Memory Retrieval Strategies to Help Retain STEM Content Knowledge

Prof. Osman Yasar, State University of New York, Brockport

Osman Yasar is an endowed professor and director of the CMST Institute at The College at Brockport, SUNY. He established the first undergraduate degree program in computational science in the United States, and his research interests include engineering and science education, computational pedagogy, fluid and particle dynamics, engine ignition modeling, and parallel computing. Yasar has a PhD in engineering physics and an MS in computer science from the University of Wisconsin–Madison. Contact him at oyasar@brockport.edu.

Dr. Peter Veronesi, The College at Brockport

Peter Veronesi is program coordinator and lead faculty for the secondary science education programs at The College at Brockport.

Dr. Jose Maliekal, The College at Brockport, SUNY

Jose Maliekal is the Dean of the School of Arts and Sciences at The College at Brockport, State University of New York. He is an atmospheric scientist and have been teaching meteorology and climatology courses. His research interests include climate change and computational pedagogy.

Address: School of Arts and Sciences, The College at Brockport, Brockport, NY 14420 Phone: 585-395-5598 Email: jmalieka@brockport.edu

Dr. Leigh J. Little, SUNY Brockport

Leigh Little is currently a member of the Earth Sciences Department at SUNY Brockport. His research interests include technology education, high performance computing and preconditioning of linear systems.

Dr. John W. Tillotson, Syracuse University

John W. Tillotson is an associate professor of STEM education and chair of the Department of Science Teaching in the College of Arts & Sciences at Syracuse University. He currently serves as the director of the Ph.D. in College Science Teaching doctoral program. His research focuses on STEM instructors' beliefs and instructional practices in undergraduate environments, as well as the socialization experiences of undergraduate STEM majors and how they relate to retention within the STEM career pipeline. Dr. Tillotson earned his Ph.D. in science education and his M.S. in chemistry from the University of Iowa, and prior to that received a B.S. in Chemistry/Secondary Education from SUNY Cortland.

Memory Retrieval Strategies to Help Retain STEM Content Knowledge

Abstract

We report about memory retrieval experiences to help students retrieve content they learned in class, retain it, and apply it in different contexts to solve novel problems. Supported by multi-year fall/spring professional development opportunities for teachers, these technological and pedagogical experiences range in complexity from simple electronic flashcards for basic retrieval strategies to low-stakes quizzes for spaced-out (initial exposure and retrieval effort are spaced out) and interleaved (two or more spaced-out topics are interleaved) practices. A sequential mixed-methods approach was used to collect quantitative data from a large number of participating teachers (N=180), followed by an enriched case study with a qualitative component to explore the meaning of the quantitative trends/findings in the first part of the study. Participants reported that they gained a greater understanding of the science behind the concept of interleaving, a greater understanding of how it can be implemented and tested in the classroom, and a higher level of confidence in the effectiveness of interleaving on knowledge retention than they had prior to training. While deployment of retrieval strategies in the classroom has been required of all participants, those who attended additional training in the summers (N=68) have also conducted Action Research to measure the effect of new strategies on learning. These teachers randomly selected control and target student groups within the same school, grade and course environment. They also self-selected an area of content within their respective science disciplines or mathematics curriculum and created two different retrieval practices – a blocked practice that examines student knowledge and skills for applying a certain method to the solution of various questions on only one topic or type, and the interleaved practice that involves questions on two or more topics that need different methods to solve. Results from the first summer cohort (N=16) show that students who learned math and science topics through interleaved practices consistently scored 5-30% better than those who learned it in the more traditional blocked practice. In many cases, the differences were statistically significant ($p < 0.05$). While the second summer cohort (N=42) continues its action research, our future work will attempt to reduce confounding variables in research experiments and repeat them with more robust techniques and another level of memory retrieval strategy to help students not only recall what they learned in a classroom but also apply their content knowledge and computational skills to problem solving in a generative fashion beyond just answering multiple-choice questions.

Keywords: Memory retrieval, interleaved practice, computational thinking, teacher training, professional development,

1. Introduction

There are yet to be any content standards for teacher professional development and student learning outcomes in engineering, however, recent national efforts[11-12] have helped build some momentum for standardization in engineering education. While a few states have taken bold steps to make engineering education accessible to all K-12 students, others are also using current content standards to promote science and engineering (S&E) practices such as:[12]

1. Asking questions (for science) and defining problems (for engineering)
2. Developing and using models
3. Planning and carrying out investigations
4. Analyzing and interpreting data
5. Using mathematics and computational thinking
6. Constructing explanations (for science) and designing solutions (for engineering)

