# Using Interleaving to Promote Inductive Learning in Educational Contexts: Promises and Challenges

Tobias Richter[1], Lea Nemeth[2], Roland Berger[3], Rita Borromeo Ferri[4], Martin Hänze[5] & Frank Lipowsky[2]

[1] Department of Psychology IV, University of Würzburg

[2] Department of Educational Science, University of Kassel

[3] Department of Physics, University of Osnabrück

[4] Department of Mathematics and Natural Sciences, University of Kassel

[5] Department of Psychology, University of Kassel

**Accepted for publication in *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie/German Journal of Developmental and Educational Psychology* (2022)**

**Author Note**

Tobias Richter https://orcid.org/0000-0002-0467-9044

Lea Nemeth https://orcid.org/0000-0002-6212-6042

Roland Berger https://orcid.org/0000-0002-3539-8827

Rita Borromeo Ferri https://orcid.org/0000-0001-8223-3729

Martin Hänze https://orcid.org/0000-0002-9345-6175

Frank Lipowsky https://orcid.org/0000-0002-0881-6704

We have no known conflict of interest to disclose. This article is product of a collaboration of the authors within the research unit “Lasting learning: Cognitive mechanisms and effective instructional implementation” (FOR 5254), funded by the German Research Foundation (DFG) (project number 450142163; grants BE 2652/6-1, BO 2949/6-1, HA 3509/9-1, LI 1639/5-1, RI 1100/16-1, RI 1100/17-1).

Correspondence concerning this article should be addressed to Tobias Richter, University of Würzburg, Department of Psychology IV, Röntgenring 10, 97070 Würzburg, Germany. Email: tobias.richter@uni-wuerzburg.de

**Abstract**

Inductive learning, that is, the abstraction of conceptual knowledge, rules, or principles from exemplars, plays a major role in educational settings, from literacy acquisition to mathematics and science learning. Interleaving exemplars of different categories compared to a blocked presentation might be a simple but powerful way to improve inductive learning by supporting discriminative contrast. Although a consistent advantage of interleaving has been demonstrated for visual materials, relatively few studies have examined educationally relevant materials, such as mathematical tasks, science problems, and verbal materials, and their results are mixed. We discuss how interleaving could be made fruitful for school learning of mathematics, science, and literacy acquisition. We conclude that interleaving should be tailored to the specific learning content and combined with supportive instructional measures that assist students in comparing exemplars for discriminating features. Finally, we will sketch research gaps that revolve around the use of interleaved learning in the classroom.

*Keywords*: exemplars, inductive learning, learning sequence, literacy acquisition, mathematics, science

**Zusammenfassung**

Induktives Lernen, d. h. die Abstraktion von konzeptuellem Wissen, Regeln oder Prinzipien aus Beispielen, ist zentral für das schulische Lernen– vom Schriftspracherwerb bis hin zur Mathematik und den Naturwissenschaften. Das Verschachteln von Beispielen verschiedener Kategorien ist verglichen mit einer geblockten Präsentation eine wirkungsvolle Methode zur Optimierung induktiver Lernprozesse, indem es Vergleichsprozesse zwischen den Beispielen evoziert. Empirisch hat sich konsistent ein Vorteil des verschachtelten Lernens für visuelle Materialien gezeigt, es gibt allerdings nur wenige und uneinheitliche Befunde zur Wirksamkeit des Verschachtelns schulisch relevanter Materialien. Wir gehen der Frage nach, wie verschachteltes Lernen im unterrichtlichen Kontext – beispielhaft anhand der Domänen Mathematik, Naturwissenschaften und dem Schriftspracherwerb – nutzbar gemacht werden kann. Neben der Wahl geeigneter Inhalte scheint die Implementation weiterer instruktionaler Maßnahmen, mit denen Schülerinnen und Schüler beim Vergleichen unterstützt werden, dessen Wirksamkeit zu begünstigen. Abschließend werden Forschungslücken skizziert, die im Hinblick auf die Nutzung des verschachtelten Lernens im Unterricht bestehen.

*Keywords*: Beispiele, induktives Lernen, Lernabfolgen, Schriftspracherwerb, Mathematik, Naturwissenschaften

# Using Interleaving to Promote Inductive Learning in Educational Contexts: Promises and Challenges

Students encounter different content in a certain sequence when learning. Much research in instructional psychology and subject didactics has addressed the question of how topics should be arranged when learning to optimize the learning process. For example, one core element of adaptive instruction is to create individualized sequences of content that are tailored to the individual learner's needs and abilities (e.g., Skinner, 1986). Approaching the issue from a different perspective, direct instruction embraces the principle that teachers should present subtopics in a factually based and logically structured sequence (Huitt et al., 2009). Similarly, a major issue in subject-matter education is how to arrange the topics specified in curricula to enhance learning. The merits of these approaches notwithstanding, they usually focus on the arrangement of learning content on a macrolevel, telling us little about how learning content should be arranged in inductive learning settings on a microlevel, for example, different tasks, procedures, or subskills.

Inductive learning refers to the process of abstracting and consolidating conceptual knowledge (e.g., generic facts, rules, principles, or categories) from studying exemplars. It is a major form of human learning that occurs in many formal and informal learning settings, from babies learning new words to doctors learning to interpret X-rays (Kornell & Bjork, 2008). Inductive learning also plays an important role in school learning, although the term is seldom addressed explicitly in educational psychology textbooks. When students practice solving math problems (e.g., Rohrer & Taylor, 2007), when they derive scientific principles from a set of observations (e.g., Prince & Felder, 2007), when they practice reading to grasp a grapheme-phoneme correspondence rule (e.g., Müller et al., 2020), or when they work through examples to understand a grammatical rule in learning a foreign language (e.g., DeKeyser, 1995), they perform inductive learning. In most classroom applications, inductive learning is not used in its pure form but is combined with deductive elements. Thus, introducing the rule or concept to be learned first in a more or less explicit and comprehensive fashion before students practice it by means of examples is commonplace. Herron and Tomasello (1992) coined the term "guided induction" for this procedure. In guided induction, the inductive phase may not be misunderstood as a mere application and consolidation of the concept taught in the beginning. Instead, the bulk of learning takes place in the inductive learning phase, which also takes up most of the class time.

