval practice opportunities can be pushed farther apart in time to ensure that retrieval continues to be effortful and not trivial or with a fading-out of supports (Eglington & Pavlik, 2020; Fiechter & Benjamin, 2018; Storm et al., 2010).

Tailoring of Guidance Relies on Teacher Expertise

In education research, heterogeneity is often viewed through the lens of demographic factors, such as race/ethnicity, age, sex, socio-economic status, and geographic locale (Schudde, 2018; Tipton et al., 2019). However, these demographic factors are unlikely to be causal factors, instead reflecting intermediary factors in the process (e.g., math fluency is one of the background knowledge pieces that allow students to succeed at math in early years, but girls are less likely to be exposed to math concepts as children; Byrge et al., 2014). Identifying these demographic gaps can show where and for whom intervention is most needed, but it does not tell *how* to best intervene. As demonstrated with the examples in the previous sections, the answers to the *how* resides in understanding the cognitive processes and which practices will most likely engage those processes in a given context.

Research from cognitive psychology can only point to general practices that might be likely to activate these cognitive processes; it is the teachers' expertise—their knowledge of their content and of their students—that is required to appropriately tailor their instructional strategies to the unique needs of their students, moving away from one-size-fits-all pedagogies. In doing so, they can avoid the pitfalls of implementing strategies ineffectively. This approach to attending to heterogeneity not only elevates the quality of education but also fosters a deeper understanding of the underlying cognitive processes of instructional strategies.

Opportunities for Policy from Cognitive Science

Learning is complex and effective recommendations therefore cannot be one-size-fits-all. Several opportunities for policy follow from building on cognitive principles.

Incorporate Cognitive Principles into Teacher Training and Certification

Teacher training and teacher certification should include understanding the underlying cognitive processes of learning, and not just knowing about particular pedagogical techniques. Educational buzzwords will come and go; training should equip teachers to see through marketing language and to evaluate and tailor pedagogies based on the cognitive processes they engage. On the other side of the coin, misconceptions about learning should also be removed from teacher training and teacher certification: In 29 US states, government-distributed test-preparation materials on teacher certification exams include the debunked theory of "learning styles" (Furey, 2020; for more, see Robinson et al., 2022).

Fund and Incentivize Educational Interventions Based on the Science of Learning

Investing in educational interventions based on the cognitive science of learning is important. In a recent review of 747 randomized control trials of educational interventions in K-12 classrooms, Kraft (2020) found an average effect size of $g = 0.16$. The effect size at the 90th percentile was $g = 0.47$. In contrast, meta-analyses of interleaving, spacing, and retrieval practice ranged from $g = 0.42$ to 0.74. Interventions based on cognitive science-of-learning are therefore promising. However, these numbers are not directly comparable, as the majority of the cognitive science-of-learning studies are not carried out in actual classroom settings (although those conducted in classrooms often yield similar effect sizes). When studies are conducted in classrooms, the outcomes are often assessed using narrow experimenter-created tests, rather than broad teacher-created or standardized tests.

Currently, conducting cognitive science-of-learning educational interventions in classrooms relies on researchers having good relationships with individual teachers (e.g., Agarwal et al., 2012), and conducting these studies *at scale,* especially in a way that is sensitive to underlying cognitive processes, is near impossible (but see Fyfe et al., 2021). Funding of cognitive science-of-learning educational interventions at scale should be a key priority. Moreover, the focus should extend beyond mere generalizability to actively building theory for a strong foundation for understanding how interventions can be tailored to diverse contexts, content, and student needs.

Funding the researchers alone is not enough; teachers and schools should also be funded in ways that incentivize and foster research-practice partnerships. Structural incentives may take many forms, such as reducing individual teachers' teaching time to focus on research or rewarding schools for demonstrating evidence-based, data-driven improvements. Policymakers should consider the ways of incentivizing research-practitioner collaborations focused on the cognitive science of learning; this should lead to more impactful education research, more impactful, evidence-informed instructional practices, and better educational outcomes.

