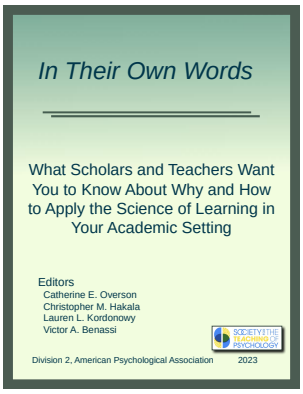


# Interleaved Training and Category Learning

***Sean H. K. Kang***
The University of Melbourne

## Overview

The ability to categorize is fundamental to human existence. Learning to assign objects in our environment to separate categories allows us to recognize the things around us, to use language (e.g., common nouns refer to classes of objects rather than a specific object), and to respond appropriately to different organisms (e.g., edible vs. poisonous plants). Category learning is also integral to achievement across academic disciplines, whether in the arts and humanities (e.g., learning to identify a style of painting or music; learning grammatical categories and their associated rules), or in STEM (e.g., learning taxonomic classification in biology; learning to identify the kind of problem one is trying to solve so that the right strategies might be employed).

In many real-world situations, and especially in education, it is more useful to identify patterns and principles from past experiences and be able to apply this learning to new circumstances than to be able to remember specific past examples, simply because it is less likely that one would face the exact same example again in the future. Inductive learning of relevant previous encounters allows the formation of mental categories, and this category learning then equips the learner to go beyond the prior cases and categorize novel examples as being a member of the category (or not). In other words, the learning of categories is an important way by which learners become able to generalize or transfer their learning to new instances.

Cognitive scientists over the past decade have shown renewed interest in how category learning might ibe improved or facilitated, and the sequencing of examples during training has been a particular area of focus. Specifically, researchers have examined whether grouping together examples from the same category during study (e.g., $A_1\,A_2\,A_3\,A_4\,B_1\,B_2\,B_3\,B_4\,C_1\,C_2\,C_3\,C_4$), otherwise known as *blocked* training, or mixing examples from different categories (e.g., $A_1\,B_1\,C_1\,B_2\,A_2\,C_2\,A_3\,C_3\,B_3\,C_4\,A_4\,B_4$), also referred to as *interleaved* training, is advantageous for learning. The bulk of the evidence supports the superiority of interleaved over blocked training, although there are noteworthy boundary conditions (for recent meta-analyses, see Brunmair & Richter, 2019; Firth et al., 2021). These findings have important potential application in educational practice, especially considering that modifying how examples are presented in an instructional sequence is often within the control of the teacher and does not require additional resources or study time. In other words, with a bit of planning, interleaved training can be a feasible, cost-effective strategy for improving student learning (Roediger & Pyc, 2012). In the remainder of this chapter, I discuss the key research findings related to interleaved training and category learning (e.g., when and why interleaving is beneficial), and the practical implications for the design of instruction.

## Research Evidence

### *Learners' Beliefs and Perceptions*

Imagine that you have not read the preceding chapter overview and are naïve to the topic of interleaved training. Now, consider the following learning scenario (Morehead et al., 2016 [taken from McCabe, 2011]):

> Two art history professors present 6 paintings by each of 12 artists (72 paintings total). Professor A presents all six paintings by a single artist consecutively (i.e., grouped), and then moves on to the next artist's six paintings, and so on, until all paintings have been presented. Professor B presents the various artists' paintings in an intermingled fashion (i.e., mixed), such that a single painting by a particular artist would be followed by a different artist's. (p. 270)

Would you rate the teaching strategy of Professor A or B to be more effective for students' learning to recognize each artist's style? If you thought that Professor B's method was superior, you would be in the minority. In a survey of college students and instructors, only 16% and 13%, respectively, gave a higher effectiveness rating to Professor B than Professor A (Morehead et al., 2016; see also McCabe, 2011). What is particularly interesting about these survey results is that the learning scenario closely resembles the procedure of an actual experiment that was conducted (Kornell & Bjork, 2008), so we can compare subjective beliefs or perceptions (i.e., survey responses) against objective learning (i.e., test performance in the experiment).

