Abstract

Quizzes using test-enhanced learning features of spaced, varied, and interleaved retrieval practice have been shown to support consolidation of knowledge gains in students. In this pilot study, we examined the test-enhanced learning potential of a novel quizzing method designed to Spiral Assessment to Reinforce Knowledge (SPARK; Hageman, 2020) along Bloom taxonomy's levels of Remember, Understand, Apply, Analyze, and Evaluate in an online college course on interdisciplinarity. In addition to spiraling content from individual chapters via increasingly more difficult Bloom questions, our SPARK design cycled in and out of chapters across weekly quizzes allowing for between-chapter connections. Analyses of students' cumulative unit test scores and evaluate question accuracy scores showed no evidence that SPARK quizzes supported consolidation of knowledge gains in our students. We discuss the implications of these findings and outline a potential role for AI tools (e.g., ChatGPT; OpenAI, 2020) to play in creating multiple-choice quiz questions like those on SPARK.

AUTHORS

Omer Ari, Ph.D.
Georgia State University

Kara Hageman, Ph.D.
University of Iowa

Andrea Lofgren, Ph.D.
Georgia State University

Effects of a Novel Quizzing Method on Consolidated Knowledge Gains of Undergraduates Enrolled in an Online College Course

Quizzes are widely used for assessment of student learning outcomes in higher education. Reports suggest a surge may have occurred in the use of quizzes during the pandemic-induced online teaching and learning of the past few years (Beason-Abmayr et al., 2021). Quizzes are likely to attract more attention from post-secondary instructors concerned about students resorting to OpenAI's (2020) artificial intelligence (AI) chatbot ChatGPT and other similar AI-powered tools, which can produce a reasonably good response to many essay or short-response prompts in a matter of seconds. Quizzes, if designed with cognitively engaging features, may offer an alternative to assignments vulnerable to AI-crafted response submissions and can support instructors with logistical advantages including automatic grading and providing students with correct-answer feedback. These advantages may be especially appealing to instructors who teach large online courses where options for mitigating concerns around AI-produced response submissions may be limited.

With this study, which we conducted in Spring 2022 prior to ChatGPT's release in Fall 2022, our goal was to examine test-enhanced learning through weekly online quizzes and whether or not such learning is strong enough to deepen undergraduates' conceptual understanding of interdisciplinary studies content by piloting a novel quizzing method called SPARK (Spiraling Assessment to Reinforce Knowledge) designed by Hageman (2020).

CORRESPONDENCE
Email
oari1@gsu.edu

Test-Enhanced Learning and SPARK

Other than their typical implementation as an assessment of student learning, quizzes have the potential to serve as a test-enhanced learning device for college students (Lawson, 2022). When conducted at spaced-out intervals, quizzing of previously encountered content strengthens memory traces associated with that content and results in stronger retrieval (Glass, 2009; Lyle & Crawford, 2011; McDaniel et al., 2012; Trumbo et al., 2016; Thomas et al., 2018). Furthermore, quizzing appears to promote the acquisition of course material at cognitively higher levels than simple retrieval of factual knowledge. Evidence shows transfer of learning from factual knowledge of course content to application of that content (Butler, 2010; Glass, 2009; Jensen et al. 2014; McDaniel et al., 2013; McDaniel et al., 2015; Thomas et al., 2018). Effects are particularly strong when quizzing is utilized with correct answer feedback (McDaniel et al., 2015) and when a group or whole class discussion of quiz answers is provided (Mayer et al., 2009).

Application of knowledge is only the third level of cognitive processes on the revised Bloom's Taxonomy of Thinking followed by higher level processes of Analyze, Evaluate, and Create (Anderson & Krathwohl, 2001). It does not appear that the current literature has consistently examined test-enhanced learning effects at higher levels of thinking such as Analyze, Evaluate, and Create on Bloom's levels, and if it has, these higher levels of thinking are often lumped together (Richmond et al., 2022). Moreover, in some of the studies, students experienced the quizzing regimen on short-answer questions which demand more effortful responding and may in fact engage some of the knowledge consolidation processes through retrieval and elaboration of previously acquired content. However, these constructed-response assignments are harder to grade than multiple-choice question format quizzes (e.g., Thomas et al., 2018) and they are now vulnerable to AI-generated submissions.

SPARK (Hageman, 2020) combines spaced varied retrieval practice with interleaving of content through a design that spirals questions from previous chapters that are increasingly more difficult and cognitively more demanding. Questions are created around a chosen topic along the levels of thinking identified by Bloom's taxonomy starting with lowest *Remember* level all the way up to the highest *Create* level. First assessed with questions on Bloom's lower levels, content continues to be assessed with increasingly higher-level questions on the ensuing assessments until it is cycled out at the highest level. SPARK's design not only allows students to retrieve previous knowledge, but also to use that knowledge toward higher-level cognitive processes elicited by questions on the R-U-Ap-An-Ev-Cr (Remember-Understand-Apply-Analyze-Evaluate-Create) sequence. Starting with the second quiz, questions from previous chapters are spiraled allowing for between-chapter connections at increasingly higher cognitive levels. This cumulative revisiting of previously encountered content at cognitively more demanding thinking levels allows for unique and idiosyncratic ways of assimilating new information with previously learned content through a strengthening of existing connections and creation of novel ones within each student's long-term memory (Brown et al., 2014; Steyvers & Tenenbaum, 2005).

SPARK's design not only allows students to retrieve previous knowledge, but also to use that knowledge toward higher-level cognitive processes elicited by questions on the R-U-Ap-An-Ev-Cr sequence.

In one study related to SPARK, college students enrolled in an educational psychology course section submitted weekly SPARK assessments on Thursdays and then reviewed and discussed question answers in class the following Tuesday. Their final exam scores were compared to students enrolled in another educational psychology course section who were assessed with three non-comprehensive tests. At the outset, the sections were closely matched on average ACT scores (25.1 for treatment group and 25.4 for control group). Students in the section with SPARK assessments were found to achieve significantly greater final exam scores than the section with three tests with a large effect size of $d = 1.21$ (Hageman, 2016). In another study, Hageman (2020) compared end-of-course (EOC) Algebra I test scores of 8th-graders who completed the SPARK method to peers who completed the study in non-SPARK control classrooms. The SPARK classroom students took a SPARK test every other week and discussed their responses in class the following class session. There were outside support sessions for students who needed additional explanation of SPARK test questions. With standard scores on a standardized mathematics test entered as a covariate, analyses

of EOC Algebra I scores showed SPARK classroom students significantly outperformed their peers in three non-SPARK classrooms with effect sizes ranging from $d = .66$ to $d = 1.68$ (Hageman, 2020).