The question is then how inductive learning settings at school should be designed to make them most effective, in particular for lasting learning. One possible and simple way to improve inductive learning might be to change the sequence in which exemplars are presented. Research suggests that interleaving exemplars from different categories (e.g., A – B – C – A – C – A – C – B – C) might support the acquisition of the underlying concepts under certain circumstances compared to "blocking" the same exemplars (e.g., A – A – A – B – B – B – C – C – C), that is, presenting all items of one category or concept before the items of another category or concept are presented (e.g., Kang & Pashler, 2012; Kornell & Bjork, 2008; Rohrer & Taylor, 2007).

In this article, we discuss the principle of interleaved learning with a focus on its suitability for school-based learning. We start by reviewing theories to explain why and when interleaving is effective for learning and discuss empirical studies that shed light on this approach to learning. Based on the discussion of existing theoretical and empirical work, we propose that to unfold its full potential for school-based learning, interleaving should be tailored to the specific learning content and combined with supportive instructional measures that assist students in cognitive processes conducive to inductive learning. We substantiate and illustrate this proposition by sketching possible applications of interleaving in three central areas of school learning: mathematics learning, science learning, and literacy acquisition. The discussion of these applications will reveal open questions that revolve around the use of interleaved learning in the classroom, which deserve increased research efforts in psychology and the educational sciences.

## When and Why is Interleaving Effective for Learning?

Interleaved learning was popularized by Kornell and Bjork (2008), who showed that students learned the painting styles of various impressionistic painters considerably better when different painters' works were presented interleaved rather than blocked (interleaving effect), although the participants in the experiment reported that they learned better in the blocked compared to the interleaved condition. Accordingly, interleaving is a *desirable difficulty* that makes learning subjectively more difficult but objectively improves learning outcomes (Bjork & Bjork, 2011). The effect has been replicated with the original materials (e.g., Metcalfe & Xu, 2016; Kang & Pashler, 2012), and other experiments have demonstrated interleaving effects with mathematical tasks, photographs, pictures of artificial objects, and physics problems (e.g., Birnbaum et al., 2013; Rohrer & Taylor, 2007; Samani & Pan, 2021 Wahlheim et al., 2011). Still, other studies with various types of materials have reported no interleaving effect (e.g., Carpenter & Mueller, 2013; Dobson, 2011; Sorensen & Woltz, 2016). Brunmair and Richter (2019) conducted a meta-analysis of all published studies (cut-off date June 2018) and found a medium-sized positive overall effect of interleaved learning (Hedges' $g$ = 0.42). However, they also found substantial heterogeneity of effect sizes between studies, suggesting that the advantage of interleaving over blocking occurs in only some learning situation, depending on certain conditions.

Theoretical accounts of interleaved learning provide a key to better understanding the divergent results. The extant theories explain the potential benefits of interleaving by assuming that the interleaved presentation facilitates the detection of differences between exemplars. The discriminative-contrast hypothesis (Kang & Pashler, 2012) posits that presenting exemplars from different categories in temporal proximity facilitates discriminative contrasts, which augments category learning. Discriminative contrasts refer to comparisons directed at distinguishing features of exemplars such as features of paintings that distinguish the styles of different painters in the experiments by Kang and Pashler (2012). Evidence for the hypothesis comes from experiments showing that temporal spacing between items hinders inductive learning in an interleaved presentation but not in a blocked presentation. The sequential attention theory (Carvalho & Goldstone, 2017; also called attentional bias framework, Carvalho & Goldstone, 2015) adopts the idea of discrimination as the mechanism of interleaved learning and additionally specifies the conditions that render discriminative contrasts effective. According to this theory, an interleaved presentation of exemplars facilitates the detection of differences, whereas a blocked presentation facilitates the detection of similarities. Thus, the effect of interleaving emerges when the to-be-learned categories are difficult to discriminate. In contrast, the acquisition of clearly distinguishable categories, which are defined more by similarities of category members, works well—or even better—with a blocked presentation of exemplars. Carvalho and Goldstone (2017) provided evidence for these assumptions in experiments in which participants learned to distinguish between species of aliens by looking at sequentially presented pictures. An interleaved presentation directed the attentional focus to features discriminating between species and to a better encoding of such features, whereas a blocked study led to a better encoding of features common to one category (for a presentation and discussion of further evidence, see also Carvalho & Goldstone, 2014, 2015).

In their meta-analysis, Brunmair and Richter (2019) found support for the basic predictions of the sequential theory of attention. In a meta-regression, a positive effect of the similarity of exemplars between categories and a negative effect of the similarity of exemplars within categories on the size of the interleaving effect emerged. A second finding of particular importance was that the type of learning material moderated the effects. The positive interleaving effect was strongest (Hedges' $g = 0.67$) for visual materials, such as naturalistic paintings (such as those used by Kornell & Bjork, 2008; or Kang & Pashler, 2012), and pictures of artificial objects (*g* = 0.36) that are typically used as stimuli in category learning experiments (e.g., Carvalho & Goldstone, 2014). In comparison, the effect of interleaving was still positive but smaller for mathematical tasks (*g* = 0.34), not significantly different from zero when texts were used as learning materials, and even negative when words were used as learning materials (*g* = -0.39).

These findings suggest that for learning materials that are typically used in inductive learning in school contexts, interleaving might not be sufficient to induce the comparison processes that enable learners to successfully discriminate between different concepts. Possible causes lie in the higher complexity and dependence on prior knowledge that characterize such learning materials. These features are likely to make comparison processes not only more cognitively demanding but also give them the character of cognitive learning strategies that students may not be expected to apply routinely, either because they lack relevant declarative or procedural knowledge, or because they lack the relevant conditional knowledge of when the strategies are useful (Schneider et al., 2022). With instructional interventions that specifically support students in using comparison strategies tailored to the students' abilities, the learning materials, and the learning objectives, interleaving could perform to its full potential for school learning. In the following sections, we will elaborate on these assumptions for inductive learning in three areas: mathematics, the natural sciences, and literacy acquisition.

