Interleaving area problems in the 4th grade classroom: What is the role of context and practice?

Rachael Todaro (rtodaro@kent.edu) & Bradley J. Morris (bmorri20@kent.edu)
Department of Educational Psychology, 405 White Hall
Kent State University, Kent, OH 44242 USA

Abstract

Typical mathematics instruction involves blocked practice across a set of conceptually similar problems. Interleaving, or practice across a set of conceptually dissimilar problems, improves learning and transfer by repeatedly reloading information and increasing discrimination of problem features. Similarly, comparing problems across different contexts highlights relevant and irrelevant knowledge. Our experiment is the first to investigate the relative effects of interleaving geometry problems and interleaving contexts. Thirty-three fourth-grade students received the same practice problems but were randomly assigned to one of three conditions: interleaved by math skill, interleaved by context, and interleaved by math skill *and* by context (i.e., hyper-interleaved). Afterward, each participant was exposed to tests assessing declarative and procedural knowledge. The results suggest that interleaving math skill within varying contexts enhances the acquisition of mathematical procedures.

Keywords: interleaving, cognitive development, mathematics instruction

Introduction

Mathematics has been subject to a broad array of interventions and techniques that could potentially improve learning, retention, and transfer of knowledge to novel contexts. One promising intervention is known as interleaved practice, in which exposures to concepts (e.g., math skills) are followed by exposures to dissimilar concepts (Rohrer, 2012). Another promising technique is presenting concepts in multiple contexts, which supports generalization (Vlach, Sandhofer, & Kornell, 2008). The purpose of the present study was to examine the effects of interleaving geometrical problems across two different, yet familiar, contexts. We investigated the effects of interleaving context, math skill, or context *and* math skill by presenting this information in blocks (i.e., the same format across a series of examples). The main hypothesis of the present study was that interleaving across both context and math skill (hereafter hyper-interleaving) produces an additive effect, which would increase learning, retention, and transfer beyond other conditions because such presentation highlights differences between examples and supports greater discrimination among math skills.

Mechanisms underlying Interleaving

Presenting math problems in an interleaved fashion improves performance outcomes because this type of presentation supports two fundamental mechanisms for successful learning: discrimination training and repeated reloading. Discrimination training involves comparing and contrasting problem features, which may leading to a higher likelihood of increased learning of concepts and procedures as well as an increased ability to transfer solution strategies to novel problems (Kang & Pashler, 2012; Rittle-Johnson & Star, 2007). Repeated reloading occurs when a student revisits the same type of problem and supports effortful recall of the information, which increases the likelihood for successful encoding (Bjork & Bjork, 2011).

Presenting interleaved problems typically consists of two components. The first component is the presentation of conceptually dissimilar problem types during the practice session (e.g., a problem about the area of a square following a problem about the area of a triangle). The second component is the distribution of those problems across multiple practice sessions. That is, the student returns to practice the interleaved problems on more than one instance (Rohrer et al, 2014). Presenting interleaved problems supports comparisons and contrasts between members of different categories (e.g., perimeter of squares vs. triangles). In this manner, comparing and contrasting perceptual and conceptual information not only promotes learning regarding how to perform each procedure, but trains the learner to discriminate which solution strategy is appropriate for each problem example (i.e., discriminative contrast; Birnbaum, Kornell, Bjork, & Bjork, 2013; Kornell & Bjork, 2008).

Initially, interleaved practice appeared to owe most of its effectiveness to spacing. Spacing is inherent in interleaved practice because there is time between each opportunity to practice concepts. This is distinct from massed practice as it allows students to forget irrelevant information between learning events, which increases the potency of encoding on subsequent presentations (Bjork and Allen, 1970; Cuddy and Jacoby, 1982). However, interleaving does not solely rely on the benefits and efficacy of spacing. Given the same amount of temporal space between each exposure, interleaved presentation produces greater gains in learning than blocked presentation (Kang & Pashler, 2012). More recent literature emphasizes the importance of repeated reloading (Bjork & Bjork, 2011), which suggests that accurate memory retrieval is enhanced when a learner must repeatedly reload specific concepts from long-term to short-term memory. In fact, interleaved practice may provide its benefit more from repeated reloading rather than to the amount of time between successful reloads (e.g. temporal spacing; Kang & Pashler, 2012).