In this study, we sought to expand on this emerging research base on SPARK by examining its effects on knowledge consolidation of undergraduates enrolled in two sections of an online course on interdisciplinarity. Our study of SPARK quizzes was also an opportunity to explore testing effects in academic topics and courses outside the field of psychology, the traditional testing ground for hypotheses of test-enhanced learning (Richmond et al., 2022).

Method

Participants

The study took place in one of the core courses offered through an undergraduate Interdisciplinary Studies (IDS) degree at a large urban university in southeast U.S. According to data from the College of Education, where the IDS program is housed, most of the students in the program are female (81.4%), non-white (73.8%), non-Hispanic (87%), local residents (91.4%), and are transfers from nearby open-access two-year community colleges (67.2%) representing a predominantly minority background including non-traditional college student status. To gain entry into the IDS program, students must have completed their general education core courses and have a GPA of at least 2.0.

In this study, we sought to expand on this emerging research base on SPARK by examining its effects on knowledge consolidation of undergraduates enrolled in two sections of an online course on interdisciplinarity.

The IDS course introduces students to the program and to interdisciplinarity through textbook study, site visits to community partners, and interactive video recordings of point faculty describing career and future prospects for IDS graduates. During a regular semester, IDS course students study and submit assignments on chapters from a textbook on interdisciplinarity (*Intro to Interdisciplinary Studies*, 3rd Ed., Repko et al., 2019). Students' engagement and learning is facilitated through content modules put together by the instructional team around each chapter. Included in each content module are short videos explaining concepts the instructional team has found to be unfamiliar to previous students, video recordings of interviews conducted with faculty on the topic of the chapter, and study guides. These modules also include other public-domain resources (e.g., podcasts, blog posts, scholarly articles) curated to further elaborate on the topics/concepts from the assigned chapter. For example, for Chapter 5, which focused on the topic of disciplinary perspective and elements that together make up a discipline's perspective (e.g., theories, epistemology, assumptions, theories, concepts, methods, etc.), the content module started with a "Main Ideas" video recording created by a graduate research assistant (GRA) to outline for students the main points covered in the chapter. Another GRA-created video discussed the concept of taxonomies of disciplines by natural sciences versus social sciences and the humanities. A third video we located on the Internet as a public-domain resource provided explanations of the concepts of ontology, epistemology, and methodology. Yet another video featured the course instructor interviewing a biology professor on her thoughts on her discipline's perspective. A study guide was included in the module with instructions for students to download or save it to their computer and use it while studying the chapter. This was followed by an assignment students completed on VoiceThread, a collaborative multimedia presentation platform (VoiceThread LLC, n.d.). On a VoiceThread file created for the chapter, students found an overview of chapter topics through embedded PowerPoint slides. The students were asked to respond to a critical thinking/application question on the final slide by submitting either a 150-word text comment or a 1 min 30 s-long audio response and to reply to a classmate's response by posting a short thoughtful reply. They were asked to submit their VoiceThread assignment and the assigned SPARK quiz before the Sunday midnight deadline.

Recruitment

Institutional review board approval (IRB # H22287) for the study was obtained in late Fall 2021 from the institution of the first author where this study took place. Students enrolled in the two Spring 2022 offerings of the IDS course were invited to participate in the study with an enrollment of 33 in one section (02) and 16 in the other (04). As part of recruitment, students were informed of study purposes in the course syllabus and on the course Learning

Questions on the SPARK quizzes, all multiple-choice, were created along Bloom's levels of cognitive engagement to ensure knowledge acquisition

Management System (LMS) website. They were told: a) they would experience a new method of quizzing in the IDS Course called SPARK; b) they would experience SPARK method either during the first or the second half of the semester; and c) cumulative unit test scores would be used to examine effectiveness of SPARK in promoting college students' cognitive engagement with interdisciplinary content covered in the IDS course. Students' consent was sought through an online consent form. Students were offered extra credit (2 percentage points towards their final grade) for completing the consent form. To mitigate concerns of coercion to participate, an alternative extra-credit opportunity was offered to students who wanted to take advantage of the extra-credit opportunity but did not want to consent to their data being used for analysis.

Research Design and Procedures

Using a within- and between-subjects design, we counterbalanced the SPARK and control conditions across the two sections of the IDS course offered in the Spring 2022 semester using random assignment. As seen in Figure 1, Section 02 was assigned to the SPARK method during the first half of the semester and Section 04 to the non-SPARK control condition. During this time, weekly SPARK quizzes were completed only by Section 02 students on Chapters 1-4 with multiple-choice questions spiraled from previous chapters. Section 04, assigned to the control condition, did not submit the weekly SPARK quizzes. At Week 6, both sections took the same cumulative unit test (Unit Test 1) covering content from Chapters 1-4.

Figure 1
IDS Course Sections and Study Conditions

IDS Course Section 02

Week 1	Week 2	Week 3	Week 4	Week 5	Week 6	Week 7	Week 8	Week 9	Week 10	Week 11
Welcome! Module	Chapter 1 Module	Chapter 2 Module	Chapter 3 Module	Chapter 4 Module	Careers Module	Chapter 5 Module	Chapter 6 Module	Chapter 7 Module	Chapter 8 Module	Chapter 9 Module
	SPARK CONDITION					NON-SPARK CONDITION				
Checkin Survey	VoiceThread	VoiceThread	VoiceThread	VoiceThread	Careers Quiz	VoiceThread	VoiceThread	VoiceThread	VoiceThread	VoiceThread
	SPARK Quiz 1	SPARK Quiz 2	SPARK Quiz 3	SPARK Quiz 4	UNIT TEST 1					UNIT TEST 2
	Ch1 R	Ch 1 R	Ch 1 R	Ch 1R	Ch 1R					Ch 5R
	Ch1 R	Ch 1 U	Ch 1 U	Ch 1U	Ch 1U					Ch 5U
	Ch1 R	Ch 1 Ap	Ch 1 Ap	Ch 1 Ap	Ch 1 Ap					Ch 5 Ap
	Ch1 U	Ch 2 R	Ch 1 An	Ch 1 An	Ch 1 An					Ch 5 An
	Ch1 U	Ch 2 R	Ch 2 R	Ch 2 R	Ch 1 Ev					Ch 5 Ev
	Ch1 U	Ch 2 R	Ch 2 U	Ch 2 U	Ch 2 R					Ch 6 R
		Ch 2 U	Ch 2 Ap	Ch 2 Ap	Ch 2 U					Ch 6 U
		Ch 2 U	Ch 3 R	Ch 2 An	Ch 2 Ap					Ch 6 Ap
		Ch 2 U	Ch 3 R	Ch 3 R	Ch 2 An					Ch 6 An
			Ch 3 R	Ch 3 U	Ch 2 Ev					Ch 6 Ev
			Ch 3 U	Ch 3 Ap	Ch 3 R					Ch 7 R
			Ch 3 U	Ch 3 An	Ch 3 U					Ch 7 U
			Ch 3 U	Ch 4 R	Ch 3 Ap					Ch 7 Ap
				Ch 4 R	Ch 3 An					Ch 7 An
				Ch 4 R	Ch 3 Ev					Ch 7 Ev
				Ch 4 U	Ch 4 R					Ch 8 R
				Ch 4 U	Ch 4 U					Ch 8 U
				Ch 4 U	Ch 4 Ap					Ch 8 Ap
					Ch 4 An					Ch 8 An
					Ch 4 Ev					Ch 8 Ev