Provide Guidelines for the Development of Educational Technologies

Technology provides a unique opportunity to create engaging instruction that harnesses the importance of adapting approaches to different content, students, and situations. The same group of students could have personalized and very different experiences in a learning technology environment (e.g., Koedinger et al., 1997; Lindsey et al., 2014). However, this adaptability should include the whole learning situation when making instructional decisions, and not only the student's performance in the current task (Schuetze & Yan, 2023; Soderstrom & Bjork, 2015). For example, it is important to take into consideration when the information learned will have to be used to decide how much spacing to include, how well the student already understands the concepts, and pivot to interleaving if concepts are being confused. This is still an unfulfilled promise of educational technology, but an important next frontier (Schuetze, 2023).

Using Existing Time Efficiently to Address Learning Gaps and Learning Loss

Achievement gaps are a long-standing issue: Students come into schools with varying background preparation; schools and districts themselves are subject to differing levels of resources. Many students are behind where they should or could be. This issue has only been exacerbated following the COVID-19 pandemic (Kuhfeld et al., 2022). Most of the proposed solutions—before-school and after-school remediation, tutoring programs, summer school, extended school days—all take one common form: more instructional time. But more time cannot always be the answer, not from the students nor from the teachers. Instead, proposed solutions should also incorporate principles from the cognitive science of learning to make use of existing study and instruction time more effectively and efficiently.

Declaration of Conflicting Interests

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

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The preparation of this manuscript was supported in part by National Science Foundation award #2238567 to VXY, by Social Sciences and Humanities Research Council award # 435-2021-0426 to FS and VXY, and by National Science Foundation award #2301130 to PFC. Social Sciences and Humanities Research Council of Canada, National Science Foundation, (grant number 435-2021-0426, 2238567 , 2301130).

ORCID iD

Veronica X. Yan https://orcid.org/0000-0002-3988-3184

References

Agarwal, P. K. (2019). Retrieval practice & Bloom's taxonomy: Do students need fact knowledge before higher order learning? *Journal of Educational Psychology*, *111*(2), 189–209. https://doi.org/10.1037/edu0000282

Agarwal, P. K., Bain, P. M., & Chamberlain, R. W. (2012). The value of applied research: Retrieval practice improves classroom learning and recommendations from a teacher, a principal, and a scientist. *Educational Psychology Review*, *24*(3), 437–448. https://doi.org/10.1007/s10648-012-9210-2

Appleton-Knapp, S. L., Bjork, R. A., & Wickens, T. D. (2005). Examining the spacing effect in advertising: Encoding variability, retrieval processes, and their interaction. *Journal of Consumer Research*, *32*(2), 266–276. https://doi.org/10.1086/432236

Barrett, M., Czerwinski, E., Goodell, J., Jacobs, D., Ritter, S., Sottilare, R., & Thai, K. P. (2022). *Learning Engineering Uses Data (Part 2): Analytics*. Routledge.

Benjamin, A. S., & Tullis, J. (2010). What makes distributed practice effective? *Cognitive Psychology*, *61*(3), 228–247. https://doi.org/10.1016/j.cogpsych.2010.05.004

Bjork, R. A. (1975). Retrieval as a memory modifier: An interpretation of negative recency and related phenomena. In R. L. Solso (Ed.), *Information processing and cognition: The loyola symposium* (pp. 123–144). Lawrence Erlbaum.

Bjork, R. A., & Bjork, E. L. (2006). Optimizing treatment and instruction: Implications of a new theory of disuse. In L. G. Nilsson, & N. Ohta (Eds.), *Memory and society: Psychological perspectives* (pp. 116–140). Psychology Press.

Brunmair, M., & Richter, T. (2019). Similarity matters: A meta-analysis of interleaved learning and its moderators. *Psychological Bulletin*, *145*(11), 1029–1052. https://doi.org/10.1037/bul0000209

Byrge, L., Smith, L. B., & Mix, K. S. (2014). Beginnings of place value: How preschoolers write three-digit numbers. *Child Development*, *85*(2), 437–443. https://doi.org/10.1111/cdev.12162

Carpenter, S. K., Pan, S. C., & Butler, A. C. (2022). The science of effective learning with spacing and retrieval practice. *Nature Reviews Psychology. Nature Reviews Psychology*, *1*, 496–511. https://doi.org/10.1038/s44159-022-00089-1

Carvalho, P., Rachatasumrit, N., & Koedinger, K. R. (2022). Learning depends on knowledge: The benefits of retrieval practice vary for facts and skills. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *44*(44), 10–175. https://escholarship.org/uc/item/5cm7582q