7. Engaging in argument from evidence
8. Obtaining, evaluating, and communicating information

There are many similarities between the practices of scientists and engineers – e.g., both include using computational tools to test scientific theories and predict outcomes of engineering designs. While new technologies and pedagogies now afford us many opportunities to cultivate students' S&E habits of mind,[4,5,18] developing novel approaches to integrate technology, pedagogy, and content knowledge (TPCK, or TPACK) has been a difficult and complex problem whose solution is often custom-designed by teachers.[8] Luckily, when teaching with *computational* technology, the integration occurs naturally because computation inherently involves not only multiple content areas (e.g., math, computing science) but also inductive and deductive pedagogies.[15] Yet, professional development (PD) to support meaningful uses of computational technologies and pedagogies is still sorely needed. It is important that teachers are educated in the most effective ways to sustain their *computational* PCK (or CPACK) skills and experiences.[16-17]

Educational psychologists and neuroscientists have recently compiled a critical mass of empirical findings to change the way teachers are prepared and supported in their classroom instruction. One of the benefits of these findings is that they provide foundational support for us to better understand strategies and theories that have long been reported in educational psychology journals, including the optimum flow, scaffolding, and zone of proximal development (ZPD) that suggest to challenge learners in incremental steps as they build more skills and simultaneously enjoy learning with the help of a *cognitive push*.[1,10] At the same time, cognitive and computational scientists are trying to consolidate all these numerous cognitive and neuropsychological findings into a single framework to explain how our brain works as a computational device to facilitate thinking.[9] Besides structural (hardware) factors, it turns out that there are non-structural (software) factors that determine how information is stored, retrieved, and processed by a computing device. The information that is acted upon by a computer, be it electronic or biological, allows certain affordances and places various restrictions. For example, like granular matter, information constructs tend to undergo two fundamental actions; they either unite *associatively* to make bigger constructs or break down *distributively* to smaller ones.[19]

Though quite simplistic, this dual (associative/distributive) nature of information may help us understand how it may optimally be stored, retrieved, and processed by a computing device, regardless of its core hardware. Such a need-and-capacity driven optimization naturally forces computing devices to evolve. For example, the inductive/deductive way our minds currently operate[4] could very well be an evolutionary response, shaped over many years to optimally deal with incoming sensory information whose quantifiable nature is receptive only to a dual set of computable operations. The evolution in electronic computing devices from centralized hardware to today's distributed structure also appears to be of similar nature in order to more optimally handle a growing need for faster processing and more storage. So, today's electronic computing devices, our cognitive view of the mind, and neuroscientific understanding of the brain are all converging to a point, indicating that common principles may be at work regarding information storage, retrieval, and processing.[19]

Accordingly, if we consider an information construct (e.g., a concept) as a combination of other smaller concepts, these sub-concepts might also involve another level of even smaller concepts and details that may be stored and mapped onto the brain's neural network in a hierarchical way.[19] Then, instead of creating a new record for incoming sensory information, the brain hardware would rather store it in the form of a specific (distributed) pattern of neurons placed on a pathway and linked to all other associated patterns of previously stored relevant concepts and memories.

This is consistent with neuropsychological findings[2,6]. So, when new information arrives, it lights up all related neurons and pathways in a ***distributive*** process that is similar to a top-down action, where a concept/memory is broken up into related pieces. On the other hand, retrieving a memory is like a reassembly of its original pattern of neurons and pathways in an ***associative*** process that is similar to a bottom-up action. To ultimately improve on the storage and retrieval of information, one would simply want the quantity and strength of both pathways to increase. That is, the more links to associated concepts, the higher the chances of recalling the newly acquired concept when needed later. Cognitive retrieval practices attempted at different times, various settings and multiple contexts would be more effective because every time a recall is attempted it would establish more significant links that would help the remembering. Consequently, exposure to new concepts, through links to multiple views from different fields, is an effective retrieval strategy recommended by cognitive psychologists.[2]

From these cognitive views of distributive storage and associative retrieval of information, it is logical to think that recalling a memory is merely an act of creative re-imagination (thinking) and what is retrieved would depend on the effort – that is, it will not necessarily be the original pattern but one with some holes or extra bits. This view is consistent with those by neuroscientists who now see little or no distinction between the acts of information storage/retrieval and the act of thinking.[2,9] A distributive act of information storage appears to be no different than deductive thinking and an associative act of retrieval being no different than inductive thinking. If so, like the act of inductive thinking, retrieval then would require an effort, and the harder the learner tries, the better recall occurs. As described below, this one simple tenet of (computational) thinking[19] can dramatically change what we have known to this date about remembering.

