Research Article

Psychological Science
2022, Vol. 33(5) 782–788

DOI: 10.1177/09567976211057507
www.psychologicalscience.org/PS
SAGE

Interleaving Retrieval Practice Promotes Science Learning

Faria Sana[1] and Veronica X. Yan[2]
[1]Centre for Social Sciences, Athabasca University, and [2]Department of Educational Psychology, The University of Texas at Austin

Abstract

Can interleaved retrieval practice enhance learning in classrooms? Across a 4-week period, ninth- through 12th-grade students (N = 155) took a weekly quiz in their science courses that tested half of the concepts taught that week. Questions on each quiz were either blocked by concept or interleaved with different concepts. A month after the final quiz, students were tested on the concepts covered in the 4-week period. Replicating the retrieval-practice effect, results showed that participants performed better on concepts that had been on blocked quizzes (M = 54%, SD = 28%) than on concepts that had not been quizzed (M = 47%, SD = 20%; d = 0.30). Interleaved quizzes led to even greater benefits: Participants performed better on concepts that had been on interleaved quizzes (M = 63%, SD = 26%) than on concepts that had been on blocked quizzes (d = 0.35). These results demonstrate a cost-effective strategy to promote classroom learning.

Keywords

interleaving, retrieval practice, classroom learning, open data, open materials

Received 6/8/21; Revision accepted 9/21/21

Conceptual learning is the backbone of education. Yet one of the challenges that science teachers face is promoting conceptual understanding rather than simple memorization. That is, the goal is to have students not only learn definitions of new terms but also be able to understand the underlying idea, to abstract general principles, and to apply these principles across superficially different situations. For example, students in physics must not only learn the different types of circuits; they must also understand how a series circuit is similar to and different from a parallel circuit and identify when each of these circuits applies in new problems.

In the present study, we leveraged two principles from learning science—retrieval practice and interleaving—to examine whether a simple manipulation has the potential to increase conceptual understanding: brief, weekly quizzes with interleaved concepts. The benefits of retrieval practice are well documented (Yang et al., 2021). In comparison with restudying, the act of effortfully bringing previously taught information to mind strengthens that learning, organizes knowledge, and makes the knowledge more easily retrievable and transferable in the future (Karpicke & Blunt, 2011; McDaniel et al., 2013; Roediger et al., 2011). Retrieval practice can take many forms (frequent quizzing, brain dumps, teaching other people) and can easily be implemented both in and out of the classroom. In the present study, we focused on in-class quizzing.

Research on the interleaving effect is relatively recent, but studies have reliably demonstrated a striking and counterintuitive finding—namely, that practicing problems (e.g., 1, 2, 3) of related concepts (e.g., A, B, C) in a mixed-up order ($A_1C_1B_1C_2B_2A_2B_3C_3A_3$; interleaved) can lead to better concept learning than practicing problems one concept at a time ($A_1A_2A_3B_1B_2B_3C_1C_2C_3$; blocked; Brunmair & Richter, 2019; Kang, 2017). One of the leading explanations for the interleaving benefit is that when

Corresponding Authors:
Faria Sana, Athabasca University, Centre for Social Sciences
Email: fsana@athabascau.ca

Veronica X. Yan, The University of Texas at Austin, Department of Educational Psychology
Email: veronicayan@austin.utexas.edu

problems of different concepts are juxtaposed, as is the case with interleaved sequences, learners' attention is drawn to the features that help discriminate between the concepts (Carvalho & Goldstone, 2017; Kang & Pashler, 2012). In contrast, when problems are blocked by concept, there are fewer opportunities to notice the critical features that differ between concepts. The interleaving effect has been shown to be remarkably robust in controlled laboratory settings across a range of materials from perceptual categories (e.g., artists, butterflies, radiology) to cognitive concepts (e.g., statistics, clinical diagnoses, comma rules) and across a range of age groups (e.g., elementary school, college, medical school, older adults; Brunmair & Richter, 2019; Kang, 2017).