Carvalho, P. F., & Goldstone, R. L. (2014). Putting category learning in order: Category structure and temporal arrangement affect the benefit of interleaved over blocked study. *Memory & Cognition*, *42*(3), 481–495. https://doi.org/10.3758/s13421-013-0371-0

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*(1), 281–288. https://doi.org/10.3758/s13423-014-0676-4

Carvalho, P. F., & Goldstone, R. L. (2017). The sequence of study changes what information is attended to, encoded, and remembered during category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*(11), 1699–1719. https://doi.org/10.1037/xlm0000406

Carvalho, P. F., & Goldstone, R. L. (2021). The most efficient sequence of study depends on the type of test. *Applied Cognitive Psychology*, *35*(1), 82–97. https://doi.org/10.1002/acp.3740

Cepeda, N. J., Coburn, N., Rohrer, D., Wixted, J. T., Mozer, M. C., & Pashler, H. (2009). Optimizing distributed practice: Theoretical analysis and practical implications. *Experimental Psychology*, *56*(4), 236–246. https://doi.org/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*(3), 354–380. https://doi.org/10.1037/0033-2909.132.3.354

Cepeda, N. J., Vul, E., Rohrer, D., Wixted, J. T., & Pashler, H. (2008). Spacing effects in learning: A temporal ridgeline of optimal retention. *Psychological Science*, *19*(11), 1095–1102. https://doi.org/10.1111/j.1467-9280.2008.02209.x

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*(1), 4–58. https://doi.org/10.1177/1529100612453266

Eglington, L. G., & Pavlik, P. I. Jr (2020). Optimizing practice scheduling requires quantitative tracking of individual item performance. *Npj Science of Learning*, *5*(1), 1–10. https://doi.org/10.1038/s41539-020-00074-4

Fiechter, J. L., & Benjamin, A. S. (2018). Diminishing-cues retrieval practice: A memory-enhancing technique that works when regular testing doesn't. *Psychonomic Bulletin & Review*, *25*(5), 1868–1876. https://doi.org/10.3758/s13423-017-1366-9

Furey, W. (2020). The Stubborn Myth of "Learning Styles." *Education Next*. Retreived from https://www.educationnext.org/stubborn-myth-learning-styles-state-teacher-license-prep-materials-debunked-theory/

Fyfe, E. R., de Leeuw, J. R., Carvalho, P. F., Goldstone, R. L., Sherman, J., Admiraal, D., Alford, L. K., Bonner, A., Brassil, C. E., Brooks, C. A., Carbonetto, T., Chang, S. H., Cruz, L., Czymoniewicz-Klippel, M., Daniel, F., Driessen, M., Habashy, N., Hanson-Bradley, C. L., Hirt, E. R., & Motz, B. A. (2021). Manyclasses 1: Assessing the generalizable effect of immediate feedback versus delayed feedback across many college classes. *Advances in Methods and Practices in Psychological Science*, *4*(3), 25152459211027576. https://doi.org/10.1177/25152459211027575

Goldstone, R. L., Landy, D. H., & Son, J. Y. (2010). The education of perception. *Topics in Cognitive Science*, *2*(2), 265–284. https://doi.org/10.1111/j.1756-8765.2009.01055.x

Guadagnoli, M. A., Holcomb, W. R., & Weber, T. J. (1999). The relationship between contextual interference effects and performer expertise on the learning of a putting task. *Journal of Human Movement Studies*, *37*(1), 19–36.

Hall, K. G., Domingues, D. A., & Cavazos, R. (1994). Contextual interference effects with skilled baseball players. *Perceptual and Motor Skills*, *78*(3 Pt 1), 835–841. https://doi.org/10.2466/pms.1994.78.3.835

Hebert, E. P., Landin, D., & Solmon, M. A. (1996). Practice schedule effects on the performance and learning of low- and high-skilled students: An applied study. *Research Quarterly for Exercise and Sport*, *67*(1), 52–58. https://doi.org/10.1080/02701367.1996.10607925

Janiszewski, C., Noel, H., & Sawyer, A. G. (2003). Meta-analysis of the spacing effect in verbal learning: Implications for research on advertising repetition and consumer memory | journal of consumer research | Oxford academic. *Journal of Consumer Research*, *30*(1), 138–149. https://doi.org/10.1086/374692