While deliberations of computational thinking in the context of expert thinking (e.g., scientific and engineering thinking) has been previously reported in the literature,[18] examining its validity in information retrieval/retention and learning by novices is also equally interesting. A contentious area it might be tested on is the role of tests in education and whether they help or impede student learning. For example, while tests have often been designed as high-stakes exams to evaluate student performance, new findings[2] provide empirical evidence that they can also serve as important (low-stakes) learning tools to help students retrieve newly taught concepts in effortful ways that will, in essence, burn new knowledge into memory through connected understandings rather than rote memorization. When learning is based in challenges, it is more strongly held in the brain and it lasts longer.[10] Another common misconception held by many teachers is that learning that is repeated is better or perhaps, more long-lasting. Many studies have found that while repetition or review of a lecture by teachers and review of lecture notes or rereading of a textbook by students might help remember the content for a short while, such repeated exposure does not lead to long-term memory and meaningful, connected recall. For sure, repetition of conceptual ideas helps, especially if done in multi modal ways. Actually, a cognitive and constructive effort to recall recently learned concepts through connections to what had been previously stored in long-term memory might have a much greater chance of being retained longer. In that sense, *to support self-constructive recalls, some actual forgetting needs to happen* and because of this forgetting every time such recall is practiced it reassembles sought-after concepts through different and more learning pathways or links to one's knowledge that is already known and easily retrievable. There are multiple ways to accomplish these learning pathways or links according to a recent book by Brown, Roediger and McDaniel.[2] One of them is *spaced-out retrieval practice* through low-stakes quizzes, self-testing, or flashcards. Spacing allows some forgetting that will trigger a cognitive effort for retrieval while repeated retrieval leads to more durable memories along with additional knowledge produced by connecting the dots and linking these memories and knowledge to a greater number of concepts, situations, and problems.

Another one is *interleaved practice* that also helps link newly learned concepts to different contexts, changing conditions and parameters, and even multiple STEM subjects.

Both spaced-out and interleaved retrieval practices have now been already tested in social sciences[2] and mathematics[13] against the usual *blocked practice* whereby students learn to apply a certain method to solution of various questions of the same type on only one topic. An example of blocked practice would be to apply the Pythagorean Theorem to compute the hypotenuse of a right-angle triangle, $d^2= x^2 + y^2$. Here, students need not learn to choose a solution method because every problem within a blocked practice would require the same strategy. In an *interleaved retrieval practice*, two or more types of questions are mixed, and students are faced with choosing a strategy to solve a problem. Blocked practice is still prevalent in schools for many reasons, including the belief that repeated practice of the same drill builds up skills. The interleaved practice, on the other hand, would require a re-arrangement of class schedules and the order in which topics are taught and practiced. Teachers would obviously need to devote more time and energy in some cases. More importantly, they would need to be trained to implement the interleaved practice. As benefits of the interleaved practice are being documented in the literature, this study, to our knowledge, may be one of the first to test it in natural sciences and engineering education. Also, it not only puts interleaving on a strong theoretical (cognitive and computational) foundation,[19] but it also involves teacher professional development (PD) to help implement it. The following sections describe our research design, along with findings on the effectiveness of interleaving in learning as well as quantitative/qualitative feedback on the PD program that helped secondary-school teachers implement and research it in their classrooms.

2. Research Design

In the past two school years, during fall and spring, we offered: a) introductory training on basic retrieval practices[2] and Google Forms (to prepare and conduct practice tests) to 180 teachers from 29 local school districts (SDs) in our region, including 33% from urban, 19% from rural, and 48% from suburban SDs, and b) additional training on interleaving retrieval strategies to returning 91 teachers, including 42% from urban, 22% from 6 traditionally underrepresented rural, and 36% from 14 suburban SDs. Participation required deployment of retrieval strategies in the classroom after the training. Additional summer training and yearlong support were offered to interested teachers who wanted to undertake Action Research and conduct a control/target comparison in classrooms that they would teach the following school year. During the past two summers, previously trained teachers were invited to the college to share ideas and work together to develop lesson plans and retrieval practices for classroom implementation. Sixteen returning teachers attended the 2017 summer workshop and 42 attended the 2018 summer workshop.

A sequential mixed-methods approach[3] was used to collect quantitative data from participating teachers (e.g., pre- and post-activity teacher surveys, classroom artifacts, and student test scores), followed by an enriched case study with a qualitative component (e.g., focus group interviews, teacher activity logs, and classroom observations) to explore the meaning of the quantitative trends/findings in the first part of the study. Quantitative data from students (surveys and unit test scores) were collected to get an initial response from a large number of students. A control/target experiment was conducted by summer teachers following their training. Results from the first summer cohort is presented in this article under the Action Research. The overall experiment evolved in phases as the program staff developed, in collaboration with participating teachers, a database of curricular modules, lesson plans, and related assessment instruments and rubrics with good psychometric properties. The data was collected by two professional evaluators, who also coded open-ended responses and used an inductive analysis to identify major themes emerging from teacher activity logs, questionnaires, and journals.