There have been only a handful of classroom studies on interleaved retrieval practice, all of which have focused on mathematics learning (Ostrow et al., 2015; Rohrer, Dedrick, Hartwig, & Cheung, 2020; Rohrer et al., 2014, 2015). The largest classroom study to date is that of Rohrer, Dedrick, Hartwig, and Cheung (2020), which showed a large benefit of interleaving (Cohen's $d = 0.83$) across 54 classrooms. In this study, seventh-grade students completed worksheets on which questions were either interleaved or blocked by concept across 4 months, followed by an interleaved review worksheet and then a final test 1 month later. However, interleaving was manipulated together with spacing. That is, not only were practice problems of a given concept interleaved with the practice problems of other concepts, but they were also distributed across every worksheet (eight worksheets across 103 days). In contrast, in the comparison group, practice problems of a given concept were both blocked (practiced consecutively without any other intervening concept) and massed (practiced in just one worksheet on a single day within the 103-day time frame). This confound poses a potential theoretical problem: Did improved learning occur because of the mixing of different concepts, the spacing of problems from the same concept across months, or both? Moreover, it is unclear whether effects found for mathematics will generalize to science-concept learning. In each of these studies, the target skill was mathematics problem solving, which consists of both conceptual and procedural components—hence, benefits could arise from greater practice in using formulae, deepening conceptual understanding, or a combination of both.

Clearly, interleaved retrieval practice is a promising strategy to promote learning. And yet blocked instruction dominates academic programs and materials (Rohrer, Dedrick, & Hartwig, 2020). Although students often receive opportunities to practice applying what they learn, these opportunities often do not involve retrieval (in the case of homework, where they can refer to their notes) and often are blocked (i.e., practice on the one concept that was just learned), or they are treated as high-stakes summative assessments (e.g., final exams) rather than low-stakes, formative, learning opportunities. What are some barriers in scaling interleaved retrieval practice from the lab to the classroom? First, there is a lack of empirical classroom-based evidence demonstrating the benefit of the interleaving, beyond mathematics problem-solving, and separate from the spacing. Second, given the dearth of classroom studies, there is also a lack of evidence for the impact of interleaved retrieval practice across educationally relevant retention intervals. Hence, in the present study, we examined whether interleaved practice of science concepts through brief end-of-week, in-class quizzes can deepen and sustain students' conceptual learning.

Statement of Relevance

Conceptual learning is the backbone of education. Yet one of the challenges that science teachers face is promoting long-lasting conceptual understanding (being able to understand underlying principles and apply them to new situations) rather than simple memorization. Students might appear to understand concepts in one class, only to forget by the next. To address this challenge, we leveraged two powerful principles from the science-of-learning research in cognitive psychology: interleaving and retrieval practice. In high school science classrooms, students were given a 10- to 12-min weekly quiz on the concepts taught each week. We found that just the addition of this brief quiz significantly enhanced performance on a test of the concepts over 1 month later. Importantly, when questions about different concepts on the quiz were presented in an interleaved (mixed up) order, learning gains were even greater than when questions were blocked by concept. This shows that a small addition to classroom practice—one that does not require additional teacher training, is cost effective, and is not time intensive—can yield powerful, long-lasting effects on student learning.

Method

Participants

We recruited students from eight science classrooms ($N = 155$ students), ranging from ninth to 12th grades, from a midsize Canadian public high school in southern Ontario. Students were given a $3 Tim Hortons gift card