Learning with Contexts

Learning is context-dependent (Willingham, 2009). Placing math problems in familiar contexts may not only be an effective presentation method but an additive one as it activates domain knowledge that may facilitate learning and problem solving by providing a framework in which the student can make sense of the concept (Willingham, 2009; Rittle-Johnson & Star, 2007). Context also offers the learner cues to solve a problem because it draws his/her attention to the right details and improves learning in memory retrieval, problem solving (Godden & Baddeley, 1975), and reasoning tasks (Cheng & Holyoak, 1985). The main finding across studies regarding context is that when learning and testing contexts are the same, there is an improvement in performance. These findings also suggest that if a student learns within a single context, she may fail to retrieve information outside the context (i.e., context dependency, Godden & Baddeley, 1975; Vlach et al., 2011). Distributing learning across multiple events that use multiple contexts can reduce context dependency (Rothkopf, Fisher & Billington, 1982; Smith, 1982). Additionally, learning across multiple contexts results in a greater number and variance of salient cues during learning, which may increase the likelihood of recall (Smith, 1982; Vlach et al., 2011). When multiple contexts are presented across multiple learning events and these contexts are similar, the shared contextual support leads to greater learning than does providing a single cue from one context (Thiessen & Saffran, 2003; Vlach et al., 2011).

While ours is the first study to interleave both math skill and context, Rau, Aleven, and Rummel (2013) investigated how interleaving specific dimensions of math problems may affect performance. By either interleaving fraction problems (i.e., dividing fractions) or their graphical representations (i.e., pie chart, number line), they found that interleaving problems was most beneficial to learning whereas interleaving the representations was not. In a follow-up study, Rau, Aleven, Rummel, and Pardos (2014) found that interleaving both dimensions significantly benefited learning over interleaving problems alone.

The present study investigated the relative contributions of interleaved sequencing across math skills and context on declarative knowledge (i.e., knowledge of facts), procedural knowledge (knowledge of how to choose and carry out a procedure), and transfer assessments (i.e., ability to apply knowledge in novel contexts). In addition, we investigated whether these elements produce an additive effect in which their combination produces greater learning gains than either element presented individually. To test these effects, we created a 2 x 2 factorial design as follows: math skill interleaved (interleaved math skill/blocked context), context interleaved (blocked math skill/interleaved context), hyperinterleaved (interleaved math skill/interleaved context). The authors chose to omit the fully condition (blocked math skill/blocked context) since it does not address the current research question, as both dimensions are blocked. As mentioned in the introduction, we hypothesized that the hyperinterleaved group would perform better than the math skill interleaved and context interleaved group on all assessments post-intervention (i.e., posttest, delayed posttest, transfer, and delayed transfer) due to increased discrimination training across contexts and procedures. We also predicted that the math skill interleaved group would perform better than the context interleaved condition since relevant problem features are highlighted.

Method

Participants

Thirty-seven children (15 girls) ranging from nine to ten years of age were recruited from an elementary school in Northeast Ohio and completed up to six separate sessions. Each child was enrolled in the 4th grade during the 2016-2017 school year. Four students completed the first session but were not present for either the second or third session and their data were excluded from all analyses. The remaining 33 students completed both sessions of the intervention and were subject to analyses.

Task

Participants were taught to define and solve for area of four, two-dimensional geometrical shapes: square, triangle, rectangle, and parallelogram. Although various problems resulted in the same numerical solution during pretest, intervention, post-test, transfer, delayed posttest and delayed transfer test, no problem appeared more than once. In each problem, the participants were given the necessary features of each shape to solve the problems successfully (e.g., for a triangle, the base, height, and length of the sides were given). Throughout all sessions of the experiment, the participants needed to select and carryout the appropriate math skill (e.g., solving for the area of each shape).