The Action Research component by the 2017 summer cohort of 16 teachers explored the impact of interleaved practice on teaching and learning during the spring of 2018. Student study included various topics in geometry, biology, chemistry, and earth sciences. While each study explored memory retrieval with slightly different arrangements (See Tables 1 and 2), they all mixed (interleaved) at least two topics in each study. Students were placed randomly into control and target groups of equal size ranging from 12 to 32, depending on the study. About half of the cases followed the research design in Table 1 where one group (A) followed the blocked strategy while the other group (B) used an interleaved strategy for practices and assignments involved in the teaching of Topics X & Y. Other cases followed the design in Table 2 where each group (A & B) was taught via both strategies (blocked and interleaved). Care was taken throughout to make sure that group placement was not visible to students; that is, all students participated in simultaneous classroom instruction. Instruction for each topic lasted for the same number of days with both strategies. An in-class review was conducted a short while after completion of instruction in all groups. The review was concluded with a test which in some cases served as a baseline. Finally, an unannounced test was conducted later to measure student retention of the subject 15-30 days after the review. In most cases, teachers conducted pre- and post-activity assessment with multiple-choice questions on all control and target groups to identify/reduce the number of confounding variables and triangulate the results as much as possible. They all used Google Forms to prepare and conduct practice tests. Many used CastleLearning™ database to generate test questions.

GROUP	Assignment 1	Assignment 2	Time Delay	Unit Review Test	Time Delay	Unannounced Test
Blocked (A)	Topic X	Topic Y	3 days		15-30 days	
	Topic X	Topic Y				
	Topic X	Topic Y				
	Topic X	Topic Y				
	Topic X	Topic Y				
	Topic X	Topic Y				
Interleaved (B)	Topic Y	Topic X	3 days		15-30 days	
	Topic Y	Topic X				
	Topic Y	Topic X				
	Topic X	Topic Y				
	Topic X	Topic Y				
	Topic X	Topic Y				

Table 1: Timeline and scheduling of a control/target study with separate retrieval strategies.

GROUP	Topic X	Topic Y	Time Delay	Unit Review and Test	Time Delay	Unannounced Test
A	Interleaved	Blocked	3 days		15-30 days	
	5 days	5 days				
B	Blocked	Interleaved				
	5 days	5 days				

Table 2: Timeline and scheduling of a control/target study with mixed retrieval strategies.

3. Quantitative Data

a) Teacher Surveys: In a teacher survey conducted after the 2017 summer workshop, the data suggested that teachers found the program to be both engaging and effective in raising their awareness of retrieval strategies as a pedagogical practice and of the research base supporting its efficacy within STEM classrooms as indicated by mean scores above 4.0 on a 5-point scale for nearly every survey item. The respondents highly valued the opportunities to collaborate with other teachers in designing online retrieval practices and considering possible research designs for studies they planned to conduct during the upcoming 2017-2018 school year. Similar data from the 2018 summer workshop participants also suggested that the bulk of participating teachers (88%) found the workshops to be valuable overall, specifically in providing them with the conceptual knowledge and practical skills necessary to effectively engage in implementation efforts during the 2018-2019 school year. Ninety-six percent (96%) of the teacher respondents highly valued the opportunity to learn about interleaving strategies and the research base supporting the effectiveness of this pedagogical approach in K-12 classrooms. The significant percentage of the teacher participants (96%) also indicated that the time devoted to accomplishing each of the primary workshop objectives was appropriate, and that the workshop was effective in helping them develop a clear understanding of their roles and responsibilities pertaining to the classroom-based Action Research component of the grant project (93%).

b) Action Research: The following quantitative data represents case studies by 2017 summer teachers who took part in Action Research during spring 2018. To compute differences between groups and within groups, a *t*-test statistic has been applied because of the small sample sizes ($n<30$). The *p* values have been computed within Excel using the TTEST function.

Biology/Living Environment:

- 7th Grade (Topic: Punnett squares, Research Design: Table 1): The average score of 4 daily assignments conducted in the same week (March 27, 28, 29, and April 2, 2018) for Group A (blocked, n=27) and Group B (interleaved, n=29) was about the same (46.79 vs 47.34 out of 100). Both groups received a unit review on April 17 that ended with a test to set a baseline for measuring how much they would retain/recall later. Group A's average on the review test was 55.88 versus Group B's average of score of 52.45. When an unannounced quiz was given to both groups 15 days later, Group B's average score (60.21) was 4.4% higher than Group A's (57.67). While such a difference is not statistically significant, Group B's average improved significantly by 8 points versus Group A's not-so-significant 2-point improvement. A similar trend was seen in the analysis of each student's progress for both sample groups. For example, 19 students in the interleaved group increased their score while 8 decreased versus 14 increasing and 10 decreasing in the blocked group.

Punnett Squares	Group A (n=27) BLOCKED	Group B (n=29) INTERLEAVED	*Is the difference statistically significant?*
Review-test	55.88	52.45	not ($p = 0.52$)
Post-review test	57.67	60.21	not ($p = 0.71$)
Is the difference significant?	not significant $p = 0.50$	significant $p = 0.01$	

- 9th Grade (Topics: Photosynthesis and Respiration, Research Design: Table 2): Two groups (n=22 each) experimented with alternating practices on different topics. While Group A used the blocked practice to learn about photosynthesis, Group B used the interleaved practice. Similarly, while Group B used the blocked practice to learn about respiration, Group A used the interleaved practice. Groups were compared to themselves to examine their review and post-review retrieval and to their counterparts (post-review) which had learned the same topic via different practices. As shown in the tables below, while groups that used the blocked practice scored higher on pre-tests, they scored lower than the groups with the interleaved practice. While pre- and post-test differences between groups are not statistically significant, the improvements by the interleaved groups from pre-test to post-test are significant and large enough to exceed their counterpart.