IDS Course Section 04

Week 1	Week 2	Week 3	Week 4	Week 5	Week 6	Week 7	Week 8	Week 9	Week 10	Week 11
Welcome! Module	Chapter 1 Module	Chapter 2 Module	Chapter 3 Module	Chapter 4 Module	Careers Module	Chapter 5 Module	Chapter 6 Module	Chapter 7 Module	Chapter 8 Module	Chapter 9 Module
	NON-SPARK CONDITION					SPARK CONDITION				
Checkin Survey	VoiceThread	VoiceThread	VoiceThread	VoiceThread	Careers Quiz	VoiceThread	VoiceThread	VoiceThread	VoiceThread	VoiceThread
					UNIT TEST 1	SPARK Quiz 5	SPARK Quiz 6	SPARK Quiz 7	SPARK Quiz 8	UNIT TEST 2
					Ch 1R	Ch5 R	Ch 5 R	Ch 5 R	Ch 5R	Ch 5R
					Ch 1U	Ch5 R	Ch 5 U	Ch 5 U	Ch 5U	Ch 5U
					Ch 1 Ap	Ch5 R	Ch 5 Ap	Ch 5 Ap	Ch 5 Ap	Ch 5 Ap
					Ch 1 An	Ch5 U	Ch 6 R	Ch 5 An	Ch 5 An	Ch 5 An
					Ch 1 Ev	Ch5 U	Ch 6 R	Ch 6 R	Ch 6 R	Ch 5 Ev
					Ch 2 R	Ch5 U	Ch 6 R	Ch 6 U	Ch 6 U	Ch 6 R
					Ch 2 U		Ch 6 U	Ch 6 Ap	Ch 6 Ap	Ch 6 U
					Ch 2 Ap		Ch 6 U	Ch 7 R	Ch 6 An	Ch 6 Ap
					Ch 2 An		Ch 6 U	Ch 7 R	Ch 7 R	Ch 6 An
					Ch 2 Ev			Ch 7 R	Ch 7 U	Ch 6 Ev
					Ch 3 R			Ch 7 U	Ch 7 Ap	Ch 7 R
					Ch 3 U			Ch 7 U	Ch 7 An	Ch 7 U
					Ch 3 Ap			Ch 7 U	Ch 8 R	Ch 7 Ap
					Ch 3 An				Ch 8 R	Ch 7 An
					Ch 3 Ev				Ch 8 R	Ch 7 Ev
					Ch 4 R				Ch 8 U	Ch 8 R
					Ch 4 U				Ch 8 U	Ch 8 U
					Ch 4 Ap				Ch 8 U	Ch 8 Ap
					Ch 4 An					Ch 8 An
					Ch 4 Ev					Ch 8 Ev

Note: Ch refers to a chapter of the *Introduction to Interdisciplinary Studies* textbook used in EDUC 3100; *R* refers to Remember, *U* to Understand, *Ap* to Apply, *An* to Analyze and *Ev* to Evaluate in Bloom's Taxonomy of Thinking Levels. The study of chapters was completed from Week 2- Week 11. There were other assignments for IDS students to complete during the rest of the semester.

Questions on the SPARK quizzes, all multiple-choice, were created along Bloom's levels of cognitive engagement to ensure knowledge acquisition with questions at lower levels of Remember (R) and Understand (U) and to elicit higher-order cognitive engagement with acquired content knowledge at higher levels with Apply (Ap), Analyze (An), and Evaluate (Ev) questions. *Remember* questions tested factual knowledge retrieval. *Understand* questions were created to test students' conceptual understanding of those facts and phenomena. When students answered *Apply* questions, they were tested on their ability to apply that understanding to a novel context, situation, or object. With *Analysis* questions, we measured students' ability to analyze an unrelated topic/phenomenon using the knowledge and skills they had gained around the main topic. *Evaluate* questions, the highest level of Bloom's taxonomy included in these multiple-choice quizzes, required students to check the qualities of a phenomenon, topic, or event against their knowledge of a chapter's topic that had been reinforced in their schema through the previous SPARK questions. Our design did not include *Create* questions which are more appropriate for project-based learning assignments or written essay questions than multiple-choice quizzes.

Course instructor and a GRA met weekly to create SPARK quiz questions. They first individually read the assigned chapter before meeting online to create questions. At the meeting, they identified a topic they thought was the most important theme in the chapter. To do this, they reviewed the learning objectives that came with each chapter and used their own inferences from reading the chapter. After they agreed on a target theme, they created lower-level questions and made their way up on Bloom's levels creating increasingly cognitively demanding questions. When they finished creating questions for an upcoming quiz, the team then sent their questions to the second author, an expert in test question development, for her feedback. With the test expert's final approval, the questions were included in the SPARK quiz assigned to students the following week. Table 1 shows select questions created by the team on the RUApAnEv sequence for Chapter 5's "disciplinary perspective" topic along with the SPARK quiz/unit test where they appeared.

Once content from a chapter started to be quizzed, that content continued to appear at increasingly higher cognitive levels on each of the following SPARK quizzes leading up to the unit test. This spiraling method resulted in an uneven distribution of chapter questions across Bloom's levels as seen in Table 2. Content from Chapters 1 and 5 were cycled across three quizzes and, therefore, were the most heavily reinforced each with a total of 22 questions. Chapters 4 and 8 content, on the other hand, were the final chapters to be introduced before each unit test and, therefore, were the least reinforced with only 11 questions each.

During Week 7, the midpoint of the semester, the study conditions were switched between the IDS course sections: Section 02 was assigned to the control condition, taking no more quizzes, while Section 04 was assigned to the SPARK condition with weekly SPARK quizzes on Chapters 5-8. Both groups submitted the same Unit Test 2 during Week 11 on content from Chapters 5-8. Everything else was held constant between the IDS course sections with respect to the instructor, course assignments, and format of teaching and learning experiences as conducted on the course LMS.