Design & Procedure

The current study included five sessions: a pretest, first and second intervention, a one-day delayed posttest and transfer test, and a 30-day delayed posttest and transfer test. Each of the first four sessions occurred across four consecutive days, one session per day. The fifth session occurred 30 days after the fourth session ended. All sessions occurred in the 4th graders' classroom. Each participant was randomly assigned to one of three conditions: math skill interleaved, context interleaved, or hyper-interleaved. Problems in the math skill interleaved condition were presented such that the area problems were interleaved but blocked by context. Problems in the context interleaved condition were presented such that problem contexts were interleaved but the math skill was blocked. Problems in the hyperinterleaved condition were presented such that both context *and* math skill were interleaved. See Table 1. The math skills used in the study were formulas in which to solve area (e.g., area of a square). The contexts used in the current study were "indoor maintenance" (How much carpet is needed to cover a bedroom?) and "outdoor renovations" (How many feet of

the playground is covered by mulch). The first author administered all procedures described below.

Coding Rubric

The coding rubric was created for the purpose of representing the performance expectations for all assessments. The rubric was separated into two component parts (i.e., declarative and procedural knowledge) and provided clear descriptions of criteria needed to satisfy each one. The declarative knowledge component assessed adequate shape drawing as well as accurate defining of geometrical terms (i.e., area). The rubric included acceptable answers (i.e., "space inside of a shape") ranging from one to two points and unacceptable answers (i.e., area around the shape"), which were worth zero points. For the procedural knowledge component, the following criteria were included: two points for the application of the correct procedure, one point for correct answer, and one point for correct unit notation. Inter-rater reliability was strong for the declarative and procedural components, with Cohen's kappa = .90 and .96, respectively.

Table 1: Groups were presented problems interleaved by math skill, context, or by both math skill and context.

Group	Sequence of Math Skill	Sequence of Context
Math Skill Interleaved	Interleaved	Blocked
Context Interleaved	Blocked	Interleaved
Hyperinterleaved	Interleaved	Interleaved

Session 1: Pretest

During the first session, participants were administered a pre-test in order to assess prior declarative knowledge regarding the concept of area. One question was asked regarding the definition of area. Students earned up to two points if they included key terms listed on the scoring rubric. They were also asked to draw each shape, which was worth one point. Participants were also asked to define the shapes with the possibility of earning up to two points for each shape's definition. Additionally, within the pretest the participants were given two area problems for each shape to solve, which assessed prior procedural knowledge. Each procedural problem was worth four points. As per the scoring rubric, the participants needed to demonstrate the correct procedure, correct answer, and correct notation of units (e.g., $ft.^2$). In total, participants were able to earn up to 46 points on the pretest. The participants were given 30 minutes to complete the pretest. The pretest problems did not contain a context but a box below asking for the area. The participants were expected to show their work inside the box.

Session 2: Intervention Phase 1

During the second session, students were given a supplemental packet with the definitions of concepts (e.g., height, base, area, etc.) and worked examples (Sweller, 1988) of area problems across all four shapes.

Shortly afterward, students were given a brief lesson about the area of each shape encouraging the students to identify key words such as "cover" for area, show their work in the specified boxes below each figure, and to use their supplemental packet to follow along in the lesson. The lesson lasted between 10-15 minutes.

After the lesson, participants were administered the first phase of the training packet. The entire training packet consisted of 24 area problems to be divided between three sessions of training. Two area problems of each shape were included in each training packet. The first phase of the training packet consisted of a mixture of 8 area problems across either one or both contexts depending on which condition the participants pertained. If the participants had questions, they were directed to the supplemental packet and were advised to focus on the cue words in order to solve for area of the four various shapes. Students were given minimal feedback for the duration of the intervention. The first phase of the intervention lasted approximately 30 minutes.

Session 3: Intervention Phase 2

The second intervention session was similar to the first. The students received the second set of eight area problems of four different shapes across one or both contexts. The participants were given 25-30 minutes to complete the second session of the intervention.