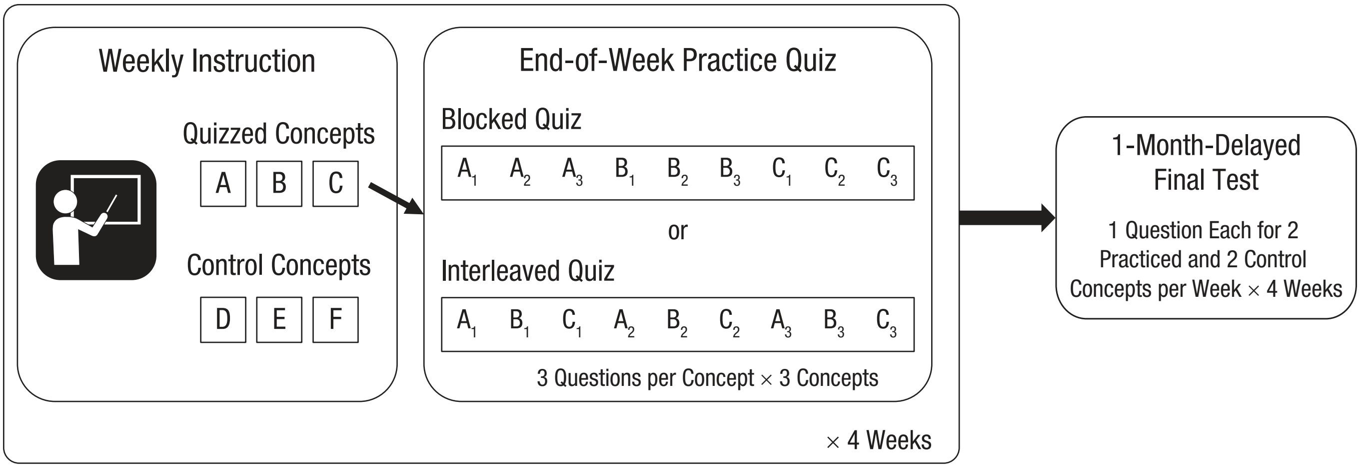

Fig. 1. General design of the study. For four weeks, students were quizzed at the end of each week about some of the concepts they had learned that week. Quiz questions were ordered such that the same concept was either tested consecutively (blocked) or intermixed with the other concepts (interleaved). All students experienced two blocked quizzes and two interleaved quizzes. Students were not quizzed on an additional set of control concepts. A final test was administered 1 month after this 4-week instructional period.

if they completed the consent form, regardless of whether they chose to participate in the study. The consent rate (assent obtained from both students and their guardians) was 100%. The school was unable to provide any other demographic information. However, according to 2016 Canadian Census data for the town, English is the mother tongue of roughly 61% of the town's population (Statistics Canada, 2017). Teachers were recruited following a workshop that provided an overview of the evidence-based strategies that promote student learning. Six teachers who taught science, biology, chemistry, and physics agreed to participate in the study, which was conducted midsemester in the winter term of 2019. Each teacher was given a $500 (Canadian) honorarium. We implemented the study in as many classrooms as we could recruit after the workshop. Power analyses using G*Power (Version 3.1; Faul et al., 2007) revealed that 52 participants would provide sufficient power to detect a conservative, medium effect size (d) of 0.40 at an α of .05 with 80% power (given a within-subjects design). The current study was approved by the Athabasca University Institutional Research Ethics Board and by the board of the high school where the study took place.

Design

The study followed a repeated measures design with three conditions: concepts that were practiced blocked, concepts that were interleaved, or concepts that were not practiced (control). Figure 1 illustrates the design of the study. The initial quizzing phase spanned 4 weeks in which the concepts taught in each week were assigned to the practiced or nonpracticed condition. Across weeks, the quizzes alternated between presenting the practice questions blocked by concept or with concepts interleaved. Each classroom was randomly assigned to start with a blocked quiz or an interleaved quiz on Week 1. There were two teachers who both taught two sections of the same class (ninth-grade science and 11th-grade biology). For these classes, one section of each pair was randomly assigned to start with a blocked quiz and the other was assigned to start with an interleaved quiz. A final test occurred 1 month after the end of the 4-week quizzing phase.

Materials

All questions were created by two graduate students who had recently received their teaching accreditation in science and who were serving as substitute teachers in a similar school district teaching the same curriculum. These questions were then vetted by the teacher of each class to ensure that they were conceptually accurate and were written appropriately for their student level. Questions were randomly selected to be on either the weekly quiz or the final test. All quizzes and the final test were administered on pen and paper and were graded by a research assistant who was blind to the study hypotheses and manipulations. Materials, data, and analyses can be accessed on OSF (https://osf.io/aqng6/).