Once content from a chapter started to be quizzed, that content continued to appear at increasingly higher cognitive levels on each of the following SPARK quizzes leading up to the unit test. This spiraling method resulted in an uneven distribution of chapter questions across Bloom's levels.

Quizzes were completed by students asynchronously on the password-protected course LMS which presented questions on a SPARK quiz in a randomized order for each student. Students had access to an assigned quiz any time from 12:01 am on Monday to Sunday midnight during that week. When students started a quiz, they saw only one question on the screen. To minimize opportunities for students to work together during online administration of the quizzes, LMS quiz design features were used to prevent students from going back to a previous question after they entered a response. There was a time limit of two minutes per question each worth one point. Students were allowed only one attempt on each SPARK quiz. After they submitted their SPARK quiz, students were shown their responses and the correct answer to any questions they missed. There was additional explanation to some of the higher-level questions via a link; students had to click that link to see the explanation. The same format and procedures were used in administering the unit tests. Finally, on a voluntary end-of-study survey following the submission of Unit Test 2, students entered responses on how often they checked the correct answer feedback and reviewed material in their textbook for any of the questions they missed.

Table 1
The RUApAnEv Sequence for Chapter 5's Topic of "Disciplinary Perspective"

SPARK 5 (Ch. 5- Remember [R]). A discipline's perspective toward reality is reinforced by its favored phenomena, epistemology, assumptions, concepts, theories, methods, and data. What are these called? **a. A discipline's defining elements b. A discipline's defining approaches c. A discipline's defining characteristics	**UNIT TEST 2 (Ch. 5- Evaluate [Ev]). Read the following researchers' descriptions of their research agenda and answer the question.** **Dr. A:** My research covers drug design and synthesis. I develop molecules that interact with nucleic acids, which may lead to biological activity. My research may lead to treatment of a variety of diseases. My projects involve important collaborations with other scientists at [my university], at other universities and institutes, and in industry.
SPARK 5 (Ch. 5- Understand [U]). Which of the following is not appropriate for study by Natural Sciences disciplines? a. Complex phenomena b. Real world problems **c. Value-laden issues	**Dr. B:** My research explores 1) transportation as a social determinant of health and potential solutions to accessibility-related health disparities; 2) the role of shared mobility and micromobility (personal mobility devices such as bikes, scooters, skateboards, etc.) in a safe, sustainable and equitable post-pandemic urban transportation system; 3) socio-spatial isolation and segregation in individuals' daily lives and activity spaces; and 4) how mandatory inclusionary housing policies can help low-income households achieve home ownership and accumulate wealth.
SPARK 6 (Ch. 5- Apply [Ap]). Which of the following is the most accurate metaphor for describing disciplinary perspective? **a. Maslow's Hammer: If all you have is a hammer, then everything starts to look like a nail b. Fancy Smoothie: A smoothie blended to perfection with exotic and novel ingredients c. Golden Gate Bridge: A bridge to connect two pieces of land separated by a body of water	**Dr. C:** My research investigates the cognitive processes underlying improvisation and related therapeutic applications. Some of my research has shown that middle school children who receive training in musical improvisation score higher on measures of executive function compared to students who receive traditional music training.
SPARK 7 (Ch. 5- Analyze [An]). Read the following excerpt and answer the question. Dr. Tray's research work has focused on the biological and environmental factors that impact developing cognitive, learning and language systems in typically developing children and adults, and those with atypical development or acquired neurological disorders (including dyslexia, autism, ADHD, mitochondrial disease, brain tumors). His current projects include intervention studies focused on studying treatment outcomes of children and adults with dyslexia and reading disabilities, and related projects focused on using state-of-the-art technology to provide reading instruction where there are limited instructional resources. **Which of the following is most evident in this excerpt about Dr. Tray's disciplinary perspective?** **a. Natural sciences and social sciences perspectives b. Natural sciences and humanities perspectives c. Natural sciences and arts perspectives **Feedback:** Dr. Tray uses both biological/neurological and social science (cognitive psychology) perspectives in his research to understand outcomes of interventions he develops for those with reading/learning disabilities with a neurobiological origin.	**The researchers above differ in the breadth of their disciplinary perspective. Some use a broader perspective in their research than others. Which of the following is the most accurate rank ordering of these researchers' disciplinary perspective *from broadest to narrowest*?** **a. Dr. B > Dr. C > Dr. A b. Dr. B > Dr. A > Dr. C c. Dr. C > Dr. B> Dr. A d. Dr. C > Cr. A > Dr. B **Feedback:** With linkages between transportation, health, accessibility and socio-spatial effects of urban planning (urban planning) on daily functioning of individuals and the role that policies (political science) affect low-income households (sociology), Dr. B's perspective is broader than Dr. C's whose work combines cognitive processing as a result of music improvisation and that of Dr. A who has the narrowest perspective with a sole focus on chemical processes.

Note: The ** is used here to mark the correct answer.

Table 2
Distribution of SPARK Quiz and Unit Test Questions across Bloom's Levels by Chapter

	Remember (R)	Understand (U)	Apply (Ap)	Analyze (An)	Evaluate (Ev)	Total
Chapter 1	7	7	4	3	1	22
Chapter 2	6	6	3	2	1	18
Chapter 3	5	5	2	2	1	15
Chapter 4	4	4	1	1	1	11
Chapter 5	7	7	4	3	1	22
Chapter 6	6	6	3	2	1	18
Chapter 7	5	5	2	2	1	15
Chapter 8	4	4	1	1	1	11

Note: R refers to Remember, U to Understand, Ap to Apply, An to Analyze and Ev to Evaluate in the revised Bloom's Taxonomy of Thinking Levels.

Hypotheses

Based on findings from the test-enhanced learning literature and previous work on SPARK, we expected there would be significantly different unit test scores between the IDS sections favoring the SPARK condition over the non-SPARK control condition. We also expected to find significantly different unit test scores within sections through an analysis of each section's SPARK performance compared to their own non-SPARK control condition.

In addition, because of the unique design of our SPARK quizzes, we formulated two hypotheses of knowledge consolidation and we expected to see these effects on the Evaluate questions, the highest level of cognitive engagement in our SPARK quiz questions. We called our first hypothesis "knowledge consolidation through spaced, varied retrieval practice."

According to this hypothesis, we expected the most quizzed chapters to show the greatest number of correct responses on the Evaluate question. We reasoned that spaced retrieval of topics from most-quizzed chapters would experience the greatest consolidation in students' long-term memory organization as students kept engaging with those topics at cognitively higher levels of SPARK quiz questions presenting the content in varied contexts (Brown et al., 2014; Glass, 2009). We thus expected a descending pattern of accuracy scores on Evaluate questions across Chapters 1-4 with the highest number of students answering Chapter 1 Evaluate question correctly followed by Chapter 2 Evaluate question followed by Chapter 3 and Chapter 4 Evaluate questions. We anticipated this pattern of responding to replicate on Chapters 5-8 with Chapter 5 Evaluate question showing the highest accuracy scores followed by Chapters, 6, 7, and 8 Evaluate questions.