## The Potential Benefits of Interleaved Learning in Mathematics

A central goal of mathematics instructions is to enable students to switch flexibly between different solution methods, strategies, or formulas and thus to choose an appropriate solution approach leading to faster and more accurate problem solving (Baroody & Dowker, 2003; Heinze et al., 2009). Therefore, learners need to know not only about different strategies (declarative knowledge) and how to use them (procedural knowledge) but also when and why to use which strategy for which type of task (conditional knowledge; Paris et al., 1983). Choosing appropriate methods for solving mathematical tasks is often very challenging for students, especially when the tasks are superficially similar (Rohrer et al., 2015). For example, subtraction tasks, such as 514 – 199 and 342 – 338, are superficially similar but differ in the most appropriate solution method. Adding one to both integers is more efficient for the first task (i.e., 515 – 200), subtracting by adding (i.e., 338 + _ = 342) is more efficient for the second task. When solving algebra problems, such as x + 2x – 1, students also must choose a useful strategy (e.g., factoring, quadrating, adding; Rohrer et al., 2015). In such situations, interleaved learning can be an appropriate instructional strategy.

**Studies on Interleaving in Mathematics Learning**

When introducing new solution methods, strategies, or formulas for a specific mathematical topic, teachers usually teach them and their application conditions in a blocked fashion (Rohrer, Dedrick, & Hartwig, 2020). Thus, students often practice different solution methods separate from each other. However, students need to learn to recognize and discriminate the differences between superficially similar mathematical tasks to match the problems to an appropriate strategy.

Interleaving mathematical tasks should directly evoke these discrimination processes (Birnbaum et al., 2013) and help students in abstracting rules and principles about when to use which strategy for which kind of task. However, the meta-analysis by Brunmair and Richter (2019) only revealed a small effect of interleaving mathematical tasks, with primary studies ranging from strongly negative to positive effects. Still, several studies have shown an advantage of interleaving over blocking mathematical tasks for different topics such as graph and slope problems (Rohrer et al., 2015), equations and terms (Rohrer, Dedrick, Hartwig, & Cheung, 2020), algebraic addition and multiplication (Ziegler & Stern, 2014, 2016), surface and volume calculations of solids (Rohrer & Taylor, 2007), and subtraction strategies (Nemeth et al., 2019, 2021).

**How to Implement Interleaving in Mathematics Learning**

Some studies on interleaving mathematical tasks present the mathematical tasks in an interleaved fashion but also implemented further instructional guidance, such as explicit prompts to contrast and compare or to self-explain, or they presented the tasks simultaneously instead of successively (e.g., Nemeth et al., 2019, 2021; Ziegler & Stern, 2014, 2016). In these studies, longer-lasting effects up to 10 weeks of interleaving mathematical tasks have been observed, whereas studies investigating a pure form of interleaved practice in mathematics have predominantly focused on shorter time intervals with a maximum of about 30 days (e.g., Rohrer, Dedrick, Hartwig, & Cheung, 2020). Therefore, additional instructional support might be necessary or at least helpful for lasting positive effects of interleaving in mathematics.

Providing indirect support for this assumption, Ziegler et al. (2018) compared the effectiveness of interleaving algebraic multiplication and addition problems in an implicit and explicit learning condition. In their study, interleaving mathematical tasks was more effective when different forms of additional instructional guidance (i.e., self-explanation prompts and comparison prompts) were included, and the effects showed even after 10 weeks. However, systematic research disentangling the effects of these different forms of instructional support on interleaving mathematical tasks is lacking. We will discuss three types of instructional support, self-explanation prompts, comparison prompts, and simultaneous presentation and how they might be combined with interleaved learning to boost its effectiveness in mathematics learning.

*Generating self-explanations* is a learning technique that promotes deep processing and transfer (Chi et al., 1994). Several studies have shown positive effects when including prompts to self-explaining in mathematics (Rittle-Johnson et al., 2017). Chi (2000) proposed two main mechanisms explaining the positive impact of self-explaining: (1) Self-explaining promotes knowledge integration by either connecting different pieces of new information or integrating them into existing knowledge structures. (2) Self-explaining facilitates the recognition of central structural features, promoting knowledge transfer of the learned content. We suppose that prompting students to self-explain why they use a specific solution method for a specific mathematical task can support students in abstracting rules and principles and in acquiring conditional knowledge about when and why specific solution methods should be used. Therefore, combining interleaved practice, which augments students' ability in choosing appropriate solution methods, with self-explanation might strengthen the interleaving effect.

Comparison processes are expected to be underlying learning mechanisms that explain the advantage of interleaved over blocked practice (Birnbaum et al., 2013; Carvalho & Goldstone, 2015, 2017). However, individuals often neglect to compare different categories unless they are explicitly prompted (Durkin et al., 2017). Previous research on learning by comparison has shown that supporting learners' comparison processes with explicit *comparison prompts* results in higher learning gains than offering only opportunities to compare (e.g., Catrambone & Holyoak, 1989). Studies demonstrate that prompting students to compare different problem types or solution methods for solving mathematical tasks can promote their flexible strategy use (e.g., Rittle-Johnson & Star, 2007, 2009). Therefore, augmenting interleaved practice with explicit comparison prompts could support students when comparing and discriminating interleaved categories.

In most studies, interleaved exemplars have been presented sequentially (Brunmair & Richter, 2019). According to the sequential attention theory, interleaved practice facilitates the identification of differences between categories (Carvalho & Goldstone, 2017). Research on comparison learning in mathematics has shown that a *simultaneous presentation* is often superior to a sequential presentation (Begolli & Richland, 2016; Rittle-Johnson & Star, 2007). Hence, we assume that detecting differences when presenting the interleaved examples or tasks simultaneously instead of sequentially makes comparisons easier for students because recalling previous tasks from working memory is not required to draw comparisons. However, whether simultaneous presentation can augment comparison processes is an open question. Previous research regarding the role of a simultaneous vs sequential presentation in interleaved learning is inconsistent (Brunmair & Richter, 2019) and largely lacking in the domain of mathematics learning.

## Interleaved Learning in the Natural Sciences

The natural sciences of physics, chemistry, and biology are typically regarded as important domains in education. Moreover, inductive learning plays a large role in science education (Prince & Felder, 2007). Despite their importance, the number of studies on interleaved learning in the science classroom is small. For example, a search in the PsychInfo database using the keywords "interleav*" and "physics" or “interleav*” and “chemistry” each yielded only one and “interleav*” and “biology” yielded no relevant hits (as of November 9, 2022). Given the scarcity of studies on interleaving in science learning, we focus our discussion on three exemplary studies and discuss these studies and their implications in the broader context of the meta-analysis by Brunmair and Richter (2019).