Weekly quizzes. Because each teacher taught a different class and hence covered different material, we created four end-of-week multiple-choice quizzes for each teacher. For each of the 4 target weeks during the quizzing period (March 25, 2019–April 18, 2019), teachers identified at least six concepts that students typically find confusing. Three of these concepts were randomly

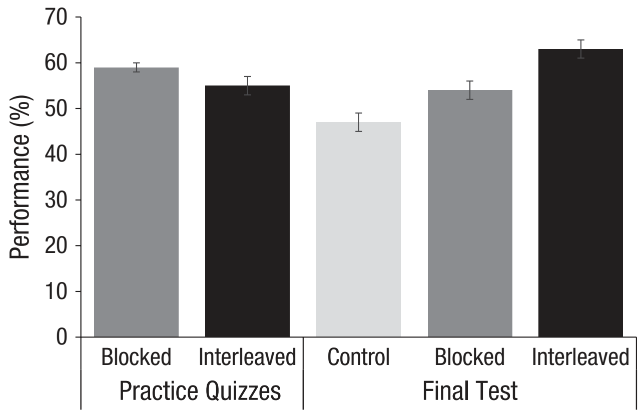

Fig. 2. Mean percentage of correct responses on blocked and interleaved practice quizzes (left) and mean percentage of correct responses in each condition (control, blocked quiz, interleaved quiz) on the final test (right). Error bars represent ±1 *SE*.

selected to appear on the end-of-week quiz (three questions per concept). The questions were ordered such that they were blocked by concept or interleaved so that no two questions on the same concept were consecutive. For half of the classes, the concepts on Quizzes 1 and 3 were interleaved and the concepts on Quizzes 2 and 4 were blocked. For the other half of classes, the concepts on Quizzes 1 and 3 were blocked and the concepts on Quizzes 2 and 4 were interleaved.

Final test. To keep the length of this final test manageable, instead of testing every concept taught in the 4-week quizzing period, we randomly selected four concepts from each week—two that were on the quiz (practiced) and two that were not (control). Hence, the final test consisted of 16 questions, which could be classified into three conditions: eight questions from the control condition, four questions that appeared on a blocked quiz, and four questions that appeared on an interleaved quiz.

Procedure

For 4 weeks, students were given these quizzes on the last class day of each week (usually Friday; occasionally Thursday if it was the last class day of the week, e.g., on the week of Good Friday). Test packets were delivered to the teachers either the day before or the morning of test administration. In any given week, all the students in one class received the same quiz. Students had 10 to 12 min to complete the quiz each week and were encouraged to answer all questions in the order in which they were presented. Although the intent was for teachers to be unaware of the sequence conditions, it was easy to guess the condition on the basis of the question order on a quiz. One month after the fourth practice quiz, students were given a surprise final test and were given 20 min to complete it.

We requested that the teachers not provide any aid to the students while they were completing the quizzes or the test and not give feedback to the students after the quizzes were completed and collected.

Results

First, we examined whether the sequencing of the practice quizzes yielded different levels of performance on the quizzes themselves. The left side of Figure 2 shows the overall performance on blocked and interleaved quizzes; Table 1 lists the details by classroom. To account for the nested nature of the data (different concepts were taught within each course-grade combination), we conducted a linear mixed-effects regression analysis, predicting practice-quiz score from condition (blocked, interleaved), with individual ID and concepts nested within course grade entered as random effects[1] (for a full summary of results, see Table S1 in the Supplemental Material available online). This analysis revealed that the scores on the interleaved quizzes were significantly lower than those on the blocked quizzes, $b = -0.07$, $SE = 0.02$, $t(819.76) = -3.43$, $p < .001$, Cohen's $d = 0.21$. In other words, the interleaved practice quizzes were more difficult than the blocked practice quizzes.