In a seminal study, Kornell and Bjork (2008, Experiment 1b) presented college participants with 6 paintings each by 12 artists in either a blocked or interleaved fashion (the artist's name accompanied each painting during training). When later asked to classify previously unseen paintings by the same 12 artists, participants in the interleaved training group were more accurate at identifying the correct artist than those in the blocked training group (59% vs. 36%). In a similar experiment (Kornell and Bjork, 2008, Experiment 1a), participants experienced both training conditions (the paintings by half of the artists were studied in an interleaved manner while the paintings by the other half were blocked), and again interleaved training produced better category learning. A striking mismatch between subjective and actual learning was also revealed: despite the majority of participants exhibiting better learning after interleaved training, when surveyed after the test whether they thought interleaved or blocked training was more effective for their learning, most judged that blocked training was superior.

Learners also show a preference for blocking when they have control over the sequencing of examples during training. Tauber et al. (2013) found that when learning to categorize different types of birds, participants overwhelmingly chose to view multiple examples of a particular bird family before switching to another bird family. The preference for blocking or the belief that it is more effective for learning (relative to interleaving) is likely due to (1) a greater fluency experienced when several examples of the same category are studied consecutively, wherein this subjective ease of processing is interpreted as learning, and (2) the common use of blocked training in education (Yan et al., 2016). Inaccurate metacognitive beliefs or judgments, such as these, can be resistant to correction; I will discuss the pedagogical implications later.

### *Interleaving Benefit for Category Learning*

Since the study by Kornell and Bjork (2008), dozens of other studies have replicated the main finding that interleaved (relative to blocked) training yields superior category learning (e.g., Kang & Pashler, 2012; Kornell et al., 2010). A key point is that across studies, the interleaved and blocked training conditions each involved the same amount of training, identical presentation durations, and the exact

same examples; the only difference was the sequencing of the examples. Most of these studies have investigated the learning of visual categories, in which participants were presented with examples of images belonging to different categories during training and later asked to categorize previously unseen examples from these categories. The visual materials have included paintings (e.g., Sana et al., 2018), birds (e.g., Walheim et al., 2011), butterflies (e.g., Birnbaum et al., 2013), and various artificially generated categories (e.g., Zulkiply & Burt, 2013b).

I will draw attention to a couple of other studies featuring visual categories that have clear educational relevance. Eglington and Kang (2017) examined college students' learning to categorize diagrams of organic chemical compounds (e.g., alkanes, alkynes, carbamates), and they found that interleaving the categories benefited learning, even when the diagnostic features were highlighted during study. This finding is notable because it suggests that (1) the interleaving advantage can generalize to rule-based categories (i.e., category membership can be defined by verbalizable rules; cf. much of the materials used in previous studies, such as paintings, birds), and (2) interleaving is still effective when paired with another useful instructional strategy (highlighting or using visual cues to draw learners' attention to particular features). There is also evidence that interleaving (relative to blocking) examples of different radiographic (or x-ray) patterns (e.g., emphysema, pleural effusion, pneumonia) during training enhances medical students' ability to correctly interpret x-rays (Rozenshtein et al., 2016). Importantly, the presentation of the x-ray examples during training was accompanied by a voice-over description of the disease and the radiographic pattern, akin to how the cases would be presented in typical medical education. Similar results have been found with medical students learning to interpret electrocardiograms (Hatala et al., 2003).

Researchers have also examined the effects of interleaved training on auditory category learning. In an auditory analogue of Kornell and Bjork's (2008) study, Wong et al. (2020) found that interleaving (relative to blocking) excerpts of classical music pieces by various composers during training facilitated listeners' ability to correctly classify novel pieces by the same composers. A follow-up study found an interleaved training advantage for identifying musical intervals (Wong et al., 2021). Being able to recognize composers' styles and musical intervals are important skills for musicianship, and these findings suggest that interleaved training can have useful applications in music education. In addition, a study found that nursing students' auscultation skills (using a stethoscope) improved more with interleaved than blocked training, and students were better able to accurately classify various cardiac and respiratory sounds, especially novel sounds (Chen et al., 2015, Study 3). These studies demonstrating interleaved training benefits for auditory categorical learning provide a nice parallel to the previously mentioned findings on interpreting x-rays and electrocardiograms.