Session 4: Posttest and Transfer Test

The posttest was similar to the pretest but differed in numerical measurements across all four shapes in order to avoid practice effects. The first section required that the students define the concepts of area along with drawing and defining the four different shapes. Again, this section assessed declarative knowledge and was worth a possible 14 points. The participants then completed four area problems of each shape for a total of eight posttest problems. Six questions that contained novel contexts were included in the posttest, in order to assess for transfer of procedural knowledge. The purpose of including these problems was to investigate whether any of the presentation conditions would lead to increased transfer to novel contexts. The transfer problems dealt with contexts distinct from those seen previously in the experiment (baseball fields, cooking, distances riding a bike, etc.). Like the pretest, participants could earn up to four points on each posttest and transfer procedural problem provided that they demonstrate the correct procedure, answer, and unit notation. In total,

participants were able to earn up to 46 points on the posttest and 24 points on transfer problems.

Session 5: 30-day Delayed Posttest and Transfer Test
The delayed post-test and the delayed transfer test were similar to the original post-test and transfer test. The first section of the delayed posttest was identical to the one-day posttest: the participants defined, drew, and described the concepts and shapes listed. The problems on both the delayed post-test and delayed transfer test only differed slightly on the numerical measurements of the shapes and novel figures in order to avoid practice effects. In total, participants were able to earn up to 46 points on the delayed posttest and 24 points on delayed transfer problems.

Table 2: Means and Standard Deviations (N = 33)

Test	Math Skill Interleaved (*n = 12*)	Context Interleaved (*n = 10*)	Hyper-interleaved (*n = 11*)
Pretest			
Declarative	5.1 (3.0)	5.5 (2.8)	5.7 (2.8)
Procedural	5.1 (4.6)	1.7 (1.8)	4.4 (4.5)
Posttest			
Declarative	6.3 (1.6)	7.4 (2.2)	7.1 (2.2)
Procedural	18.2 (8.2)	9.3 (5.9)	20.0 (9.3)
Transfer	12.3 (6.9)	9.2 (5.8)	15.8 (9.0)
Delayed Posttest			
Declarative	7.8 (1.9)	7.0 (2.3)	7.3 (2.7)
Procedural	19.4 (11.8)	14.3 (5.8)	23.2 (5.2)
Transfer	12.9 (7.0)	11.1 (5.1)	14.8 (7.0)

Results

Table 2 displays the means and standard deviations for the pretest, posttest, transfer, delayed posttest, and delayed transfer test separated by declarative items, problems, and transfer problems. It is clear in the table that there was a great deal of variability within each group on each of the problem and transfer test scores.

Three repeated measures ANOVAs were conducted to examine the differences in number of points scored as a function of (1) 3 Interleaving Types (between; Math Skill Interleaved, Context Interleaved, and Hyperinterleaved x 3 Procedural Problems (within; Pre, Post, Delayed); (2) 3 Interleaving Types x 2 Transfer Problems (within; Transfer & Delayed Transfer); (3) 3 Interleaving Types x 3 Declarative Knowledge Problems (within; Pre, Post, Delayed).

Interleaving x Declarative Knowledge

In this analysis we used the three interleaving types as a between-subjects variable and pretest, posttest, and delayed posttest as a within subject variable. This was to examine whether or not interleaving types influenced declarative knowledge. The test of within subjects effects indicated a main effect of test, $F(2, 54) = 7.746$, $MSe = 27.033$, $p < .001$, $\eta_p^2 = .223$. Within subjects contrasts indicated that participants performed better on the posttest declarative questions compared to the pretest declarative questions, $F(1, 27) = 6.133$, $MSe = 45.633$, $p = .020$, $\eta_p^2 = .185$. Although the mean for delayed posttest declarative questions was higher than posttest, this difference was not statistically significant, $F(1, 27) = 3.381$, $MSe = 12.033$, $p = .077$, $\eta_p^2 = .111$. There was not a significant interaction between test and group from posttest to delayed posttest declarative questions, $F(2, 27) = 2.931$, $MSe = 10.433$, $p = .07$, $\eta_p^2 = .178$. Tests of between-subjects effects determined that there was not a significant effect of group, $F < 1$.