More importantly, we examined whether there were significant differences between the three conditions on the final test—the mean percentage of correct responses on the final test for each condition are represented on the right side of Figure 2 and are detailed by classroom in Table 1. We conducted a linear mixed-effects regression analysis, predicting final test score from condition (control, blocked quiz, interleaved quiz), with individual ID and concepts nested within course grade entered as random effects. The blocked-quiz condition was set as the reference condition (for a full summary of results, see Table S2 in the Supplemental Material). Results revealed two critical findings: First, the blocked-quiz condition led to significantly better final test performance than the no-quiz control condition, $b = -0.09$, $SE = 0.03$, $t(119.07) = -2.66$, $p = .009$, Cohen's $d = 0.30$. This result replicated the retrieval-practice effect. Second, the interleaved-quiz condition led to significantly better final test performance than the blocked-quiz condition, $b = 0.08$, $SE = 0.03$, $t(602.67) = 2.68$, $p = .008$, Cohen's $d = 0.35$. This result reveals an additional benefit of interleaving on top of retrieval practice.

Discussion

The current study investigated the long-term effects of interleaved retrieval practice in ninth- through 12th-grade science classrooms. We found that on a surprise test administered 1 month after the last practice quiz, students performed better on concepts that were

Table 1. Mean Percentage of Correct Responses (and Standard Deviations) on the Weekly Quizzes and Final Test, by Class and Sequence Condition

		Weekly practice quiz		1-month-delayed final test		
Classroom	Subject and grade	Blocked	Interleaved	Control	Blocked	Interleaved
1	Science, Grade 9[a]	58 (19)	57 (18)	48 (18)	55 (35)	59 (25)
2	Science, Grade 9[a]	66 (16)	48 (17)	50 (25)	55 (19)	67 (29)
3	Science, Grade 10	56 (14)	34 (14)	39 (18)	46 (24)	53 (21)
4	Biology, Grade 11[a]	69 (13)	51 (25)	45 (19)	50 (25)	59 (25)
5	Biology, Grade 11[a]	35 (14)	49 (9)	38 (22)	44 (37)	48 (31)
6	Physics, Grade 11	57 (14)	57 (11)	45 (18)	59 (22)	66 (22)
7	Chemistry, Grade 12	68 (16)	70 (18)	56 (17)	61 (24)	78 (22)
8	Physics, Grade 12	62 (11)	76 (12)	54 (19)	63 (28)	75 (25)
	Overall	59 (18)	55 (20)	47 (20)	54 (28)	63 (26)

[a]Science (Grade 9) and Biology (Grade 11) had two sections taught by different teachers; content assignment to the blocked or interleaved condition in each week was counterbalanced between sections.

quizzed compared with concepts that had not been quizzed. This finding is consistent with those of previous studies that show the learning benefits of retrieval practice (Yang et al., 2021). We also found that students performed better on concepts that appeared in the interleaved quizzes, in which the order of the questions was mixed up, than on concepts that appeared in the blocked quizzes, in which questions were organized by concept.

Only some studies have previously examined interleaved retrieval practice in classrooms, most of which included students completing weekly *cumulative* math assignments (i.e., assignments that assess concepts covered in both the current and previous weeks). Such manipulations combine interleaving and spacing benefits. In the present study, we manipulated interleaving without also manipulating spacing to show that interleaving alone can enhance student learning.

Our study builds on prior studies in two ways. First, we separated interleaving from spacing by quizzing students only once on the concepts taught in each week. Second, most of the prior studies have focused on content that either has not yet been taught by teachers (Ziegler & Stern, 2014) or is not currently being taught by the teacher (e.g., concepts learned in a prior semester or in a prior school year; Rohrer et al., 2014). In our study, the weekly quizzes were administered as the students were learning the concepts in class. In other words, the weekly quizzes were taking place in the context of other rich learning activities and study behaviors (e.g., in-class lessons, discussions, homework assignments). On one hand, this context reduced experimenter control because students were being exposed to the concepts repeatedly outside of our quizzes. On the other hand, it makes the results of our study even more surprising and meaningful. These quizzes added only 10 to 12 min of class time each week and yet led to sizable and sustained learning benefits a month later. In fact, the potential benefit of setting aside only 10 to 12 min per week is particularly striking when one considers the large difference between performance in the interleaved-retrieval-practice condition (63%) and the control condition (47%; a large effect size of d = 0.71).