Another broad type of material that has been researched in interleaved training studies are verbal concepts. Zulkiply et al. (2012) examined college students' learning of six types of psychological disorders by studying three case studies of each type. The case studies were presented either interleaved or blocked by disorder type and accompanied by made-up names of the disorders to minimize the influence of prior knowledge or assumptions. Participants were then asked to categorize new case studies, and the interleaved training group exhibited superior performance relative to the blocked group. A similar pattern of results was observed by Sana et al. (2017, Experiment 1) in their study of college students' learning of statistics concepts (three nonparametric statistical tests). Word problems were presented interleaved or blocked by concept during training, and each problem would describe the research design appropriate for the listed statistical concept. Test results demonstrated that participants in the interleaved training group were more accurate at classifying new problems. In the above two studies, the definitions of the verbal concepts were not provided during the case study or word problem presentations. It is probably more common for educators to present the definition before

giving examples to illustrate the concept. There is evidence that when students first study the definitions for each concept followed by the presentation of illustrative examples, interleaving (compared to blocking) the examples leads to more accurate categorization of new examples (Rawson et al., 2015, Experiment 2).

A relatively new area of research on interleaved training has targeted second language (L2) grammar learning, and the emerging findings are promising. In an experiment conducted with Japanese college students who had been studying English for the past 6 years, Nakata and Suzuki (2019) found that interleaved practice on five different English grammatical categories (simple tense, past tense, first, second, and third conditional) produced more accurate grammatical judgments than blocked training immediately and 1-week after training. In another study conducted with US college students who were complete novices in Spanish (Pan, Lovelett, et al., 2019), when learning to conjugate Spanish verbs, an interleaving advantage (over blocking) on a test (2-days later) emerged only when there was systematic alternation (ABABAB) during the initial study trials and randomization during the subsequent practice trials. This finding suggests that how interleaving is implemented may matter, especially when you consider that novices may benefit from greater predictability during initial study. Given that many factors influence L2 learning (Long & Doughty, 2009), more research is needed in order to gain greater clarity on the optimal way to interleave for grammatical category learning.

The selection of studies discussed above demonstrates the breadth of learning materials that have been featured in the research supporting an interleaving advantage for category learning, which increases confidence that the effect will likely generalize to a wide range of educationally relevant materials. In addition, a few of these studies had the categorization test occur a couple of days to a week after training (rather than shortly after training), and the interleaving effect did replicate even with the greater test delays (e.g., Eglington & Kang, 2017; Zulkiply, 2013; Zulkiply & Burt, 2013a), providing some assurance that the benefit of interleaving persists over educationally-relevant time. There are, however, a few known moderating factors which may eliminate or reverse the interleaving advantage; these will be acknowledged in the next section.

A limitation of the existing research is that most of it was conducted on college students in lab-based studies. There is evidence that the interleaving benefit for visual category learning and verbal concept learning does not vary as a function of working memory capacity – i.e., individuals who measure high or low in working memory capacity both benefit from interleaving (Sana et al., 2018; Yan & Sana, 2021). Also, the interleaving advantage for learning painting styles has been replicated in older adults (Kornell et al., 2010), and there is strong evidence that interleaved practice benefits children's mathematics problem solving (e.g., Nemeth et al., 2019; Rohrer & Hartwig, this volume). Future classroom studies assessing the impact of interleaved training on the learning of curriculum-related categories and in specific populations (e.g., students with learning difficulties, autism) would facilitate greater translation of research findings into educational practice.