Based on findings from the test-enhanced learning literature and previous work on SPARK, we expected there would be significantly different unit test scores between the IDS sections favoring the SPARK condition over the non-SPARK control condition.

Alternatively, because chapter topics were related, as students started to develop a foundation of knowledge from the initial chapters, we anticipated this knowledge to aid their consolidation of knowledge gains on subsequent chapters. This was an artifact of our design: following the very first quiz, content from initial chapters was quizzed by increasingly higher-level questions together with newer and less cognitively complex content from subsequent chapters. As our students' understanding of topics from initial chapters began to be deepened and reinforced by increasingly higher-level SPARK questions, the knowledge network they developed around those topics might have allowed for an easier way to assimilate knowledge from subsequent chapters (Steyvers & Tenenbaum, 2005) and might have led to a stronger consolidation of knowledge covered in the subsequent chapters. This, we thought, might result in the reverse pattern of Evaluate question accuracy scores we outlined for our first hypothesis of knowledge consolidation. This formed our second hypothesis of knowledge consolidation which we called "knowledge consolidation through interleaving retrieval practice."

Results

Twenty students from Section 02 and 11 students from Section 04 consented to their data being used for this study for extra course credit. One of these students, enrolled in Section 02, did not submit Unit Test 1.

Analyses of Unit Test Means

Neither group did statistically better under the SPARK condition compared to their own non-SPARK control condition. For both groups, SPARK and non-SPARK means were statistically indistinguishable.

Because content coverage varied between Unit Test 1 (Chapters 1-4) and Unit Test 2 (Chapters 5-8), instead of a Repeated Measures ANOVA, we used separate independent or dependent-samples t-tests, where appropriate, to look for between- and within-group effects. Overall, Unit Test 2 scores were slightly lower compared to Unit Test 1 scores. There was a drop of .663 mean scores (SD = 3.377) from Unit Test 1 to Unit Test 2, which was not significant, $t(29) = 1.081$, $p = .288$. Mean differences between the sections were analyzed to identify differential effects of the SPARK method. First, group means were compared on Unit Test 1. Section 02 had a slight edge in Unit Test 1 means over Section 04; however, this difference was not statistically significant ($t < 1$). Similarly, Unit Test 2 means between the two sections were statistically indistinguishable as shown in Table 3 ($t < 1$). Our analyses of between-group differences thus revealed no effects of the SPARK method on unit test performance of IDS course sections.

Table 3
Group Means and SD on Unit Tests 1 and 2

EDUC 3100 Section	Unit Test 1 (max = 20)		Unit Test 2 (max = 20)	
	Mean	SD	Mean	SD
02 (n =19)	15.21	2.936	14.35	3.573
04 (n = 11)	14.82	3.060	13.91	3.754
Total (n = 30)	15.07	2.935	14.40	3.450

Note: SD = Standard Deviation

Next, we compared each section's performance under SPARK versus non-SPARK control conditions for a within-group analysis of SPARK effects by running dependent samples t-tests. Neither group did statistically better under the SPARK condition compared to their own non-SPARK control condition. For both groups, SPARK and non-SPARK means were statistically indistinguishable (all t's < 1).

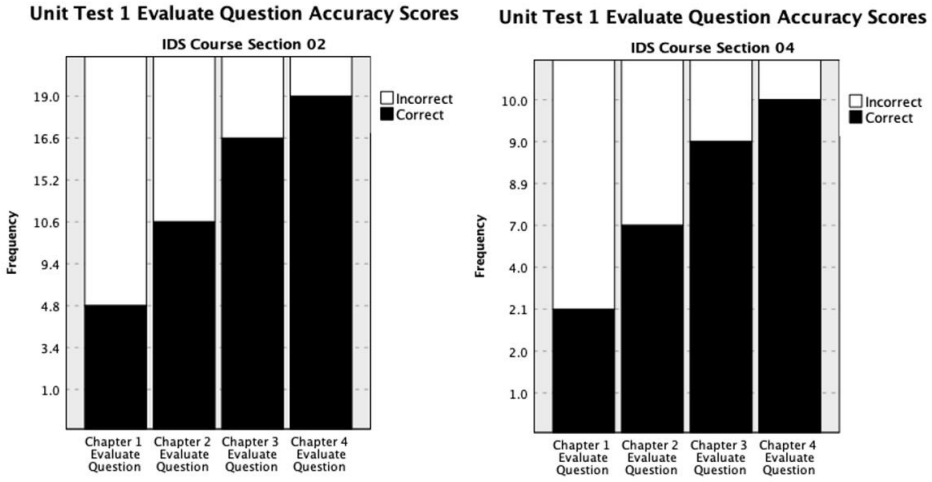
Analyses of Evaluate Question Accuracy Scores

To test our hypotheses of consolidation of knowledge gains from SPARK quizzes, we analyzed responses on Evaluate questions and expected our results to support either of the two hypotheses: consolidation through spaced, varied retrieval practice hypothesis or consolidation through interleaving retrieval practice hypothesis.

We first analyzed Evaluate questions from Unit Test 1. As seen in Figure 2, our results showed that the section that submitted SPARK quizzes on Chapters 1-4, Section 02, achieved the lowest accuracy scores on Chapter 1 Evaluate question and maintained an upward trajectory of accuracy scores on Evaluate questions across Chapters 2, 3, and 4, appearing to support the second hypothesis of knowledge consolidation through interleaving retrieval practice.

Because data on these Evaluate questions were entered dichotomously as correct or incorrect, we used the non-parametric Cochran's Q test to analyze proportions between Evaluate questions on Unit Test 1. There were 17 cases in Section 02 without missing data. The Cochran's Q test on Section 02 Unit Test 1 Evaluate questions was significant at $X^2(3) = 19.642$, $p < .001$ indicating there were significant differences between at least two of the Evaluate question proportions for Section 02 students. Post-hoc pairwise comparisons using Bonferroni

Figure 2
Evaluate Question Accuracy Scores of IDS Course Sections on Unit Test 1



adjustments showed the significant differences were between Chapter 4 Evaluate – Chapter 1 Evaluate questions and between Chapter 3 Evaluate – Chapter 1 Evaluate questions ($X^2(1) = .706$, $p < .0001$ and at $X^2(1) = .588$, $p = .005$, respectively).