### Studies on Interleaving in Science Learning

In a study by Sana and Yan (2022), 155 students in eight classrooms (physics, biolology, chemistry, and integrated science) were taught for four weeks with different materials. At the end of each week, nine multiple-choice quiz tasks in the respective subject were administered blocked by concept (A1A2A3 B1B2B3 C1C2C3) or interleaved with different concepts (A1B1C1 A2B2C2 A3B3C3). One month after the final quiz, students were finally tested on the concepts covered in the four-week period. The results revealed that the scores on the interleaved quizzes were significantly lower than those on the blocked quizzes, albeit with a small effect size (Cohen’s $d = 0.21$). However, the interleaved-quiz condition led to significantly better final test performance than the blocked-quiz condition, with a small to medium effect size ($d = 0.35$). Despite the use of very different scientific content, the results by Sana and Yan (2022) correspond to findings showing in which interleaving is initially less favorable than blocking in the short term but more effective for lasting learning (e.g., Ziegler &

Stern, 2014). Accordingly, interleaved learning in this study would qualify as a desirable difficulty (Bjork & Bjork, 2011).

Another important study was conducted by Samani and Pan (2021) in introductory physics courses at the university level. These authors investigated the interleaving of physics problems that were given as homework to 576 students in 20 lectures over eight weeks. After each lecture, students received blocked and interleaved homework assignments to foster factual knowledge and problem-solving ability. In one condition, the problems were blocked with three practice problems each, for example, on the electric capacitor a) with different filling materials (P1P2P3), b) with different sizes (Q1Q2Q3), and c) their stored energy (R1R2R3). In the second condition, the problems were interleaved, for example, P1Q1R1 M2N2O2 J3K3L3. Two unannounced criterial tests examined content from the first 10 and last 10 lectures, respectively. Both tests showed a significant positive interleaving effect with medium to large effect sizes ($d = 0.40$ and $d = 0.91$). These results are remarkable in several respects. First, similar to the study by Sana and Yan (2022), students performed slightly worse in the interleaved assignments and found the interleaved homework assignments more difficult than the blocked assignments, which is in line with a classification of interleaved learning as a desirable difficulty. Second, the criterial tests were administered after learning phases of four weeks, implying that at least some of the questions referred to materials learned and practiced several weeks before, which underscores the potential of interleaving augmenting lasting learning. Third, the bandwidth of tested physics concepts is relatively large (20 lectures in electrodynamics, atomic and quantum physics), suggesting that the effect generalizes across different content. Fourth, the physics problems used by Samani and Pan were relatively complex, meaning that the kind of inductive learning instigated by the homework assignments and the to-be-learned conceptual knowledge were quite different from the relatively simple tasks and concepts used, for example, in the category learning experiments based on visual materials.

A study by Eglington and Kang (2017) is a third instructive example of a study that has examined interleaved learning in science learning. The authors found advantages of interleaving over blocking in learning chemical compounds in three consecutive experiments with undergraduate students. In Experiment 1, simple chemical structural diagrams of five categories of hydrocarbons of similar structure (e.g., alcohol and alkynes) were presented as visual material. Unique visual critical features defined the categories (e.g., alcohols have an oxygen atom; alkynes have a carbon-carbon triple bond). In the posttest, students in the interleaving condition outperformed students in the blocked condition with a medium effect size ($d = 0.61$). In Experiment 2, the materials included four new, more complex chemical categories, leading to more irrelevant features than in the materials used in Experiment 1. This change of materials allowed manipulating within-category similarity. The benefit of interleaving persisted in spite of increasing the complexity of the chemical categories with a medium effect size ($d = 0.61$). However, no main effect of within-category similarity or interaction of similarity and presentation condition on test performance emerged. Experiment 3 used the same materials, but the critical features of each chemical category were highlighted in red. Similar to the previous results, participants in the interleaved condition outperformed those in the blocked condition, again with a medium effect size ($d = 0.57$). The advantage of interleaving over blocking found by Englington and Kang fits the meta-analytical results reported by Brunmair and Richter (2019) who found a small-to-medium effect size for pictures of artificial objects. At first glance, it might seem surprising and at odds with the sequential theory of attention (Carvalho & Goldstone, 2014). That is, despite an increase in intra-category similarity, interleaving again had an advantage over blocking. However, as Eglington and Kang argued, this finding might be traced back to the specific manipulation of the within-category similarity (limited by the complexity of real-world chemical categories) that was not strong enough for an effect to emerge. The highlighting of distinctive features in Experiment 3 can be regarded as a kind of instructional support that helped students to focus on differentiating features between categories.

**How to Implement Interleaving in the Science Classroom**

Content taught in science classes is usually quite complex and requires prior knowledge to be understood properly, which indicates the need to support students in interleaving tasks. Moreover, learning in the natural sciences often includes the categorization of materials based on visual or other features. The categorization of structural diagrams in Eglington and Kang's (2017) study is a case in point. Further examples from physics and biology are the differentiation of states of matter (solid, liquid, or gas) or the categorization of animals or plants in biology. However, prototypical scientific knowledge is characterized by a complex interplay of observations, its generalization into laws, and the principle-based interpretation within theoretical frames (Chalmers, 1999; Lederman, 2007), with the overarching goal to foster students' scientific reasoning. For example, consider the movement of charge carriers in electric and magnetic fields. On the surface, the two categories are similar, for example, because of their vectorial character. This similarity is probably why students have difficulty discriminating between electric and magnetic fields (Maloney et al., 2001). However, the different motions of charge carriers in these two fields can be easily observed by students (e.g., by using simulations), which also makes the two categories of fields easily distinguishable. However, one key aspect of physics, even in physics education at school, is the prediction of the future behavior of objects based on a theoretical frame. Consequently, students' acquisition of complex scientific concepts requires comprehending the underlying principles. To master this complexity, students need to acquire adequate principle-based cognitive skills (Renkl, 2015). The underlying principles and laws of scientific concepts are often complex and thus difficult to learn.