## *Why Does Interleaving Promote Category Learning, and What are the Boundary Conditions?*

Although the interleaving benefit is generally reliable and robust across individuals in the specific domains and with the learning materials described earlier in this chapter, there are certain boundary conditions under which interleaving is no longer advantageous or blocking becomes superior. In order to better appreciate what these conditions are (or when they arise), it is helpful to understand why interleaving promotes category learning. The dominant theory is that interleaving juxtaposes examples from different categories, which facilitates appreciation of differences among categories (discriminative contrast; Kang & Pashler, 2012; Wahlheim et al., 2011).

In learning a category, it is not enough to figure out how it differs from others; it is also crucial to identify the commonalities within each category, a process that is facilitated when examples are blocked. Therefore, whether a blocked or interleaved sequence will be optimal for learning depends critically on where the difficulty of categorization lies (Carvalho & Goldstone, 2015). If the categories are very similar to each other and it is very challenging to tell them apart, then interleaved training will be superior, as it promotes discrimination. However, if within each category there is high variability, then blocked training will be beneficial, as it directs attention toward the similarities within a category. Indeed, there is evidence to support these predictions. Carvalho and Goldstone (2014) showed that when they used artificially created visual categories that had low within- and between-category similarity, blocking was superior for learning (see also Zulkiply & Burt, 2013b).

Category similarity also likely explains other prominent failures to obtain an interleaving benefit. For instance, Carpenter and Mueller (2013) examined the learning of various French spelling-to-sound rules, and found that blocked (relative to interleaved) presentation of French words pronounced aloud, accompanied by their spelling, produced better learning of French pronunciation rules. Importantly, there was no highlighting in the visual presentation of the spelling to indicate which part of the word might be relevant for the target rule. It is likely that the French spelling-to-sound rules had minimal overlap (low between-category similarity), and the items also had low within-category similarity (each word contained other sounds and letters aside from the target spelling-to-sound rule), and interleaving interfered with commonality abstraction and thus the learning of the categories.

In another study, Yan and Schuetze (2022) presented US college students with Chinese characters and their English translations. Participants were told that the Chinese characters came from 5 different categories (e.g., *branch*, *bridge*, *cabinet*, *chair*, *pole* belong to the WOOD category), but they were not told what the categories were or of which category each character was a member. Perhaps unsurprisingly, category learning was low, and interleaved training did not yield an advantage. However, when participants were made aware that the characters contained a visual feature (a radical) that indicated category membership, and they were further informed about the semantic category to which each character belonged, interleaved training produced a strong benefit over blocked training for accuracy at identifying the meaning of new characters and the semantic category of each of the 5 radicals. It seems that for these verbal categories, if the categories are too opaque, there needs to be scaffolding to help learners become aware of the categories (e.g., making clear during training the category to which each example belongs), without which interleaved training will not be effective.

Another factor that presents a boundary condition for the interleaving advantage is the time gap between examples during training. If the advantage derives from the juxtaposition of examples from different categories, then adding a time delay between successive examples would hinder the comparing and contrasting process which promotes discrimination. Birnbaum et al. (2013) found that inserting a trivia question between each successive example eliminated the interleaving advantage (see also Sana et al., 2017), thus providing a warning against the inadvertent disruptive effects of superfluous additions to the learning sequence.

In the next section, I will generate implications for instruction by synthesizing the empirical findings, theoretical accounts, and boundary conditions related to interleaved training.

## Practical Advice for Educators

There is substantial research literature on interleaved vs. blocked training and its impact on category learning, a topic that has garnered considerable interest over the past 10 plus years. Most of the extant research has been conducted in lab settings, and although that may raise some questions about generalizability to educational settings, these lab studies often have strong methodological rigor and

control extraneous variables. They include replications of past work, allow for causal conclusions (in the case of experiments), and contribute to theoretical advancement. Overall, the weight of the evidence favors interleaved training as a beneficial instructional strategy (e.g., Firth et al., 2021), especially given the ubiquity of category learning across academic disciplines and the low cost of incorporating interleaving in one's teaching practice.