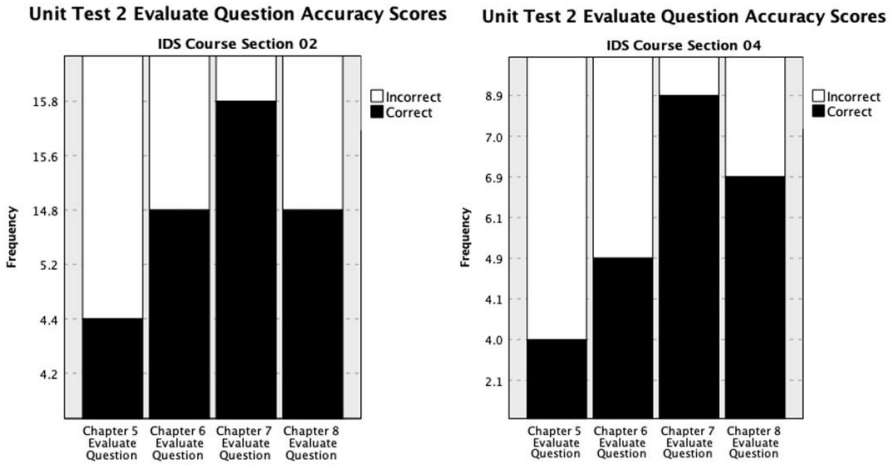
There appears to be support for the consolidation through interleaving retrieval practice hypothesis. However, attribution of this finding to SPARK quizzes is contingent upon Section 04 not showing a similar pattern of responding on the same Evaluate questions from Unit Test 1 as they were in the control condition during this time. As seen in Figure 2, a similar pattern was observed for Section 04's Evaluate questions on Unit Test 1 with a significant Cochran's Q statistic at $X^2(3) = 15.200$, $p = .002$ for the overall test of proportions of correct versus incorrect responses from 11 students who provided data for this section. Similar to Section 02, Bonferroni-corrected post-hoc tests of pairwise comparisons revealed significant differences between Chapter 4 Evaluate – Chapter 1 Evaluate questions ($X^2(1) = .727$, $p = .002$) and Chapter 3 Evaluate - Chapter 1 Evaluate question proportions ($X^2(1) = .636$, $p = .010$).

Next, as a test of the relationship between Evaluate question accuracy scores and the IDS sections, we performed separate Pearson Chi-square tests of independence on each Evaluate question on Unit Test 1. None of the Chi-square tests were significant at the .05 level suggesting no association between accuracy levels and IDS sections: $X^2(1, N = 29) = .068$, $p = .794$ for Chapter 1 Evaluate question; $X^2(1, N = 30) = .741$, $p = .389$ for Chapter 2 Evaluate question; $X^2(1, N = 30) = .036$, $p = .850$ for Chapter 3 Evaluate question; and $X^2(1, N = 29) = .133$, $p = .715$ for Chapter 4 Evaluate question. Due to a violation of the expected cell count assumption with expected cell counts less than 5 for all tests except for Chapter 2 Evaluate question, we followed up with Fisher's exact tests, which confirmed Chi-square tests of independence (i.e., 2-tailed $p = 1$ for Chapters 1, 3, and 4; 2-tailed $p = .466$ for Chapter 2) ruling out the hypothesis of SPARK-induced knowledge consolidation gains for Section 02.

Unlike positive results from previous in-person implementations of SPARK, we did not find evidence for SPARK-enhanced learning effects in our students enrolled in an online asynchronous course on interdisciplinarity.

The same pattern of results was observed for Unit Test 2 Evaluate questions. On Unit Test 2, there were 19 cases in Section 02 and 11 cases in Section 04 without missing data. Both sections exhibited similar patterns of proportions on Chapters 5-8 Evaluate questions. Other than Chapter 8, an upward trend of entering correct answers to Evaluate questions was evident for both sections as seen in Figure 3. However, only the distribution of Evaluate question proportions from Section 02, which was in the control condition during this time, was significant according to a Cochran's Q test ($X^23 = 17.000$, $p = .001$) with Bonferroni-corrected post-hoc tests showing significant differences between Chapter 5 Evaluate and Chapter 6 Evaluate questions ($X^2(1) = .526$, $p = .007$); between Chapter 5 Evaluate and Chapter 7 Evaluate questions ($X^2(1) = .579$, $p = .002$); and between Chapter 5 Evaluate and Chapter 8 Evaluate questions ($X^2(1) = .526$, $p = .007$). For Section 04, which experienced the SPARK quizzes during this time, there were no significant differences among the four Evaluate question proportions as the Cochran's Q test was not significant ($X^2(3) = 4.53$, $p = .209$).

Figure 3
Evaluate Question Accuracy Scores of IDS Course Sections on Unit Test 2



None of the Pearson Chi-square tests of independence were significant at the .05 level suggesting no association of accuracy scores on chapter Evaluate questions and IDS sections: $X^2(1, N = 31) = .992$, $p = .319$ for Chapter 5 Evaluate question; $X^2(1, N = 31) = 2.706$, $p = .100$ for Chapter 6 Evaluate question; $X^2(1, N = 30) = .036$, $p = .850$ for Chapter 7 Evaluate question; and X2(1, $N = 31$) = .445, $p = .505$ for Chapter 8 Evaluate question. Fisher's exact tests, which we ran due to the violation of the expected cell count assumption, confirmed these results with 2-tailed $p = .405$ for Chapter 5 Evaluate question; 2-tailed $p = .132$ for Chapter 6 Evaluate question; 2-tailed $p = 1$ for Chapter 7 Evaluate question; and 2-tailed $p = .683$ for Chapter 8 Evaluate question.

Our analyses of Evaluate questions failed to support either of two hypotheses of knowledge consolidation we derived from our design of SPARK quizzes.

Discussion

Unlike positive results from previous in-person implementations of SPARK (Hageman, 2016; 2020), we did not find evidence for SPARK-enhanced learning effects in our students enrolled in an online asynchronous course on interdisciplinarity. None of the between- and within-group comparisons of unit test scores were significant with group means nearly identical in all cases. Furthermore, our analyses of Evaluate questions failed to support either of two hypotheses of knowledge consolidation we derived from our design of SPARK quizzes. Accuracy levels on Evaluate questions were the lowest on Chapter 1 and exhibited an upward trajectory on the remaining chapters across Chapters 1-4 for both IDS course sections. These Evaluate response accuracy patterns were mirrored for Chapters 5-8 except for the Chapter 8 Evaluate question which departed from the upward trajectory possibly due to its difficulty. Thus, the overall upward trajectory of Evaluate question responses across both Chapters 1-4 and Chapters 5-8 appeared to support our second hypothesis of knowledge consolidation through interleaving retrieval practice. However, as our analyses revealed, these patterns of Evaluate responses were experienced to the same degree by both SPARK and non-SPARK conditions. There was no association between any of the Evaluate question accuracy scores and IDS sections. In the absence of a finding showing such an association, we were not able to attribute this finding to SPARK quizzes.