Photosynthesis	Group A BLOCKED	Group B INTERLEAVED	*Is the difference significant?*
Review-test	53.66	45.86	not; $p = 0.19$)
Post-review test	57.47	62.04	not; $p = 0.55$)
Is the difference significant?	not significant $p = 0.59$	significant $p = 0.001$	

Respiration	Group A INTERLEAVED	Group B BLOCKED	*Is the difference significant?*
Review-test	15.59	20.27	not; $p = 0.14$
Post-review test	57.72	50.90	not; $p = 0.33$
Is the difference significant?	yes; $p < 0.01$	yes; $p < 0.01$	

Environmental/Earth Sciences:

- 7th Grade (Topics: Erosion, Planetary Motion, Research Design: Table 2): As shown in the table below, Group A (interleaved, n=30) outperformed Group B (blocked, n=31) by 9% on a post-test on weathering and erosion; a difference that is statistically significant. In another experiment on planetary motion and the effect of mass on the gravity of an object, Group B (interleaved, n=21) outperformed Group A (blocked, n=29) by 38%; a difference that is statistically very significant. According to the teacher, in the same experiment, the interleaved group also outperformed the blocked group by 30% on levels of organization (progression of levels by cells to reach an organism).

Topic	Post-review test Group A	Post-review test Group B	*Is the difference significant?*
Erosion	INTERLEAVED (n= 30) 8.166/10	BLOCKED 7.516/10	yes; $p = 0.013$
Planetary motion	BLOCKED (n=29) 3.068/6	INTERLEAVED (n=21) 4.238/6	yes; $p = 0.008$

Geometry:

- 10th Grade (Topics: Properties of Quads, Equations of Lines, Circles-Angles/Arcs, Circles-Segment Lengths, Research Design: Table 2): Group A (n=23) and Group B (n=28) used blocked and interleaved practices in an alternating fashion for various topics. Both groups were local geometry classes with special education (SPED) students in them (6 in Group A and 11 in Group B). The following table shows class test averages, along with t-test statistic and p values. Students were all given a review (and a test) right after the blocked and interleaved practices ended. As shown in the upper part of the table, interleaved groups outperformed the blocked groups across the board (All/Gen Ed/SPED). While the differences range from 3.4% to 13.4 %, for some of them such as Equations of Lines, the differences are statistically very significant. 30 days later, an unannounced test was given. While average scores went down for all groups, the interleaved groups again outperformed their counterparts. The differences were statistically significant for topics such as Equations of Lines and Circles-Segment Lengths.

REVIEW-TEST (after treatment, out of 12 points)							
N=23	N=28	N=23	N=28	N=23	N=28	N=23	N=28
Class A	Class B	Class A	Class B	Class A	Class B	Class A	Class B
Properties of Quads		Equations of Lines		Circles-Angles/Arcs		Circles-Segment Lengths	
Interleaved	*Blocked*	*Blocked*	*Interleaved*	*Interleaved*	*Blocked*	*Blocked*	*Interleaved*
Average test results for All, General Education (Gen Ed) & Special Education (SPED) students							
8.87 (All) 8.94 (Gen Ed)	8.07 8.24	7.87 8.05	8.928 9.35	9.17 9.71	8.61 9.24	8.52 8.76	8.92 9.24

8.67 (SPED)	7.82	7.33	8.27	7.67	7.64	7.83	8.45
Diff (I-B);All	0.7981		1.0590		0.5667		0.4068
p (2 sample)	0.109		0.0164		0.173		0.279
UNANNOUNCED POST-REVIEW TEST (out of 12 points) – 30 days after the review test							
Average test results for All students, General Education (Gen Ed) & Special Education (SPED) students							
8.17 (All) 8.29 (Gen Ed) 7.83 (SPED)	7.61 8.00 7.00	7.22 7.59 6.17	7.928 8.24 7.45	8.70 9.12 7.50	8.25 8.71 7.55	7.652 8.06 6.50	8.50 8.88 7.91
p (paired samples) for post-review test and review-test comparison within group < 0.01 for all cases							
p (2 sample)	0.2		0.05		0.22		0.01

Chemistry:

- 10th Grade (Topics: Half-life, Heat formula, Research Design: Table 2): While learning about half-life, Group A (n=20, blocked) and Group B (n=18, interleaved) scored about the same (Group A at 63.5 vs Group B at 61.4) at the review test, their performance on an unannounced test given 30 days later differed significantly (p=0.014) with Group A's average being 40 vs Group B's 63.3. On the topic of heat, while Group A scored significantly (p=0.027) higher than Group B (81.1 vs 68.8) at the review test, its performance (32.6) on an unannounced test given 30 days later fell substantially (p= 0.1) below Group B's (42.4). While the drops in performance by both groups were statistically significant (p= 0.0006), the drop by the group (A) with the blocked practice was twice as much as the drop by the group with the interleaved practice.
- Grades 9-12 (Topics: Le Chatelier's Principle, Potential Energy Diagrams, Research Design: Table 2): Each student completed a pre-test before instruction and practice began. The pre-test consisted of 6 multiple-choice questions; the first 3 questions involved Le Çhatielier's Principle and the remaining 3 questions involved potential energy diagrams. A post-test consisting of similar questions was administered 30 days after the end of instruction, subsequent to an in-class review of both topics. On the LeÇhatielier's Principle, while both groups improved considerably after receiving instruction, Group A (interleaved, n=12) improved twice more than Group B (blocked, n=12). The increase in its score between pre- and post-tests was from 27 to 75 versus Group B's improvement from 27.6 to 47. The interleaved group outperformed the blocked group significantly by 60% on the post-test. On the topic of Potential Energy Diagrams, the increase between pre- and post-tests was from 8 to 69 for Group A (blocked) versus from 22 to 58 for Group B (interleaved). Although the group with the blocked practice (Group A) improved more than the other group with the interleaved practice, the margin of difference between the two groups was higher when Group A received interleaved instruction, as shown in the table below.

Le Chatelier's Principle	Group A INTERLEAVED	Group B BLOCKED	*is difference significant?* p (2 sample)
Pre-test	27.0	27.6	no; p=1.0
Post-test	75.0	47.0	yes; p=0.03
p (paired)	0.001	0.11	

Potential Energy Diagrams	Group A BLOCKED	Group B INTERLEAVED	*is difference significant?* p (2 sample)
Pre-test	8	22	no (p=0.12)
Post-test	69	58	no (p=0.47)
p (paired)	0.0001	0.015	

4. Qualitative Data

Teacher Focus Group Interviews: To explore the meaning of the quantitative trends/findings in the first part of the teacher feedback on the program, focus group interviews were conducted by external evaluators to provide more specific feedback on their experience during the 2018 summer workshop. The independent evaluator report indicated that participants reported a high degree of satisfaction with the workshops. Participants felt that they gained a greater understanding of the science behind the concept of interleaving, a greater understanding of how it can be implemented in the classroom, and a higher level of confidence in the effectiveness of interleaving on student information retention than they had prior to the workshop. Most reported having used some components of interleaving previously, although many were not aware of the name or some of the more structured practices they learned in the workshops. The consensus from all focus group participants was that the workshops were time well-spent, and there was general excitement about implementing and tracking the results of interleaving in their classrooms. Many looked to voluntarily providing feedback to the program after implementation and were pleased that continuing communication and support were available after the workshops. Most respondents felt the interleaving strategies they learned in the workshop had potential to help their students better retain information and be more successful in recalling information. Many participants noted that they were somewhat surprised by the evidence that more rigorous teaching and assessment with lessons spaced out over time and regular assessments helped students retain information better, contrary to current discussions on over-testing. They noted that one strength of interleaving was the ability to help students learn *which* strategy to apply to problems, in addition to helping them recall information better, because blocked units frequently use the same formulas and students don't have to think about which one to apply – i.e., interleaving forces them to choose an appropriate formula. Nearly all participants stated that they plan to implement interleaving strategies in the upcoming school year. Some participants noted potential difficulties in student completion of online assignments, with issues of inaccessibility for some without access to computers or smartphones, and concerns about inconsistent student attendance potentially impacting their assessment data results. Respondents were nearly unanimous that having previously trained teachers who have used interleaving in the classroom providing hands-on assistance and addressing real classroom implementation issues was the most beneficial aspect of their experience at the workshop. Many found the focus of the workshop on math and science to be much more helpful than many professional development workshops they had attended in the past. Some commented while other workshops have been a "waste of time" for them when the concepts covered could not be applied to their teaching practice, they were very pleased that this workshop was much more beneficial to them. Nearly all respondents considered the workshops "time well-spent." There were also recommendations to help improve the program further such as a) asking teachers to bring their curriculum so they can plan their workshop projects around their actual planned classroom instruction, b) making the sessions more interactive, and c) providing teachers with more clear written instructions for participating in Action Research in the upcoming school year.

Another focus group interview was conducted by a different evaluator with the 2017 summer cohort, following a gathering where teachers discussed their spring 2018 data, which was listed in the previous section. According to the independent evaluator report:

- Participants perceived the benefits of utilizing interleaving as an educational strategy to be significant, particularly when dealing with students who have traditionally struggled to succeed in learning science and math concepts.