The learning of verbal concepts (e.g., new terms and their definitions, examples which elucidate concepts) is part and parcel of every subject or discipline. Some disciplines go further and require perceptual category learning (e.g., learning to recognize visual images, objects). As such, there are many opportunities for instructors to consider how examples should be sequenced during study or training. Listed below are a few considerations.

## *Which Concepts or Content Should I Interleave?*

Based on theory and empirical findings (e.g., Brunmair & Richter, 2019; Kang & Pashler, 2012), interleaving is most effective when the categories are similar to each other. There is not much use in intermixing highly disparate content (e.g., Indonesian vocabulary and human anatomy; Hausman & Kornell, 2014). As a teacher, you are a subject matter expert and should have an idea (from assignments, assessments, student feedback) of which concepts or categories your students tend to confuse. Remember, though, that the confusion can sometimes be within a category (e.g., the category is highly variable, and a student may have difficulty figuring out what the different examples from the category have in common). Perhaps a simpler rule-of-thumb would be to juxtapose confusing examples. If the examples come from different categories, then you would be interleaving the different categories, which would draw students' attention to the differences among the categories. If the examples come from the same category, then you would be blocking (i.e., grouping the examples from that one category together), which would draw students' attention to the similarities within the category. Both these complementary processes should help reduce confusion.

Also, keep in mind that the research literature I reviewed focused on the presentation of examples (e.g., images, brief case studies, vignettes), and we should not assume that the benefits of interleaved training would extend to reading of longer texts, such as a book chapter. After all, longer chunks of content would mean that there are fewer chances for juxtaposition, which prevents the comparing across examples. Additionally, when presenting the examples, explicitly state the category name and make key features salient to students when possible (Eglington & Kang, 2017; Yan & Schuetze, 2022).

## *When Should I Interleave?*

Perhaps the obvious choice would be during classroom instruction when introducing a new concept. Most of the reviewed research examined novices provided with new material to learn. But there is no compelling reason to expect that interleaved training needs to be confined to the classroom and only at the juncture when a new concept is being taught. There are strategies that instructors can use to incorporate interleaved training (exposing students to examples intermixed across categories) in practice assignments or homework, which could amplify interleaving benefits. For example, Pan, Tajran, et al. (2019) found that interleaved training produced better Spanish grammatical category learning (i.e., ability to conjugate verbs in different tenses) than blocked training, but only if the training was across two sessions that were 1 week apart. Interleaving across multiple sessions typically means that there is an added benefit of spaced practice, which is another powerful strategy to enhance learning (see Kang, 2016, for a review).

## *How Should I Interleave?*

If you rely heavily on traditional textbooks and learning resources for your teaching, you probably will want to set aside a bit of time to plan. The conventional structure for most textbooks tends to feature a linear, topical arrangement, with relatively few interleaved examples (Rohrer et al., 2020), which means you will likely have to intentionally select and rearrange the sequencing of examples in your teaching materials.

Aside from the planning, the other critical aspect is the selection or operationalisation of the actual learning sequence. Although the vast majority of past research has compared (pure) interleaved vs. (pure) blocked training, those two options are obviously not the only way to sequence examples. For instance, having a mix of blocked and interleaved training could perhaps allow learners to benefit from both the commonality abstraction supported by blocking and the discriminative contrast fostered by interleaving. Yan et al. (2017) examined the efficacy of various hybrid schedules on participants' learning of painting styles, and found that a blocked-to-interleaved schedule (i.e., start out with blocked training and then transition to interleaved training) was just as effective as an interleaved schedule. Interestingly, when participants were asked to judge which of 6 different schedules was most effective for learning, the blocked-to-interleaved option was most frequently endorsed, suggesting perhaps some appreciation of the benefit of interleaving (but the pure interleaved schedule was the least endorsed, even though it was one of the most effective schedules).