As these knowledge consolidation analyses suggest, course components other than SPARK quizzes appear to be responsible for the upward trajectory of gains we saw on the Evaluate questions. The only other assignment that was submitted by all students was the VoiceThread assignment where, provided with the slides of a PowerPoint covering important points from the assigned chapter, students navigated through the VoiceThread file, listening to the course instructor's voiceover narration (with Closed Caption transcription) which provided an elaborated discussion of the points on the slides. When they reached the final slide, they were presented with a "critical thinking/application" prompt to respond to. On Chapter 5 VoiceThread, for example, the prompt was: "Do you think that those working in the natural sciences (e.g., biology, chemistry, earth science, mathematics, physics) can transcend

their cultural experience and personal values and make definitive and objective measurements of a controversial phenomenon such as climate change? Why or why not?" Students were instructed to enter either a text (150-200 words) or an audio (1 min 30 s to 2 min) response thoughtfully addressing the prompt. The prompt was selected from each chapter's critical thinking/application questions section designed by the textbook authors to engage students in deep reflection and application of topics covered in the chapter. After posting their response to the prompt using either the text or audio option, students were asked to reply thoughtfully to at least one classmate's response.

It is likely that the underlying cognitive processes of VoiceThread assignment replicated some of the cognitive processes that were targeted by the SPARK quizzes and resulted in the overall upward trajectory of Evaluate question accuracy scores we saw for all students. As they viewed the prompt on the final slide of the VoiceThread, it is possible that students thought back and activated the knowledge they learned from the chapter, any of the resources they consulted from the chapter's content module compiled and posted by the instructional team on course LMS, and any other related information from previous chapters. It is thus possible that working on these weekly VoiceThread assignment submissions, students engaged in spaced retrieval of previous content in the context of the current chapter's content, interleaving and integrating knowledge across chapters. When they then sat down to formulate their thoughts into a constructed response, they engaged in consolidation of their learning, organizing their schema around the larger theme of interdisciplinarity. When they came back a few days later to reply to a classmate's response, they might have further reinforced their knowledge and mental representations through reflection and reactivation of topics they included in their own response.

These patterns of knowledge consolidation are consistent with cognitive theories of semantic preference which posit new content with the greatest number of associative connections to prior knowledge is likely to be learned faster and accommodated into the existing knowledge bases more strongly than new content with fewer associations.

These patterns of knowledge consolidation are consistent with cognitive theories of semantic preference (Steyvers & Tenenbaum, 2005) which posit new content with the greatest number of associative connections to prior knowledge is likely to be learned faster and accommodated into the existing knowledge bases more strongly than new content with fewer associations. Consolidation of knowledge gains we saw on chapter Evaluate questions might have been the result of a similar process where our course assignments (e.g., VoiceThread) allowed students to make a wide range of connections to content covered in previous chapters.

It is worth noting that with affordances to post audio and video responses, VoiceThread might have facilitated some of the in-person interactions students regularly have with their classmates and instructors during in-person meetings and class sessions. It is likely that posting an audio or video response on the VoiceThread assignment and interacting with other students' audio/video responses to discuss their agreements or disagreements replicated some of the components of in-person interactions that may help with elaboration and consolidation of knowledge gains. In other words, there may be a socioemotional component to memory processes of knowledge construction (see Tyng et al., 2017 for a review) that is inherent to in-person interactions between students but that is unlikely to be recreated in the online administration of SPARK quizzes without an audio/visual component. Created by Hageman (2016) before the pandemic, SPARK was designed to be implemented in a classroom setting in conjunction with group and whole class discussions of quiz questions. In our implementation of SPARK quizzes in an online asynchronous course, we were not able to hold similar discussion sessions.

Limitations

A significant limitation of our study was the small size of our sample. Although students were offered extra credit for their consent to let us use their data for study purposes, only 31 of a combined total of 49 students filled out the consent form resulting in a low participation rate which might have diminished the power of our analyses. On Chi-square tests of independence, for example, where we tested a relationship between IDS course sections and Evaluate question accuracy scores, the expected cell count assumption was violated with expected cell counts less than five for all Chi-square tests except for that of Chapter 2 Evaluate question. Although we were able to confirm Chi-square results with Fisher's exact tests, which are more robust for small sample size analyses, this limitation underscores the need for a larger sample in future replications of this study.

Additionally, other than the 13 students who submitted survey responses (10 from Section 02 and three from Section 04), we do not know how many of our participants went back to their textbook to review or restudy content. Although nearly all respondents reported they often or very often checked their answers and viewed feedback (92%), there was less consensus about returning to restudy topics for questions they answered incorrectly with 54% reporting doing so sometimes while 31% reporting often and 15% very often. Without another attempt allowed on the quiz, students might not have felt compelled to go back to their book for a review and restudy of topics. Explanations of answers provided on some of the higher-level questions were optional and were available to students via a link accompanying the question, which represented an additional step for students to take. Students might have chosen not to access this additional information in the absence of stronger incentives.

There was no evidence in our pilot study that SPARK quizzes supported consolidation of knowledge gains in college students enrolled in an online course on interdisciplinarity above and beyond existing course components.

Implications for Practice

Students taking online courses may be more likely to go back and restudy content that they have learned less well or click the link for feedback if, for example, they were offered more than one attempt at a quiz with the highest score counting toward their grade. They might be more willing to review less well-learned content in their book after results from an initial quiz attempt show which questions they missed without revealing the correct answer. This opportunity for review might be especially compelling if, after restudying and reviewing the material, students could then retake the quiz for a better grade. Previous research has shown college students benefit from multiple attempts on a quiz when they are given an opportunity to study in between attempts and have access to quiz questions and correct answer feedback (McDaniel et al. 2015). Online study group discussions on topics that are still unclear to students are likely to boost knowledge gains for students who are motivated by having multiple attempts to improve their quiz scores.

Conclusion

There was no evidence in our pilot study that SPARK quizzes supported consolidation of knowledge gains in college students enrolled in an online course on interdisciplinarity above and beyond existing course components (e.g., VoiceThread assignment). We believe more useful insights into effectiveness of quizzes like SPARK can be gained from future research studies that incorporate group or whole class discussions of quiz questions and use incentives (e.g., up to three attempts on the quizzes) that may motivate students to go back and revisit less well-learned content from assigned chapters. Any future replications should also recruit a larger sample size to ensure there is adequate power to data analyses.