How should interleaved learning in the science classroom accommodate the complexity of science concepts? Teachers might be prudent to teach these concepts, such as the motion of charge carriers in electric and magnetic fields, or the law of equilibrium of forces and Newton's third law, in blocks. In the next step, contrasted comparisons of concepts can be triggered by the interleaved presentation of corresponding tasks. Following the sequential attention theory (Carvalho & Goldstone, 2017), students can first be supported in recognizing the commonalities within each category to acquire fundamental knowledge about the concepts (e. g., the law of equilibrium of forces and Newton's third law) before the subsequent interleaved learning phase prompts students' attention to the differences of the two concepts. Thus, difficulty increases as students' proficiency increases (Nakata & Suzuki, 2019). This combination of blocking and subsequent interleaving might be fruitful (Kang, 2017). Further instructional guidance such as comparison and self-explanation prompts, and simultaneous instead of sequential presentation might further support interleaved learning in science. Moreover, computer-based simulations of physical phenomena offer a straightforward way to point students to crucial differences between exemplars. Similarly, phases of direct instruction that convey the relevant principles before they are deepened and consolidated by solving interleaved science problems might be useful. The studies by Samani and Pan (2021) and by Sana and Yan (2022) followed this principle, using quizzes or homework assignments that referred to previously learned content and combining interleaving with retrieval practice, which is a different desirable difficulty that benefits learning independently of interleaving (see Roelle et al., 2022). Hence, teaching scientific concepts requires the combination of various learning phases (Koedinger et al., 2012, 2013; Oser & Baeriswyl, 2001) to master complexity and to secure the acquisition of adequate principle-based cognitive skills to achieve lasting learning.

## Interleaved Learning with Verbal Materials

Learning in school is often based on written verbal materials of varying complexity, from single words to texts of varying length. One central objective of elementary education is for children to learn how to read and write. This learning process involves extensive reading and writing practice, which is essentially inductive learning (in addition to routinization of reading processes), or it contains strong inductive elements. Moreover, according to the sequential theory of attention, the learning materials may easily be designed in a way to exhibit the conditions that make interleaving effective (Carvalho & Goldstone, 2015, 2017). Many learning tasks and verbal stimuli that children encounter during literacy acquisition are characterized by high similarities between categories, which makes discriminative contrasts useful. This raises the question of whether interleaving is also a useful measure for improving literacy acquisition.

**Studies on Using Interleaving with Verbal Learning Materials**

As noted above, the meta-analysis by Brunmair and Richter (2019) suggests that interleaving with verbal materials works poorly, with a medium-sized negative interleaving effect estimated for words as learning materials. However, scrutinizing the studies that were included in the meta-analysis reveals that this conclusion might be premature. First, the effect sizes for words as learning materials was based on only three studies. Second, the studies that yielded negative or null effects of interleaving were based on materials characterized by high within-category similarity and low between-category similarity, or inductive learning was not studied (only retention of information). For example, in the study by Carpenter and Mueller (2013), which yielded a clear advantage of blocking ($g = -0.85$ in Experiment 1), non-French speakers learned grapheme-phoneme correspondences in French. The words used as learning materials exemplified pronunciation rules for words ending with -eau, -er, -e, -eux, and -t. Thus, the graphemes and phonemes relevant for each pronunciation rule were clearly distinguishable but identical within each category, which should be ideal conditions for blocking to be effective. In sum, the available studies leave open the possibility that interleaving is beneficial for inductive learning when used with verbal materials.

**How to Implement Interleaving in Literacy Instruction**

When children learn how to read and write in elementary school, the acquisition of the relevant skills invariably involves a large amount of reading and writing practice, which is essentially inductive learning and therefore a domain in which interleaved learning could be fruitfully applied. Numerous approaches to reading and spelling instruction exist in German and other alphabetic languages (Müller & Richter, 2017). Critical for literacy acquisition in such languages is a developmental phase in which children learn to decode words by applying grapheme-to-phoneme conversion rules (Ehri, 2005; Frith, 1986).

In line with this development, the most prevalent and most effective approach for teaching *reading* to beginning readers is phonics instruction, which focuses on systematically teaching and practicing grapheme-phoneme correspondence rules (for meta-analytic results, see Ehri et al., 2001; Galuschka et al., 2014). In German, didactic approaches for reading instruction exist that contain systematic, well-reasoned and partly research-based guidelines on the order in which letters (and letter combinations) and the corresponding phonological structures should be introduced (e.g., see Dummer-Smoch & Hackethal, 2007). However, no systematic research and evidence-based guidelines exist on the best way to design the "micro-sequencing" of words practiced within units. This lack is precisely where interleaved learning, that is, interleaving words exemplifying different (but similar) grapheme-phoneme correspondence rules, could offer a distinct advantage over a blocked presentation, which is often implemented in practice materials.

However, many children may not spontaneously engage in comparisons and thus might profit from instructional support to carry out the processes of comparing and distinguishing between grapheme-phoneme correspondence rules, which should underly the effectiveness of interleaved learning (Birnbaum, 2013). Like in mathematics learning, comparison and self-explanation prompts seem to be an easy and straightforward way to foster such processes, especially for weaker readers. *Comparison prompts* can remind learners to look for distinguishing features between words, letters, or syllables, exemplifying different grapheme-phoneme correspondence rules. To our knowledge, comparison prompts have been applied and evaluated only in reading instruction for children with special needs (e.g., Bird et al., 2000).

Nevertheless, there is good reason to assume that they would also be effective in other groups of beginning readers. *Self-explanation prompts* are cognitively more demanding. Such prompts would invite children to explain why a certain grapheme-phoneme correspondence rule applies to a written stimulus or groups of written stimuli (e.g., how self-explanations have been implemented in category learning, Edwards et al., 2019). With increasing skill level, these prompts could be adaptively abated during a learning session, to facilitate routinization of phonological recoding.