## *Metacognitive Considerations Surrounding Interleaved Training*

As discussed earlier, there is ample evidence that students (and also instructors) tend to perceive blocked training as more effective than interleaved training (e.g., Kornell et al., 2010; Morehead et al., 2016). This metacognitive illusion is a barrier to greater uptake of interleaved practice, as students are unlikely to embrace learning strategies that they think are unproductive, and they might also regard teachers who promote such strategies as ineffective. Learning strategies that feel more effortful are interpreted by students as less effective for learning, and hence, less likely to be selected in the future (Kirk-Johnson et al., 2019). Therefore, it is crucial to find ways to correct learners' faulty beliefs. But these beliefs can be very resistant to change. Yan et al. (2016) found that when participants experienced blocked and interleaved training for learning painting styles, the majority (72%) judged blocked training to be more effective for their learning, despite actual learning showing an interleaving advantage. They tried to correct this metacognitive illusion by providing participants with information about interleaved training being more effective for 90% of people, asking them to ignore their sense of fluency during blocked training, and giving an explanation for why interleaving is beneficial, yet almost half still stuck to their belief that blocked training was superior. To change more minds, it was necessary to also have the blocked and interleaved training and test be experienced separately, so that test performance could be connected unambiguously to the appropriate training condition (Yan et al., 2016, Experiment 6).

One approach is to capitalize on students' greater openness to hybrid blocked-to-interleaved schedules (Yan et al., 2017). In other words, start with blocked training (something to which students are already quite accustomed), and then incorporate progressively greater interleaving. Another possibility is for teachers to conduct demonstrations of the interleaving effect in their classrooms with their students, either using paintings by different artists (like in Kornell & Bjork, 2008) or using other kinds of categories that might be more germane to the teacher's subject. Importantly, divide categories into two lists and have the blocked and interleaved training and test done separately. The point of the demonstration is for students to be able to clearly link their test outcome to a given training condition. Hopefully, a better outcome after interleaved training is obtained, which together with a debriefing containing information

about blocked vs. interleaved training and the purpose of the class mini-experiment, will help establish correct beliefs and knowledge about interleaved training among your students.

The prevailing teaching norms favor blocked training (Kang, 2017). Blocking feels effective for learning, instructional materials facilitate it, and teachers and students are used to it. But there is a growing body of research showing that increasing the degree of interleaving during training is often beneficial for inductive learning. As a teacher, you can decide how examples are presented in class and how practice is sequenced. Although incorporating greater interleaving will require some effort and planning, it does not require a dramatic change in your current practice and need not be onerous. May you encounter many successes as you explore inventive ways of including and promoting interleaved training and other empirically supported strategies in your classroom.

## Author Note

Correspondence concerning this article should be addressed to Sean Kang, Melbourne Graduate School of Education, 100 Leicester Street, Level 6, Carlton, VIC 3053, Australia. E-mail: sean.kang@unimelb.edu.au

## References

Birnbaum, M. S., Kornell, N., Bjork, E. L., & Bjork, R. A. (2013). Why interleaving enhances inductive learning: The roles of discrimination and retrieval. *Memory & Cognition*, *41*(3), 392–402. https://doi.org/10.3758/s13421-012-0272-7

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

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). What you learn is more than what you see: What can sequencing effects tell us about inductive category learning? *Frontiers in Psychology*, *6*, 505. https://doi.org/10.3389/fpsyg.2015.00505

Chen, R., Grierson, L., & Norman, G. (2015). Manipulation of cognitive load variables and impact on auscultation test performance. *Advances in Health Sciences Education*, *20*(4), 935–952. https://doi.org/10.1007/s10459-014-9573-x

Eglington, L. G., & Kang, S. H. K. (2017). Interleaved presentation benefits science category learning. *Journal of Applied Research in Memory and Cognition*, *6*(4), 475-485. https://doi.org/10.1016/j.jarmac.2017.07.005

Firth, J., Rivers, I., & Boyle, J. (2021). A systematic review of interleaving as a concept learning strategy. *Review of Education, 9*(2), 642–684. https://doi.org/10.1002/rev3.3266

Hatala, R. M., Brooks, L. R., & Norman, G. R. (2003). Practice makes perfect: The critical role of mixed practice in the acquisition of ECG interpretation skills. *Advances in Health Sciences Education*, *8*(1), 17–26. https://doi.org/10.1023/a:1022687404380