Kang, S. H. K. (2016). The benefits of interleaved practice for learning. In J. Horvath, J. Lodge, & J. Hattie (Eds.), *From the laboratory to the classroom* (pp. 91–105). Routledge. https://doi.org/10.4324/9781315625737-13

Kapur, M., & Bielaczyc, K. (2012). Designing for productive failure. *Journal of the Learning Sciences*, *21*(1), 45–83. https://doi.org/10.1080/10508406.2011.591717

Khajah, M. M., Lindsey, R. V., & Mozer, M. C. (2014). Maximizing Students' retention via spaced review: Practical guidance from computational models of memory. *Topics in Cognitive Science*, *6*(1), 157–169. https://doi.org/10.1111/tops.12077

Koedinger, K. R., Anderson, J. R., Hadley, W. H., & Mark, M. A. (1997). Intelligent tutoring goes to school in the big city. *International Journal of Artificial Intelligence in Education*, *8*, 30–43.

Koedinger, K. R., Booth, J. L., & Klahr, D. (2013). Instructional complexity and the science to constrain it. *Science*, *342*(6161), 935–937. https://doi.org/10.1126/science.1238056

Kraft, M. A. (2020). Interpreting effect sizes of education interventions. *Educational Researcher*, *49*(4), 241–253. https://doi.org/10.3102/0013189X20912798

Kuhfeld, M., Soland, J., Lewis, K., & Morton, E. (2022). *The pandemic has had devastating impacts on learning. What will it take to help students catch up?* Brookings. Retreived from https://www.brookings.edu/articles/the-pandemic-has-had-devastating-impacts-on-learning-what-will-it-take-to-help-students-catch-up/

Latimier, A., Peyre, H., & Ramus, F. (2021). A meta-analytic review of the benefit of spacing out retrieval practice episodes on retention. *Educational Psychology Review*, *33*(3), 959–987. https://doi.org/10.1007/s10648-020-09572-8

Lindsey, R. V., Shroyer, J. D., Pashler, H., & Mozer, M. C. (2014). Improving Students' long-term knowledge retention through personalized review. *Psychological Science*, *25*(3), 639–647. https://doi.org/10.1177/0956797613504302

Mielicki, M. K., & Wiley, J. (2022). Exploring the necessary conditions for observing interleaved practice benefits in math learning. *Learning and Instruction*, *80*, 101583. https://doi.org/10.1016/j.learninstruc.2022.101583

Minear, M., Coane, J. H., Boland, S. C., Cooney, L. H., & Albat, M. (2018). The benefits of retrieval practice depend on item difficulty and intelligence. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *44*(9), 1474–1486. https://doi.org/10.1037/xlm0000486

Noh, S. M., Yan, V. X., Bjork, R. A., & Maddox, W. T. (2016). Optimal sequencing during category learning: Testing a dual-learning systems perspective. *Cognition*, *155*, 23–29. https://doi.org/10.1016/j.cognition.2016.06.007

Patel, R., Liu, R., & Koedinger, K. R. (2016). When to Block versus Interleave Practice? Evidence Against Teaching Fraction Addition before Fraction Multiplication. *Proceedings of the 38th Annual Meeting of the Cognitive Science Society, Recogbizing and Representing Events, CogSci 2016, Philadelphia, PA, USA, August 10-13,* 2016.

Porter, J., & Saemi, E. (2010). Moderately skilled learners benefit by practicing with systematic increases in contextual interference. *International Journal of Coaching Science*, *4*(2), 61–71.

Pyc, M. A., & Rawson, K. A. (2009). Testing the retrieval effort hypothesis: Does greater difficulty correctly recalling information lead to higher levels of memory? *Journal of Memory and Language*, *60*(4), 437–447. https://doi.org/10.1016/j.jml.2009.01.004

Rachatasumrit, N., Carvalho, P. F., Li, S., & Koedinger, K. R. (2023). Content matters: A computational investigation into the effectiveness of retrieval practice and worked examples. In N. Wang, G. Rebolledo-Mendez, N. Matsuda, & O. C. Santos,, & V. Dimitrova (Eds.), *Artificial intelligence in education* (pp. 54–65). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-36272-9_5

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. https://doi.org/10.1016/j.learninstruc.2012.07.003

Robinson, D. H., Yan, V. X., & Kim, J. A. (2022). *Monographs in the Psychology of Education: Learning Styles, Classroom Instruction, and Student Achievement*. Springer.