- Participants reported that using the Google Forms and the interleaved assessments allowed for better tracking and diagnostics of student engagement and conceptual learning which translated into more differentiated instruction.
- Participants reported an increased use of problem-solving tasks, small group exercises, and more class discussions because of the focus on implementing interleaving strategies. The interview data identified common themes among teacher respondents related to shifts in their pedagogical beliefs and classroom instructional practices. Most notable were the use of student-centered teaching techniques that actively engaged students in learning about key science and mathematics concepts.
- Both veteran and novice teachers noted that the interleaving PD program caused them to be more purposeful in selecting their teaching strategies, curriculum materials, and assessments. Several teachers described in detail how they felt their teaching practices and decision-making had become more "purposeful" as a result of the PD experiences. One teacher noted, "The interleaved practice model provides students with practice of each topic throughout the entirety of a unit, not only in instruction and other activities, but also through formative assessment methods…It allows a teacher (and students) to incorporate multiple subjects into the learning process in hopes of putting more content knowledge into long-term memory. Furthermore, while the topics of genetics and evolution are generally two separate units, connections can easily be made between the two whereby genetics topics are used to supplement teaching of evolution, or vice versa." Another noted, "Interleaving was a term I had not heard prior to entering into this research. Apparently, I have always done interleaving within my classes, but this study has made me realize I need to find a way to make this more common in daily assignments I give to students."
- Teachers viewed their participation in Action Research as a professional learning experience that they felt would ultimately benefit their students.
- Participants noted a greater interest in, and use of, program's educational technology tools (e.g. Google Forms, Chromebooks, and Castle Learning) to support student learning than prior to their involvement in the project.
- Participants perceived that students demonstrated greater resiliency and comfort levels with revisiting content topics from other parts of the curriculum as a result of the regular use of interleaving in their classrooms. Teachers reflected on students' developing a sense of autonomy and persistence as learners. "I can give them a question on a test pretty much for anything for the year and it's not like, "Well, this wasn't going to be on the test so this isn't fair" because they now expect that anything they've learned is fair game throughout the whole year, so that has been a huge help for me," noted one respondent. Another teacher offered, "When I give them a test, I can put a spiral question on there and again it's not like, "Well, I'm not answering it because this wasn't on the review sheet." And especially in math, I can give them a deeper level question where it's not a simple equation that they might have just seen on their homework, but it could be worded completely differently, and again, they're not saying "Well, I never saw this exact wording before, so you didn't teach me anything." They become more resilient, I think, by just trying to answer the questions." Teacher interviewees also reported that students became more comfortable taking risks in the classroom related to their learning of STEM concepts. One teacher noted that "Students need to know they can take a shot at it and not have the fear of God if they get it wrong. It's taking a shot, you get it wrong, and you're going to see it again, but by the end of the year you'll know this stuff. That's really the intent. I see positive results. I see more results with special education students and students that struggle, IEP's, 504's… They're at least trying. They're putting the effort in. If anything, I think it has built some resilience in the majority of the students that I've worked with."

- The increased use of educational technology on a regular basis resulted in students coming to class better prepared and expecting to use technology tools in their STEM courses. One urban school teacher noted that “I'm finding my kids are holding themselves a little bit more accountable and they're more responsible with that technology piece. We have one-to-one so the kids all have a Chromebook and I found in my spring research that the kids… were coming more prepared, they were using the technology if it was Google Forms, or Castle Learning, or some kind of platform, they were engaging with it a little more than they would probably a paper and pencil warm up.”
- The focus group interviews also offered reflections and recommendation. They found it challenging to allocate the time necessary to adopt the interleaving strategies, modify course materials, and conduct action research given the many scheduling constraints they routinely face. A teacher noted that “The interleaving study was a valuable experience. However, it requires a significant amount of time to construct each assignment with the necessary balance of questions for each topic.” Teachers also indicated the need for additional time to meet with each other teacher to discuss Action Research project designs, to give and receive peer feedback, and to compare and discuss research results.

5. Conclusion

To support computational pedagogical experiences in STEM teaching, we offered professional development opportunities on memory retrieval strategies to secondary school teachers from an urban city surrounded by many suburban and rural school districts. Both quantitative and qualitative data from teachers pointed to the effectiveness of the interleaved retrieval strategy in the classroom. Many teacher participants noted that more rigorous teaching and assessment with lessons spaced out over time and regular assessments helped students retain information better, contrary to current discussions on over-testing. They viewed their participation in Action Research as beneficial to their students. Those who took part in focus group interviews reported that using online forms and the interleaved assessments allowed for better tracking and diagnostics of student engagement and conceptual learning which translated into more differentiated instruction. They also perceived that students demonstrated greater resiliency and comfort levels with revisiting content topics from other parts of the curriculum as a result of the regular use of interleaving in their classrooms.