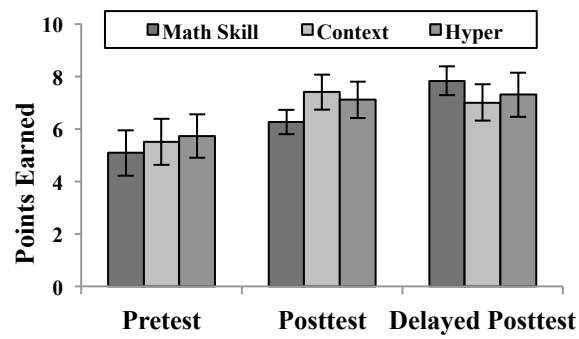

Figure 3: Mean scores of each group on declarative knowledge. Error bars indicate standard errors of the means.

Interleaving x Procedural Problems

In this analysis we used the 3 interleaving types as a between subjects variable and pretest, posttest, and delayed posttest as a within subject variable. This was to answer the question of whether or not the different interleaving types impacted the problem solving accuracy of pre, post, and delayed posttest problems. As is evident in Figure 1 regarding procedural problem accuracy, the test of within subjects effects indicated that there was a main effect of test, $F(2, 54) = 53.46$, $MSe = 2235.54$, $p < .001$, $\eta_p^2 = .66$. Within subjects contrast indicated that students performed better on the posttest problems compared to the pretest problems, $F(1, 27) = 85.19$, $MSe = 4538.70$, $p < .001$, $\eta_p^2 = .76$. Although the mean for the delayed posttest problems was higher than the posttest problems, it was not statistically significant, $F(1, 27) = 3.89$, $MSe = 407.01$, $p = .117$, $\eta_p^2 = .089$. The interaction between test and group was approaching statistical significance, $F(1, 27) = 3.157$, $MSe = 168.175$, $p = .059$, $\eta_p^2 = .190$.

The test of between-subjects effects determined that there was a significant main effect of group, $F(1, 27) = 5.209$, $MSe = 136.608$, $p = .012$, $\eta_p^2 = .278$. A Bonferroni test of multiple comparisons revealed that participants in the hyperinterleaved condition performed significantly better

from posttest to delayed posttest procedural problems when compared to the context interleaved condition, $p = .011$, whereas the math skill interleaved group did not, $p = .108$. The hyperinterleaved condition did perform better than the math skill interleaved condition on solving procedural problems from post- to delayed post test, however, these findings were not statistically significant, $p > .088$.

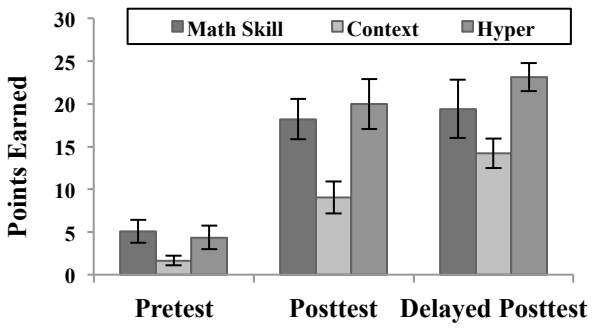

Figure 1: Mean scores of each group on procedural knowledge. Error bars indicate standard errors of the means.

Interleaving x Transfer

In this analysis we used the 3 interleaving types as a between subjects variable and transfer and delayed transfer test as a within subject variable. This was to answer the question of whether or not the different interleaving types impacted the participants' problem accuracy of transfer and delayed transfer test. The test of within subjects effects did not find a main effect of test, $F(1, 28) < 1$, $MSe = 1.492$, $p = .805$, $\eta_p^2 = .040$. Although represented in Figure 2, the context and math skill interleaved groups encountered a rise in performance from transfer to delayed transfer problems, within subjects contrasts revealed that the interaction effects were not significant, $F(1, 28) < 1$, $MSe = 13.981$, $p = .566$, $\eta_p^2 = .040$. While the hyperinterleaved group performed best on both transfer tests, tests of between-subjects effects determined that there was not a significant effect of group, $F(1, 28) = 2.25$, $MSe = 127.801$, $p = .124$, $\eta_p^2 = .138$.