In a similar way, *spelling instruction* could profit from an interleaved presentation of words during practice phases. Spelling and reading draw on the same cognitive resources, and the development of the associated abilities is closely intertwined (e.g., Frith, 1986; Houghton & Zorzi, 2003). Parallel to learning how to read, learning how to spell also starts with associations of phonemes and graphemes (phoneme-grapheme associations), but for languages with a highly transparent orthography (e.g., German), not all words can be spelled correctly based on their phonology. Therefore, children need to learn spelling rules in addition to phoneme-grapheme associations, they need to understand that words are composed of morphemes that partly govern regularities of spelling, and they need to acquire representations of orthographic word forms through reading and writing practice (Treiman, 2008; Treiman & Kessler, 2014). Interleaved learning is a promising approach for improving the effectiveness and efficiency of the intensive spelling exercises that accompany this learning process, on all levels of spelling development. Instead of practicing a single phoneme-grapheme correspondence rule repeatedly, for example, exercises could involve sequences of words that exemplify different but easy to confuse rules such as words ending with *-er* and *-a*, which are difficult to distinguish phonologically in German. Likewise, spelling rules that are easily confused could be practiced in an interleaved fashion. Again, comparison prompts and self-explanation prompts could be used to support the comparison processes that supposedly underlie the effectiveness of interleaving. To the best of our knowledge, no studies to date have systematically examined the use of interleaving or the use of comparison and self-explanation prompts in spelling instruction.

## Conclusion and Open Questions

This review and discussion of existing research underscores the potential of interleaving exemplars for school learning in subject areas as diverse as mathematics, science, and literacy instruction. Interleaving is an instructional measure that can be applied easily in all kinds of inductive learning settings in which students acquire conceptual knowledge by studying exemplars, case studies, single observations, or by working on tasks exemplifying general principles. Compared to a blocked presentation of content, interleaving incurs no extra costs on teachers or students because it requires no additional learning time or additional materials. However, the discussion of existing studies also suggests that important research questions are still open. To date, the available evidence supporting the efficacy of interleaved learning is strongest (but far from exhaustive) for mathematics learning, emerging (but still weak) for science learning, and almost nonexistent for learning with verbal materials. Therefore, at this point, interleaving should be applied in educational contexts with several caveats in mind.

First, an important limitation to note is that the benefits of interleaving seem to be confined to learning settings that involve an inductive element, or at least, the benefits of interleaving are limited to learning situations in which comparisons between tasks or examples are informative and consequently add value to learning. Learning is not enhanced when simply mixing unrelated topics (e.g., Hausman & Kornell, 2014).

Second, according to relevant theories of interleaving (Birnbaum et al., 2013; Carvalho & Goldstone, 2015, 2017), the effectiveness of interleaved learning is based on the execution of comparison processes between exemplars belonging to different categories or exemplifying different principles. Therefore, an interleaved presentation can only be expected to promote learning, if between-category comparisons of exemplars are informative and contribute to abstracting conceptual knowledge. If categories are defined through commonalities of exemplars and concepts can be better understood by grasping such commonalities, a blocked presentation might be more effective.

Third, given the complex and demanding nature of learning content presented in educational contexts, students might not always be expected to perform the comparison processes necessary to abstract the underlying concepts (see also Richter et al., 2022). For example, Gentner's structure mapping theory of analogical learning states that comparison processes required to learn a concept from studying exemplars vary in complexity (Gentner & Markman, 1997). If the underlying conceptual structure is based on surface features (such as single feature or object present), as is the case in many studies on interleaved learning with visual stimuli (see Brunmair & Richter, 2019, for an overview), interleaving is likely to be effective without further instructional support. In contrast, if the underlying conceptual structure is based on commonalities in the relational structure of features, learners need to draw an analogy to learn the underlying concept. In that case, instructional support that guides students' analogical reasoning might be necessary for the advantages of interleaving to emerge. Consider, for example, the similarities of the two arithmetic expressions 1:3 and 3:9 (see Gentner & Markman, 1997, p. 47). If the underlying concept was "Contains the number 3," focusing on whether an expression contains the number 3 would be sufficient during inductive learning. In that case, interleaving expressions that contain the number 3 and omitting expressions that do not contain the number 3 might be sufficient to foster inductive learning. However, if the underlying category is "The ratio 1:3," students should receive instructional support that guides their attention to the relation of the two numbers. The same holds for other concepts that are characterized by their relational structure, including causal relations that are often constitutive of scientific concepts.

In sum, targeted instructional measures that facilitate comparisons and specific instructional measures, such as comparison or self-explanation prompts, seem to be a promising way to assist students in comparison processes with more complex concepts that require analogical reasoning. The acquisition of such concepts is characteristic for mathematics and science learning as well as learning from text. On a related note, the application of interleaving in science, abstraction of concepts from exemplars, or solving problems often presupposes extensive prior knowledge, implying that instruction should be designed such that the relevant prior knowledge is provided before inductive learning phases are implemented.

We emphasize that to date, all recommendations of how to best implement interleaving in the classroom stand on shaky empirical grounds. Research on interleaving in educational contexts and with curriculum-based materials is scarce. Among other factors, the role of learner motivation in the effectiveness of interleaving is still largely unresolved—clearly a research gap that stands in the way of implementing interleaving in the classroom. Several studies have shown that learners judge blocking to be more effective than interleaving and would rather choose a blocked schedule when they have to decide between hypothetical scenarios, although this judgement is at odds with their actual learning gains (e.g., Kornell & Bjork, 2008; McCabe, 2011; Yan et al., 2016; Yan et al., 2017). This metacognitive illusion of blocking being superior to interleaving is difficult to mend, even if learners are informed about the benefits of interleaved practice (Yan et al., 2016). Thus, the question arises how students can be motivated to interleave their practice schedule when being informed about its effectiveness is not sufficient. In contrast to these studies, Kornell and Vaughn (2018) have shown that learners block as well as interleave categories when they can determine the order of the categories during the study phase by themselves. Lu et al. (2021) provide a deeper insight into the conditions under which learners are more likely to block or interleave in self-regulated learning. In their study, participants were allowed to choose the next exemplar to be presented during real-time-studying. The results show that students studying high-similarity categories switch more often between categories than students learning low-similarity categories. These findings provide evidence that learners do not actually plan their practice schedule beforehand, but flexibly adapt their schedule during learning depending on the similarity of the categories.

However, it is unclear whether these findings generalize across learning situations, materials, and contexts. Still, it is plausible to assume that interleaving requires more cognitive effort during learning, implying that interleaving categories in self-regulated learning can be positively influenced by motivating learners to invest this effort during learning.

With few exceptions in the areas of mathematics and science learning (e.g., Nemeth et al., 2019, 2021; Sana & Yan, 2022; Ziegler & Stern, 2014, 2016), the extant studies have not fully addressed the extent that interleaving can be used to foster lasting learning (Richter et al., 2022). There are hints, though, that interleaving might qualify as a desirable difficulty (Bjork & Bjork, 2011) and might contribute to lasting knowledge that prepares students for future learning. Future research should address this issue by focusing on experimental classroom studies with assessments of both short-term and long-term outcomes.