An overall analysis of the current research data shows that students who learned a topic via the interleaved strategy scored considerably better than those who learned in the usual (blocked) way. In many cases reported, the differences were statistically significant in favor of the interleaved practice. While the sample sizes were small (between 12 and 32) and not consistent throughout Action Research experiments, in cases where performance results by both groups were not very different, still the group with the interleaved practice appeared to have improved its performance considerably more than the group with the blocked practice. While student groups were selected randomly, some confounding variables such as suitability of content for interleaving and each group’s prior experience and background were possibly in play in our research. We suspect that the level of improvement in some cases depended on the nature of topics, grade level, as well as teacher experience and school environment. It seemed that the interleaved practice is more effective with the teaching of more discrete content items (erosion, order of operations, etc.) versus more complex or abstract content themes such as photosynthesis or organ systems functions. It was evident that when concepts were interleaved and spaced-out over time for the student, then more discrete conceptual items were far impactful with interleaving than concepts that were more abstract and non-discrete.

Our findings in natural sciences are consistent with theory and data in social sciences[2] and mathematics[16] which claim that interleaving makes the retrieval process more effortful, thereby keeping student attention alive and opening up more neural pathways for retention because of the mix of contexts in which the topics are taught. They also offer support to the simplistic view of the human brain as a computational device.[9,19] We plan to conduct additional research to reduce confounding variables in our experiments and repeat them with more robust techniques while adding another level of memory retrieval strategy to help students not only recall what they learned in the classroom but also apply their content knowledge and computational skills to problem solving in a generative fashion beyond just answering multiple-choice questions. Results of a TPACK survey[14] for teacher participants will also be presented in future publications.

References

1. Bransford, J., Brown, A., & Cocking, R. (2000). *How People Learn: Brain, Mind, Experience, and School*. Washington, D.C.: National Academy Press.
2. Brown, P. C., Roediger, H. L., & McDaniel, M. A. (2014). *Make it Stick.* Cambridge, MA: The Belknap Press of Harvard University Press.
3. Creswell, J. W. (2012). *Educational Research: Planning, Conducting and Evaluating Quantitative and Qualitative Research*. 4th Edition. Boston, MA: Pearson Education, Inc.
4. Dunbar, K. N., & Klahr, D. (2012). Scientific Thinking and Reasoning. In K. J. Holyoak and R. G. Morrison (Eds.), *The Oxford Handbook of Thinking and Reasoning* (pp. 701-718). London: Oxford University Press.
5. Hanson, J. and Lucas, B. (2018). Engineering the future. A report for the Royal Academy of Engineering. ISBN: 978-1-909327-43-6.
6. Hebb, D. (1949). *The Organization of Behavior*. New York: Wiley & Sons.
7. L-Horsley, S., Stiles, K. E., Mundry, S., Love, N., & Hewson, P. W. (2010). *Designing professional development for teachers of science & mathematics*. Thousand Oaks: Corwin.
8. Mishra, P., Koehler, M. J. (2006). Technological pedagogical content knowledge. *Teachers College Record*, 108 (6), 1017-1054.
9. Montague, R. (2006). *Your Brain Is (Almost) Perfect.* New York, NY: Plume Books.
10. Mooney, C. G. (2013). *An Introduction to Dewey, Montessori, Erikson, Piaget, and Vygotsky*. St. Paul, MN: Redleaf Press.
11. National Academy of Engineering Report (2010). Standards for K-12 Engineering Education. Wash., DC: National Academies Press.
12. National Research Council Report (2012). *A framework for K-12 science education: practices, crosscutting concepts, and core ideas*. Wash., DC: National Academies Press.
13. Rohrer, D., Dedrick, R. F., and Stershic, S. (2014). Interleaved Practice Improves Mathematics Learning. *J. of Educational Psychology, Vol. 107, No. 3; 900-908.*
14. Schmidt, D., Baran, E., Thompson, A., Koehler, M.J., Shin, T, & Mishra, P. (2009). TPACK: The development and validation of an assessment instrument for preservice teachers. *J. Research on Technology Education*, 42, 2009; 123-149.
15. Yaşar, O. & Maliekal, J. (2014). Computational Pedagogy. *Comp. in Sci.&Eng.*,16(3), 78-8.
16. Yaşar, O., Maliekal, J., Veronesi, P., Little, L. and Vattana, S. (2015). Computational Pedagogical Content Knowledge (CPACK). In L. Liu and D. C. Gibson (Eds), *Research Highlights in Technology and Teacher Education*. pp. 79-87. ISBN: 978-1-939797-19-3.
17. Yaşar, O., Veronesi, P., Maliekal, J., Little, L., Vattana, S., Yeter, I. (2016). Computational Pedagogy: Fostering a New Method of Teaching. *Comp. in Education,* 7(3), 51-72.
18. Yaşar, O., Maliekal, J., Veronesi, P., and Little, L. (2017). The essence of scientific and engineering thinking and tools to promoted it. Presented at 2017 ASEE Annual Conference.
19. Yaşar, O. (2018). A new perspective on computational thinking. *Comm. of the ACM,* 61(7); 33-39.