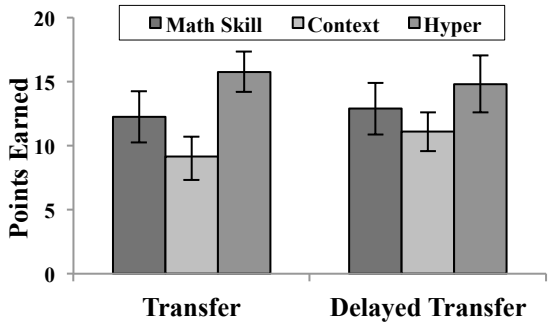

Figure 2: Mean scores of each group on transfer of procedural knowledge to novel contexts. Error bars indicate standard errors of the means.

Overall, all groups demonstrated significant learning of procedural problems from pretest to posttest. In fact, there was a main effect of group on the posttest procedural problems in which the hyperinterleaved group performed significantly better than the context interleaved group and the math skill interleaved group did not. However, the result pattern could be due to lower pretest scores for the context interleaved group.

To further examine the effect of group on assessments, change scores were computed and a one-way ANOVA was conducted. The ANOVA demonstrated that change scores from pretest to posttest of the hyperinterleaved group significantly differed from the context interleaved group, $F(2, 28) = 3.301$, $MSe = 169.563$, $p = .05$. Bonferroni tests of multiple comparisons indicated this difference was not statistically significant, $p = .07$. Change scores from pretest to delayed posttest and from posttest to delayed posttest indicated no significant differences between groups, ps $> .05$. Furthermore, there were no differences between groups in changes scores from transfer to delayed transfer, $p > .05$.

Discussion

Our results demonstrate that when math skill was interleaved (i.e., in the math skill interleaved and hyperinterleaved groups), procedural performance on posttest was significantly better than when math skill was blocked (i.e., context interleaved group). These findings provide additional support for interleaved practice as a technique that enhances memory by increasing the number of repeated reloads and by promoting discriminative contrast among problems. Recall that the context interleaved group blocked the math skill problems and interleaved the contexts. Blocking math skills does not allow the learner space between problems to reload relevant information. Additionally, blocking these problems does not allow the learner to discriminate between the features of other shapes in order to highlight key elements within the problem in order to apply the appropriate procedure.

The lack of statistical differences between the math skill interleaved and the hyperinterleaved group on posttest may be due to the lack of variation between contexts. Recall that practicing problems in multiple, varying contexts typically reduces context dependency, which supports generalization to novel situations. In the current study, the contexts may not have been different enough to decrease the level of context dependency. And, a stronger effect of hyperinterleaving may have been observed if more than two contexts were included in this study. Alternatively, the spacing of practice across two separate sessions of intervention may have equalized the effects of interleaving math skill and hyperinterleaving.

The results of our experiment align with those of Rau et al. (2013; 2014), suggesting two notions. One, math skill is the problem dimension that benefits most from an interleaved practice sequence. Two, interleaving both math skill and another dimension (i.e., context) may enhance learning when compared to interleaving math skill alone. It is important to note that the hyperinterleaved group demonstrated increased learning on all post, delayed posttest, and transfer procedural problems. One explanation

for these findings may be that interleaving context while also interleaving math skill may require more effort on the part of the learner during practice, resulting in enhanced memory. Another explanation is that shuffling familiar contexts during practice may facilitate the application of knowledge outside of the context in which it was learned. This experiment provided an important contribution in understanding the benefits of more effortful interleaved practice when learning new skills and transferring them to novel contexts.

Although the fourth-grade sample size in the current study was small, the apparent trend of interleaving math skills within different contexts that led to better performance seems to be promising. The results suggest that early learning of math skills such as solving for area may benefit from all types of practice, especially when spaced over time. For future research, it may be beneficial to examine the effect of context versus no context in interleaving experiments that evaluate retention and transfer of declarative and procedural knowledge.