**References**

Baroody, A., J., & Dowker, A. (2003). *The development of arithmetic concepts and skills: Constructing adaptive expertise*. Lawrence Erlbaum Associates.

Begolli, K. N., & Richland, L. E. (2016). Teaching mathematics by comparison: Analog visibility as a double-edged sword. *Journal of Educational Psychology, 108*(2), 194–213. https://doi.org/10.1037/edu0000056

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

Bird, E. K. R., Cleave, P. L., & McConnell, L. (2000). Reading and phonological awareness in children with Down syndrome: A longitudinal study. *American Journal of Speech-Language Pathology, 9*(4), 319-330.

Bjork, E. L., & Bjork, R. A. (2011). Making things hard on yourself, but in a good way: Creating desirable difficulties to enhance learning. In M. A. Gernsbacher, R. W. Pew, L. M. Hough, & J. R. Pomerantz (Eds.), *Psychology and the real world: Essays illustrating fundamental contributions to society* (pp. 56–64). Worth.

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

Carpenter, S. K., & Mueller, F. E. (2013). The effects of interleaving versus blocking on foreign language pronunciation learning. *Memory & Cognition, 41*(5), 671–682. https://doi.org/10.3758/s13421-012-0291-4

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

Catrambone, R., & Holyoak, K. J. (1989). Overcoming contextual limitations on problem-solving transfer. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*(6), 1147–1156. https://doi.org/10.1037/0278-7393.15.6.1147

Chalmers, A. F. (1999). *What is this thing called science?* (3rd ed.). Open University Press.

Chi, M. T. H. (2000). Self-explaining: The dual processes of generating inference and repairing mental models. In R. Glaser (Ed.), *Advances in instructional psychology: Educational design and cognitive science* (Vol. 5, pp. 161–238). Lawrence Erlbaum Associates.

Chi, M. T. H., Leeuw, N., Chiu, M.-H., & Lavancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science, 18*(3), 439–477. https://doi.org/10.1207/s15516709cog1803_3

DeKeyser, R. M. (1995). Learning second language grammar rules: An experiment with a miniature linguistic system. *Studies in Second Language Acquisition, 17*(3), 379–410. https://doi.org/10.1017/S027226310001425X

Dummer-Smoch, L., & Hackethal, R. (2007). *Kieler Leseaufbau* [Kiel reading build-up]. Veris.

Edwards, B. J., Williams, J. J., Gentner, D., & Lombrozo, T. (2019). Explanation recruits comparison in a category-learning task. *Cognition, 185*, 21–38. https://doi.org/10.1016/j.cognition.2018.12.011

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://10.1016/j.jarmac.2017.07.005

Ehri, L. C. (2005). Development of sight word reading: Phases and findings. In M. J. Snowling & C. Hulme (Eds.), *The science of reading: A handbook* (pp. 135–154). Blackwell.

Ehri, L. C., Nunes, S. R., Stahl, S. A., & Willows, D. M. (2001). Systematic phonics instruction helps students learn to read: Evidence from the National Reading Panel's meta-analysis. *Review of Educational Research, 71*(3), 393–447. https://doi.org/10.3102/00346543071003393

Frith, U. (1986). A developmental framework for developmental dyslexia. *Annals of Dyslexia, 36*(1), 67–81. https://doi.org/10.1007/BF02648022

Galuschka, K., Ise, E., Krick, K., & Schulte-Körne, G. (2014). Effectiveness of treatment approaches for children and adolescents with reading disabilities: A meta-analysis of randomized controlled trials. *PloS One, 9*(2), e89900. https://doi.org/10.1371/journal.pone.0089900

Gentner, D., & Markman, A. B. (1997). Structure mapping in analogy and similarity. *American Psychologist, 52*(1), 45–56. https://doi.org/10.1037/0003-066X.52.1.45

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

Heinze, A., Star, J. R., & Verschaffel, L. (2009). Flexible and adaptive use of strategies and representations in mathematics education. *ZDM, 41*(5), 535–540. https://doi.org/10.1007/s11858-009-0214-4

Herron, C., & Tomasello, M. (1992). Acquiring grammatical structures by guided induction. *The French Review*, *65*(5), 708-718. https://www.jstor.org/stable/395311

Houghton, G., & Zorzi, M. (2003). Normal and impaired spelling in a connectionist dual-route architecture. *Cognitive Neuropsychology, 20*(2), 115–162. https://doi.org/10.1080/02643290242000871

Huitt, W. G., Monetti, D. M., & Hummel, J. H. (2009). Direct approach to instruction. In C. M. Reigeluth & A. A. Carr-Chellman (Eds.), *Instructional-design theories and models, Vol. III* (pp. 85-110). Routledge.

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). New York: Routledge.

Kang, S. H., & 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

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

Koedinger, K. R., Corbett, A. T., & Perfetti, C. (2012). The Knowledge-Learning-Instruction framework: Bridging the science-practice chasm to enhance robust student learning. *Cognitive Science, 36*(5), 757–798. https://doi.org/10.1111/j.1551-6709.2012.01245.x

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., & Vaughn, K. E. (2018). In inductive category learning, people simultaneously block and space their studying using a strategy of being thorough and fair. *Archives of Scientific Psychology, 6*(1), 138–147. https://doi.org/10.1037/arc0000042

Lederman, N. G. (2007). Nature of science: Past, present, and future. In S. K. Abell & N. G. Lederman (Eds.), *Handbook of research on science education* (pp. 831–880). Erlbaum.

Lu, X., Penney, T.B. & Kang, S.H.K. (2021). Category similarity affects study choices in self-regulated learning. *Memory & Cognition, 49*(3), 67–82. https://doi.org/10.3758/s13421-020-01074-w

Maloney, D. P., O'Kuma, T. L., Hieggelke, C. J., & Van Heuvelen, A. (2001). Surveying students' conceptual knowledge of electricity and magnetism. *American Journal of Physics, 69* (7/Supp1), 12–23. https://doi.org/10.1119/1.1371296

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

Metcalfe, J., & Xu, J. (2016). People mind wander more during massed than spaced inductive learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 42*(6), 978–984. https://doi.org/10.1037/xlm0000216

Müller, B., & Richter, T. (2017). Förderung hierarchieniedriger Leseprozesse [Promoting basic reading processes]. In M. Philipp (Ed.), *Handbuch Schriftspracherwerb und weiterführendes Lesen und Schreiben* (pp. 236-251). Beltz.