Hausman, H., & Kornell, N. (2014). Mixing topics while studying does not enhance learning. *Journal of Applied Research in memory and Cognition*, *3*(3), 153–160. https://doi.org/10.1016/j.jarmac.2014.03.003

Kang, S. H. K. (2016). Spaced repetition promotes efficient and effective learning: Policy implications for instruction. *Policy Insights from the Behavioral and Brain Sciences*, *3*(1), 12–19. https://doi.org/10.1177/2372732215624708

Kang, S. H. K. (2017). The benefits of interleaved practice for learning. In J. C. Horvath, J. M. Lodge, & J. Hattie (Eds.), *From the laboratory to the classroom: Translating science of learning for teachers* (pp. 79–93). Routledge.

Kang, S. H. K., & Pashler, H. (2012). Learning painting styles: Spacing is advantageous when it promotes discriminative contrast. *Applied Cognitive Psychology, 26*(1)*,* 97–103. https://doi.org/10.1002/acp.1801

Kirk-Johnson, A., Galla, B. M., & Fraundorf, S. H. (2019). Perceiving effort as poor learning: The misinterpreted-effort hypothesis of how experienced effort and perceived learning relate to study strategy choice. *Cognitive Psychology*, *115*, 101237. https://doi.org/10.1016/j.cogpsych.2019.101237

Kornell, N., & Bjork, R. A. (2008). Learning concepts and categories: Is spacing the "enemy of induction"? *Psychological Science, 19*(6)*,* 585–592. https://doi.org/10.1111/j.1467-9280.2008.02127.x

Kornell, N., Castel, A. D., Eich, T. S., & Bjork, R. A. (2010). Spacing as the friend of both memory and induction in young and older adults. *Psychology and Aging, 25*(2)*,* 498–503. https://doi.org/10.1037/a0017807

Long, M. H., & Doughty, C. J. (Eds.). (2009). *The handbook of language teaching.* Wiley-Blackwell.

McCabe, J. (2011). Metacognitive awareness of learning strategies in undergraduates. *Memory & Cognition*, *39*(3), 462–476. https://doi.org/10.3758/s13421-010-0035-2

Morehead, K., Rhodes, M. G., & DeLozier, S. (2016). Instructor and student knowledge of study strategies. *Memory*, *24*(2), 257–271. https://doi.org/10.1080/09658211.2014.1001992

Nakata, T., & Suzuki, Y. (2019). Mixing grammar exercises facilitates long-term retention: Effects of blocking, interleaving, and increasing practice. *The Modern Language Journal*, *103*(3), 629-647. https://doi.org/10.1111/modl.12581

Nemeth, L., Werker, K., Arend, J., Vogel, S., & Lipowsky, F. (2019). Interleaved learning in elementary school mathematics: Effects on the flexible and adaptive use of subtraction strategies. *Frontiers in Psychology*, *10*(article 86). https://doi.org/10.3389/fpsyg.2019.00086

Pan, S. C., Lovelett, J. T., Phun, V., & Rickard, T. C. (2019). The synergistic benefits of systematic and random interleaving for second language grammar learning. *Journal of Applied Research in Memory and Cognition*, *8*(4), 450-462. https://doi.org/10.1016/j.jarmac.2019.07.004

Pan, S. C., Tajran, J., Lovelett, J., Osuna, J., & Rickard, T. C. (2019). Does interleaved practice enhance foreign language learning? The effects of training schedule on Spanish verb conjugation skills. *Journal of Educational Psychology*, *111*(7), 1172–1188. https://doi.org/10.1037/edu0000336

Rawson, K. A., Thomas, R. C., & Jacoby, L. L. (2015). The power of examples: Illustrative examples enhance conceptual learning of declarative concepts. *Educational Psychology Review*, *27*(3), 483–504. https://doi.org/10.1007/s10648-014-9273-3