Rohrer, D. (2012). Interleaving helps students distinguish among similar concepts. *Educational Psychology Review*, *24*(3), 355–367. https://doi.org/10.1007/s10648-012-9201-3

Rowland, C. A. (2014). The effect of testing versus restudy on retention: A meta-analytic review of the testing effect. *Psychological Bulletin*, *140*(6), 1432–1463. https://doi.org/10.1037/a0037559

Sana, F., & Yan, V. X. (2022). Interleaving retrieval practice promotes science learning. *Psychological Science*, *33*(5), 782–788. https://doi.org/10.1177/09567976211057507

Schudde, L. (2018). Heterogeneous effects in education: The promise and challenge of incorporating intersectionality into quantitative methodological approaches. *Review of Research in Education*, *42*(1), 72–92. https://doi.org/10.3102/0091732X18759040

Schuetze, B. A. (2023). A computational model of school achievement. *PsyArXiv*. Retreived from https://doi.org/10.31234/osf.io/be6g4.

Schuetze, B. A., & Yan, V. X. (2023). The complexity of inferring long-term, generalizable learning from immediate behavioural indicators (open peer commentary 9). *Journal of Learning Analytics*, *10*(2), 14–50. https://doi.org/10.18608/jla.2023.8197

Soderstrom, N. C., & Bjork, R. A. (2015). Learning versus performance: An integrative review. *Perspectives on Psychological Science*, *10*(2), 176–199. https://doi.org/10.1177/1745691615569000

Storm, B. C., Bjork, R. A., & Storm, J. C. (2010). Optimizing retrieval as a learning event: When and why expanding retrieval practice enhances long-term retention. *Memory & Cognition*, *38*(2), 244–253. https://doi.org/10.3758/MC.38.2.244

Tipton, E., Yeager, D. S., Iachan, R., & Schneider, B. (2019). Designing probability samples to study treatment effect heterogeneity. In *Experimental methods in survey research* (pp. 435–456). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781119083771.ch22

Toppino, T. C., Phelan, H. A., & Gerbier, E. (2018). Level of initial training moderates the effects of distributing practice over multiple days with expanding, contracting, and uniform schedules: Evidence for study-phase retrieval. *Memory & Cognition*, *46*(6), 969–978. https://doi.org/10.3758/s13421-018-0815-7

van Gog, T., & Sweller, J. (2015). Not new, but nearly forgotten: The testing effect decreases or even disappears as the complexity of learning materials increases. *Educational Psychology Review*, *27*(2), 247–264. https://doi.org/10.1007/s10648-015-9310-x

Yan, V. X., & Sana, F. (2021b). Does the interleaving effect extend to unrelated concepts? Learners' beliefs versus empirical evidence. *Journal of Educational Psychology*, *113*(1), 125–137. https://doi.org/10.1037/edu0000470

Yan, V. X., & Sana, F. (2021a). The robustness of the interleaving effect. *Journal of Applied Research in Memory and Cognition*, *10*(4), 589–602. https://doi.org/10.1016/j.jarmac.2021.05.002

Yan, V. X., & Schuetze, B. A. (2022). Not just stimuli structure: Sequencing effects in category learning vary by task demands. *Journal of Applied Research in Memory and Cognition*, *11*(2), 218–228. https://doi.org/10.1016/j.jarmac.2021.09.004

Yan, V. X., Soderstrom, N. C., Seneviratna, G. S., Bjork, E. L., & Bjork, R. A. (2017). How should exemplars be sequenced in inductive learning? Empirical evidence versus learners' opinions. *Journal of Experimental Psychology: Applied*, *23*(4), 403–416. https://doi.org/10.1037/xap0000139

Yang, C., Luo, L., Vadillo, M. A., Yu, R., & Shanks, D. R. (2021). Testing (quizzing) boosts classroom learning: A systematic and meta-analytic review. *Psychological Bulletin*, *147*(4), 399–435. https://doi.org/10.1037/bul0000309

Yeo, D. J., & Fazio, L. K. (2019). The optimal learning strategy depends on learning goals and processes: Retrieval practice versus worked examples. *Journal of Educational Psychology*, *111*(1), 73–90. https://doi.org/10.1037/edu0000268