Müller, B., Richter, T., & Karageorgos, P. (2020). Syllable-based reading improvement: Effects on word reading and reading comprehension in Grade 2. *Learning and Instruction*, *66,* 101304. https://doi.org/10.1016/j.learninstruc.2020.101304

Nakata, T., & Suzuki, Y. (2019). Mixing grammar exercises facilitates long-term retention: Effects of blocking, interleaving, and increasing practice. *Modern Language Journal, 103*(3), 629–647.

Nemeth, L., Werker, K., Arend, J., & Lipowsky, F. (2021). Fostering the acquisition of subtraction strategies with interleaved practice: An intervention study with German third graders. *Learning and Instruction, 71*, 101354. https://doi.org/10.1016/j.learninstruc.2020.101354

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*, 86. https://doi.org/10.3389/fpsyg.2019.00086

Oser, F. K., & Baeriswyl, F. J. (2001). Choreographies of teaching: Bridging instruction to learning. In V. Richardson (Ed.), *Handbook of research on teaching* (4th ed., pp. 1031–1065). American Educational Research Association.

Paris, S. G., Lipson, M., & Wixson, K. (1983). Becoming a strategic reader. *Contemporary Educational Psychology, 8*(3), 293–316. https://doi.org/10.1016/0361-476X(83)90018-8

Prince, M., & Felder, R. (2007). The many faces of inductive teaching and learning. *Journal of College Science Teaching, 36*(5), 14.

Renkl, A. (2015). Different roads lead to Rome: The case of principle-based cognitive skills. *Learning: Research and Practice, 1*(1), 79–90. https://doi.org/10.1080/23735082.2015.994255

Richter, T., Berger, R., Ebersbach, M., Eitel, A., Endres, T., Borromeo Ferri, R., Hänze, M., Lachner, A., Leutner, D., Lipowsky, F., Nemeth, L., Renkl, A., Roelle, J., Rummer R., Scheiter, K., Schweppe J., von Aufschnaiter, C., & Vorholzer, A. (in press). How to promote lasting learning in schools: Theoretical approaches and an agenda for research. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie/German Journal of Developmental Psychology and Educational Psychology.*

Rittle-Johnson, B., & Star, J. R. (2007). Does comparing solution methods facilitate conceptual and procedural knowledge? An experimental study on learning to solve equations. *Journal of Educational Psychology, 99*(3), 561–574. https://doi.org/10.1037/0022-0663.99.3.561

Rittle-Johnson, B., & Star, J. R. (2009). Compared with what? The effects of different comparisons on conceptual knowledge and procedural flexibility for equation solving. *Journal of Educational Psychology, 101*(3), 529–544. https://doi.org/10.1037/a0014224

Rittle-Johnson, B., Loehr, A. M., & Durkin, K. (2017). Promoting self-explanation to improve mathematics learning: A meta-analysis and instructional design principles. *ZDM, 49*(4), 599–611. https://doi.org/10.1007/s11858-017-0834-z

Roelle, J., Schweppe, J., Endres, T., Lachner, A., von Aufschnaiter, C., Renkl, A., Eitel, A., Leutner, D., Rummer, R., Scheiter, K., & Vorholzer, A. (2022). Combining retrieval practice and generative learning in educational contexts: Promises and challenges. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*. **[Platzhalter]**

Rohrer, D., & Taylor, K. (2007). The shuffling of mathematics problems improves learning. *Instructional Science, 35*(6), 481–498. https://doi.org/10.1007/s11251-007-9015-8

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

Rohrer, D., Dedrick, R. F., & Stershic, S. (2015). Interleaved practice improves mathematics learning. *Journal of Educational Psychology, 107*(3), 900–908. https://doi.org/10.1037/edu0000001

Rohrer, D., Dedrick, R. F., Hartwig, M. K., & Cheung, C.-N. (2020). A randomized controlled trial of interleaved mathematics practice. *Journal of Educational Psychology, 112*(1), 40–52. https://doi.org/10.1037/edu0000367

Samani, J., & Pan, S. C. (2021). Interleaved practice enhances memory and problem-solving ability in undergraduate physics. *NPJ Science of Learning, 6*(1), 1–11. https://doi.org/10.1038/s41539-021-00110-x

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

Schneider, W., Tibken, C., & Richter, T. (2022). The development of metacognitive knowledge from childhood to young adulthood: Major trends and educational implications. In J. Lockman (Ed.), *Advances in Child Development and Behavior*, Vol. 63 (pp. 273–307). Elsevier. https://doi.org/10.1016/bs.acdb.2022.04.006

Skinner, B. F. (1986). Programmed instruction revisited. *Phi Delta Kappan, 68*(2), 103-10.

Sorensen, L. J., & Woltz, D. J. (2016). Blocking as a friend of induction in verbal category learning. *Memory & Cognition, 44*(7), 1000-1013. https://doi.org/10.3758/s13421-016-0615-x

Treiman, R. (2018). Teaching and learning spelling. *Child Development Perspectives, 12*(4), 235–239. https://doi.org/10.1111/cdep.12292

Treiman, R., & Kessler, B. (2014). *How children learn to spell words*. Oxford University Press.

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/s13421-010-0063-y

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., 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

Ziegler, E., & Stern, E. (2014). Delayed benefits of learning elementary algebraic transformations through contrasted comparisons. *Learning and Instruction, 33*, 131–146. https://doi.org/10.1016/j.learninstruc.2014.04.006

Ziegler, E., & Stern, E. (2016). Consistent advantages of contrasted comparisons: Algebra learning under direct instruction. *Learning and Instruction, 41*, 41–51. https://doi.org/10.1016/j.learninstruc.2015.09.006

Ziegler, E., Edelsbrunner, P. A., & Stern, E. (2018). The relative merits of explicit and implicit learning of contrasted algebra principles. *Educational Psychology Review, 30*(2), 531–558. https://doi.org/10.1007/s10648-017-9424-4