Conclusion

Our experiment demonstrated the potential educational benefits of hyperinterleaving math skill with contexts. The results of the current study suggest that along with the advantages that interleaving area problems offers, shuffling contexts throughout this practice may also contribute to better generalization of these skills. Our study is one of the first to examine the effectiveness of combining interleaved practice with another common instructional technique. Placing examples in a familiar context is often used in classroom settings to make learning tasks recognizable for young learners and this study has provided unique insight on how this technique interacts with that of interleaved practice. More research is necessary to understand how interleaved practice interacts with context and other effective learning techniques, especially within the classroom environment.

Acknowledgements

A special thanks goes to the fourth graders at Willyard Elementary, Caitlin Fortson, Jennifer Roche, Christopher Was, Tracy Arner, and Colleen Dragovich.

References

Bjork, R.A., Allen, T.W. (1970). The spacing effect: consolidation or differential encoding? *Journal of Verbal Learning and Verbal Behavior, 9* 567-572.

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

Cheng, P.W., & Holyoak, K.J. (1985). Pragmatic reasoning schemas, *Cognitive Psychology, 17,* 391-416.

Cuddy, L.J., Jacoby, L.L. (1982). When forgetting helps memory: an analysis of repetition effects. *Journal of Verbal Learning and Verbal Behavior, 21,* 451-467.

Godden, D.R., & Baddeley, A.D. (1975). Context-dependent memory in two natural environments: On land and under water. *British Journal of Psychology, 66,* 325-331.

Kang, S.H.K., Pashler, H. (2012). Learning painting styles: Spacing is advantageous when it promotes discriminative contrast. *Applied Cognitive Psychology, 26,* 97-103.

Kornell, N., Bjork, R.A. (2008). Learning concepts and categories: Is spacing the "enemy of induction"? *Psycholgical Science, 19,* 585-592.

Rau, M.A., Aleven, V., Rummel, N. (2013). Interleaved practice in multidimensional learning tasks: Which dimension should we interleave? *Learning and Instruction, 21,* 98-114.

Rau, M.A, Rau, M.A., Aleven, V., Rummel, N., Pardos, Z. How should intelligent tutoring systems sequence multiple graphical representations? A mutli-methods study. *International Journal of Artificial Intelligence in Education, 24,* 125-161.

Rittle-Johnson, B., Star, J. (2007) Does comparing solution methods facilitate conceptual and procedural knowledge? An experimental study on learning to solve equations. *Journal of Educational Psychology, 99,* 561-574.

Rohrer, D. (2012) Interleaving helps students distinguish among similar concepts. *Educational Psychology Review, 24,* 355-367.

Rohrer, D., Dedrick, R.F., Burgess, K. (2014). The benefit of interleaved mathematics practice is not limited to superficially similar kinds of problems. *Psychonomics Bulletin and Review, 21,* 1323-1330.

Rothkopf, E., Fisher, D., & Billington, M. (1982). Effects of spatial context during acquisition on the recall of attributive information. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 8,* 126-138.

Smith, S.M. (1982). Enhancement of recall using multiple environmental contexts during learning. *Memory & Cognition, 10,* 405-412.

Sweller, J. (1988). Cognitive load during problem solving, *Cognitive Science, 12,* 257-285.

Theissen, E.D., & Saffran, J.R. (2003). When cues collide: Use of stress and statistical cues to word boundaries by 7-to-9 month old infants. *Developmental Psychology, 39,* 484-494.

Vlach, H.A., Sandhofer, C.M., & Kornell, N. (2008). The spacing effect in children's memory and category induction. *Cognition, 109,* 163-167.

Vlach, H.A., Sandhofer, C.M. (2011). Developmental differences in children's context-dependent word learning. *Journal of Experimental Child Psychology, 108,* 394-401.

Willingham, D.T. (2009). *Why Don't Students Like School?* San Fransisco: Jossey-Bass.