Roediger, H. L. III, & Pyc, M. A. (2012). Inexpensive techniques to improve education: Applying cognitive psychology to enhance educational practice. *Journal of Applied Research in Memory and Cognition, 1*(4), 242–248. https://doi.org/10.1016/j.jarmac.2012.09.002

Rohrer, D., Dedrick, R. F., & Hartwig, M. K. (2020). The scarcity of interleaved practice in mathematics textbooks. *Educational Psychology Review*, *32*(3), 873–883. https://doi.org/10.1007/s10648-020-09516-2

Rozenshtein, A., Pearson, G. D., Yan, S. X., Liu, A. Z., & Toy, D. (2016). Effect of massed versus interleaved teaching method on performance of students in radiology. *Journal of the American College of Radiology*, *13*(8), 979–984. https://doi.org/10.1016/j.jacr.2016.03.031

Tauber, S. K., Dunlosky, J., Rawson, K. A., Wahlheim, C. N., & Jacoby, L. L. (2013). Self-regulated learning of a natural category: Do people interleave or block exemplars during study? *Psychonomic Bulletin & Review*, *20*(2), 356–363. https://doi.org/10.3758/s13423-012-0319-6

Sana, F., Yan, V. X., & Kim, J. A. (2017). Study sequence matters for the inductive learning of cognitive concepts. *Journal of Educational Psychology*, *109*(1), 84–98. https://doi.org/10.1037/edu0000119

Sana, F., Yan, V. X., Kim, J. A., Bjork, E. L., & Bjork, R. A. (2018). Does working memory capacity moderate the interleaving benefit? *Journal of Applied Research in Memory and Cognition*, *7*(3), 361–369. https://doi.org/10.1016/j.jarmac.2018.05.005

Wahlheim, C. N., Dunlosky, J., & Jacoby, L. L. (2011). Spacing enhances the learning of natural concepts: An investigation of mechanisms, metacognition, and aging. *Memory & Cognition*, *39*(5), 750–763. https://doi.org/10.3758%2Fs13421-010-0063-y

Wong, S. S. H., Chen, S., & Lim, S. W. H. (2021). Learning melodic musical intervals: To block or to interleave? *Psychology of Music*, *49*(4), 1027–1046. https://doi.org/10.1177/0305735620922595

Wong, S. S. H., Low, A. C. M., Kang, S. H. K., & Lim, S. W. H. (2020). Learning music composers' styles: To block or to interleave? *Journal of Research in Music Education*, *68*(2), 156–174. https://doi.org/10.1177/0022429420908312

Yan, V. X., Bjork, E. L., & Bjork, R. A. (2016). On the difficulty of mending metacognitive illusions: A priori theories, fluency effects, and misattributions of the interleaving benefit. *Journal of Experimental Psychology: General*, *145*(7), 918–933. https://doi.org/10.1037/xge0000177

Yan, V. X., & Sana, F. (2021). The robustness of the interleaving benefit. *Journal of Applied Research in Memory and Cognition, 10*(4), 589–602. https://doi.org/10.1037/h0101863

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

Zulkiply, N. (2013). Effect of interleaving exemplars presented as auditory text on long-term retention in inductive learning. *Procedia-Social and Behavioral Sciences*, *97*(6), 238–245. https://doi.org/10.1016/j.sbspro.2013.10.228

Zulkiply, N., & Burt, J. S. (2013a). Inductive learning: Does interleaving exemplars affect long-term retention? *Malaysian Journal of Learning and Instruction*, *10*, 133-155. https://e-journal.uum.edu.my/index.php/mjli/article/view/7655

Zulkiply, N., & Burt, J. S. (2013b). The exemplar interleaving effect in inductive learning: Moderation by the difficulty of category discriminations. *Memory & Cognition*, *41*(1), 16–27. https://doi.org/10.3758/s13421-012-0238-9

Zulkiply, N., McLean, J., Burt, J. S., & Bath, D. (2012). Spacing and induction: Application to exemplars presented as auditory and visual text. *Learning and Instruction*, *22*(3), 215–221. https://doi.org/10.1016/j.learninstruc.2011.11.002