Although this study addressed a gap in the literature, it has limitations. Given that the content on the quizzes was a core part of the course, we could not control how the concepts were discussed and practiced outside of our quizzes. In fact, teachers often reviewed previous concepts across subsequent weeks in their own lessons and encouraged students to space out their study sessions. It is possible that students changed how they studied following interleaved quizzes because interleaving often feels disfluent and difficult compared with blocking (e.g., Yan et al., 2016). If a student interpreted this disfluency as insufficient learning, they could have engaged in compensatory studying. To test this hypothesis, we calculated the difference between interleaved- and blocked-quiz scores for each student and then examined whether this difference score significantly moderated the interleaving benefit on the final test. If this hypothesis were supported, then we would expect that students who experienced larger blocking benefits on the practice quiz should have larger interleaving benefits on the final test because of engaging in compensatory studying. Our analysis, however, showed that the difference between blocked and interleaved quizzes had no bearing on final test performance (for details, see Table S3 in the Supplemental Material).

Another limitation is that we could not randomly assign concepts at the student level; rather, randomization of concepts to condition occurred on a weekly basis for each course-grade combination. There were two pairs of classrooms, however, in which we could counterbalance assignment of concepts to interleaved- or blocked-quiz conditions: Grade 9 science and Grade 11 biology. Although the two classes in each pairing were taught by different teachers, the curriculum and the concepts covered in each class were identical. Hence, in a given week, when one classroom from the pair was assigned a blocked quiz, the other classroom from the pair was assigned an interleaved quiz. When we restricted our analyses to just these four classrooms ($n = 72$), we found the same pattern with a similar effect size to that of the larger sample: The final test performance for the interleaved concepts (M = 58%, SD = 28%) was higher than that for the blocked concepts (M = 51%, SD = 30%; Cohen's $d = 0.27$). Moreover, given the sheer number of concepts in each class, it is unlikely that the concepts in one condition were systematically different from the concepts in the other conditions. Indeed, the heterogeneity of the grade levels, the courses, and the concepts in our study can be perceived as a strength—the interleaving benefit is robust across all these differences.

Although findings from the current study are limited to science content and to students from Grades 9 to 12, there is no reason to expect that interleaved retrieval practice will not improve conceptual learning for other student populations and in different subject areas. Indeed, separate meta-analyses on retrieval practice (Yang et al., 2021) and interleaving (Brunmair & Richter, 2019) have demonstrated robust effects across students of various ages and topic domains. The current study offers a solution to the practical question of how interleaving could be incorporated into classroom practice. Interleaving instruction can feel disorganized and chaotic and may require a significant time investment to restructure the curriculum. However, we propose that interleaving *practice*, rather than instruction, may be an easier, yet highly effective, solution. Future research in classrooms should focus on moving beyond the dichotomy of interleaved practice versus blocked practice. Given that blocking has also been shown to have benefits for directing learners to within-concept examples (e.g., in lab experiments with artificial materials; Carvalho & Goldstone, 2017) and to provide support for novice learners (e.g., in motor-skills learning; Wulf & Shea, 2002), future research should examine ways of optimizing interleaving dosage for different learners. For example, learners with lower prior knowledge may require more blocked practice before moving on to interleaved practice (but see Ostrow et al., 2015, who found greater interleaving benefits for learners with low skills).

These findings also add to the growing body of research showing the potential for learning science to impact educational practices and education outcomes. In a recent review, Kraft (2020) found a median effect size (d) of 0.17 for educational interventions using narrow measures of achievement outcomes (i.e., researcher-designed tests that are aligned with the treatment, as we had in our design). Our effect size of 0.71 (between the control and interleaved-retrieval-practice conditions) is considerably larger. There is still work to be done to further investigate whether interleaved retrieval practice can affect broad achievement measures (e.g., standardized exams) and effects on much more heterogeneous samples. However, the data are highly promising, especially in light of the fact that the intervention is simple to implement: It does not require additional teacher training, it is not time intensive, and it is cost effective. Teachers likely have existing assignments and practice problems developed for their own classes or have access to resources they can use (e.g., test banks, end-of-chapter practice problems, resources provided by publishers); the difference would be to break up longer exams into more frequent, shorter quizzes and to change the sequence in which questions are assigned, making sure that concepts are interleaved rather than blocked.

Transparency

Action Editor: M. Natasha Rajah

Editor: Patricia J. Bauer

Author Contributions

F. Sana and V. X. Yan share first authorship of this article. Both authors designed the study. F. Sana executed the study. Both authors analyzed the data, wrote the manuscript, and approved the final manuscript for submission.

Declaration of Conflicting Interests

The author(s) declared that there were no conflicts of interest with respect to the authorship or the publication of this article.

Funding

This research was funded by the Social Science and Humanities Research Council (Insight Development Grant No. 430-2017-00593).

Open Practices

All materials, data, analysis scripts, and markdown files have been made publicly available via OSF and can be accessed at https://osf.io/aqng6/. The design and analysis plans for the study were not preregistered. This article has received the badges for Open Data and Open Materials. More information about the Open Practices badges can be found at http://www.psychologicalscience.org/publications/badges.

ORCID iDs

Faria Sana https://orcid.org/0000-0002-2202-7592
Veronica X. Yan https://orcid.org/0000-0002-3988-3184

Supplemental Material

Additional supporting information can be found at http://journals.sagepub.com/doi/suppl/10.1177/09567976211057507

Note

1. There was only one set of missing data: Quiz 4 (interleaved) for one of the two Grade 11 biology classrooms.

References

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

Carvalho, P. F., & Goldstone, R. L. (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*, *42*(3), 481–495. https://doi.org/10.1037/xlm0000406

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*, 175–191.

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

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

Karpicke, J. D., & Blunt, J. R. (2011). Retrieval practice produces more learning than elaborative studying with concept mapping. *Science*, *331*(6018), 772–775. https://doi.org/10.1126/science.1199327

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

McDaniel, M. A., Thomas, R. C., Agarwal, P. K., McDermott, K. B., & Roediger, H. L. (2013). Quizzing in middle-school science: Successful transfer performance on classroom exams. *Applied Cognitive Psychology*, *27*(3), 360–372. http://doi.org/10.1002/acp.2914

Ostrow, K., Heffernan, N., Heffernan, C., & Peterson, Z. (2015). Blocking vs. interleaving: Examining single-session effects within middle school math homework. In C. Conati, N. Heffernan, A. Mitrovic, & M. Verdejo (Eds.), *Artificial intelligence in education: AIED 2015. Lecture notes in computer science* (Vol. 9112, pp. 338–347). Springer. https://doi.org/10.1007/978-3-319-19773-9_34

Roediger, H. L., III, Putnam, A. L., & Smith, M. A. (2011). Ten benefits of testing and their applications to educational practice. In J. Mestre & B. Ross (Eds.), *Psychology of learning and motivation: Cognition in education* (pp. 1–36). Elsevier.

Rohrer, D., Dedrick, R. F., & Burgess, K. (2014). The benefit of interleaved mathematics practice is not limited to superficially similar kinds of problems. *Psychonomic Bulletin & Review*, *21*(5), 1323–1330. https://doi.org/10.3758/s13423-014-0588-3

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., 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

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

Statistics Canada. (2017). *Census profile, 2016 census*. https://www12.statcan.gc.ca/census-recensement/2016/dp-pd/prof/index.cfm?Lang=E

Wulf, G., & Shea, C. H. (2002). Principles derived from the study of simple skills do not generalize to complex skill learning. *Psychonomic Bulletin & Review*, *9*(2), 185–211. https://doi.org/10.3758/BF03196276

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

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

Ziegler, E., & Stern, E. (2014). Delayed benefits of learning elementary algebraic transformations through contrasted comparisons. *Learning and Instruction*, *33*(1), 131–146. https://doi.org/10.1016/j.learninstruc.2014.04.006