Finally, we believe that there is an important role for AI-powered tools like ChatGPT (OpenAI, 2020) to play in future work on test-enhanced learning potential of quizzes. With capabilities to generate human-like coherent text, ChatGPT offers an unprecedented opportunity to create multiple-choice questions like those on SPARK and can be trained to align the questions it produces with the levels of thinking on Bloom's taxonomy. Instructors can input text or other topic-relevant content of up to 2,000 words into ChatGPT's chatbot and ask it to generate questions along the RUApAnEv sequence we used for our SPARK quizzes. Using ChatGPT in these ways can alleviate the burden on instructors and researchers who may be interested in using SPARK-like quizzes but are discouraged by the time demands required in creating multiple-choice questions. This represents a promising way forward in harnessing the potential of AI-tools like ChatGPT in future work on examining consolidated knowledge gains from test-enhanced learning experiences of college students.

AUTHOR BIO

Omer Ari
is a clinical associate professor in the Department of Middle and Secondary Education at Georgia State University's College of Education and Human Development. He has worked with struggling college readers and has designed and taught courses on interdisciplinarity. His research interests include psycholinguistic processes of reading and the effects of educational and reading interventions on college student success.

Kara Hageman
works at the University of Iowa as the senior science test developer and a faculty member in the Regents Alternative Program for Iowa Licensure program. After earning a bachelor's degree in biology and a master's in science education, she taught high school biology and chemistry for 10 years. In 2009, she began working at the University of Iowa developing the science assessments for the state of Iowa. She completed the doctorate program in educational psychology at the University of Iowa in 2020.

Andrea Lofgren
works as an assessment and accreditation officer at Gonzaga University. She has a doctorate in language and literacy from Middle and Secondary Education at Georgia State University. Her research examines factors influencing the success of undergraduate students experiencing poverty. She also examines assessment practices in higher education, particularly how assessment may be used for equity purposes.

References

Anderson, L. W., & Krathwohl, D. R. (2001). *A taxonomy for learning, teaching and assessing: A revision of Bloom's taxonomy of educational objectives*. Longman.

Beason-Abmayr, B., Caprette, D. R., & Gopalan, C. (2021). Flipped teaching eased the transition from face-to-face teaching to online instruction during the COVID-19 pandemic. *Advanced Physiology Education*, *45*, 384-389. https://doi.org/10.1152/advan.00248.2020

Brown, P. C., Roediger III, H. L., & McDaniel, M. A. (2014). *Make it stick: The science of successful learning*. The Belknap Press of Harvard University Press.

Butler, A. C. (2010). Repeated testing produces superior transfer of learning relative to repeated studying. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(5), 1118-1133. https://doi.org/10.1037/a0019902

Glass, A. L. (2009). The effect of distributed questioning with varied examples on exam performance on inference questions. *Educational Psychology*, *29*(7), 831-848. https://doi.org/10.1080/01443410903310674

Hageman, K. (2016, April 8-12). *The SPARK Testing Method: Using an assessment instrument to enhance learning* [Conference presentation]. AERA 2016 Annual Meeting, Washington, D.C.

Hageman, K. (2020). *Performance of 8th-grade Algebra I students assessed with and without the SPARK Testing Method*. Unpublished Dissertation, University of Utah.

Jensen, J. L., McDaniel, M. A., Woodard, S. M., & Kummer, T. A. (2014). Teaching to the test… or testing to teach: Exams requiring higher order thinking skills encourage greater conceptual understanding. *Educational Psychology Review*, *26*(2), 307-329. https://doi.org/10.1007/s10648-013-9248-9

Lawson, T. J. (2022). Effects of cued-recall versus recognition retrieval practice on beginning college students' exam performance. *College Teaching*, *70*(2), 181-185. https://doi.org/10.1080/87567555.2021.1910124

Lyle, K. B., & Crawford, N. A. (2011). Retrieving essential material at the end of lectures improves performance on statistics exams. *Teaching of Psychology*, *38*(2), 94-97. https://doi.org/10.1177/0098628311401587

Mayer, R. E., Stull, A., DeLeeuw, K., Almeroth, K., Bimber, B., Chun, D., & Zhang, H. (2009). Clickers in college classrooms: Fostering learning with questioning methods in large lecture classes. *Contemporary Educational Psychology*, *34*(1), 51-57. https://doi.org/10.1016/j.cedpsych.2008.04.002

McDaniel, M. A., Wildman, K. M., & Anderson, J. L. (2012). Using quizzes to enhance summative-assessment performance in a web-based class: An experimental study. *Journal of Applied Research in Memory and Cognition*, *1*(1), 18-26. https://psycnet.apa.org/doi/10.1016/j.jarmac.2011.10.001

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*, 360-372. https://doi.org/10.1002/acp.2914

McDaniel, M. A., Bugg, J. M., Liu, Y., & Brick, J. (2015). When does the test-study-test sequence optimize learning and retention? *Journal of Experimental Psychology: Applied*, *21*(4), 370-382. https://doi.org/10.1037/xap0000063

OpenAI. (2020). ChatGPT powered by GPT-3 (3rd generation language model) [Computer software]. OpenAI. https://www.openai.com/gpt-3/

Repko, A. F., Szostak, R., and Buchberger, M. P. (2019). *Introduction to Interdisciplinary Studies*. *3rd ed*. Sage.

Richmond, A. S., Ropp, A., Bradford, J., Ignizio, G. S., Hammond, J., Mowder, D., & Bittmann, J. M. (2022). An ecologically valid study of the testing effect across academic disciplines: A focus on higher- vs. lower-level learning, *College Teaching*. Advance online publication. https://doi.org/10.1080/87567555.2022.2090488

Steyvers, M. & Tenenbaum, J. B. (2005). The large-scale structure of semantic networks: Statistical analyses and a model of semantic growth. *Cognitive Science: A Multidisciplinary Journal*, *29*(1), 41-78. https://doi.org/10.1207/s15516709cog2901_3

Thomas, R. C., Weywadt, C. R., Anderson, J. L., Martinez-Papponi, B., & McDaniel, M. A. (2018). Testing encourages transfer between factual and application questions in an online learning environment. *Journal of Applied Research in Memory and Cognition*, *7*(2), 252-260. https://doi.org/10.1016/j.jarmac.2018.03.007

Trumbo, M. C., Leiting, K. A., McDaniel, M. A., & Hodge, G. K. (2016). Effects of reinforcement on test-enhanced learning in a large, diverse introductory psychology course. *Journal of Experimental Psychology: Applied*, *22*, 148-160. https://doi.org/10.1037/xap0000082

Tyng, C. M., Amin, H. U., Saad M. N. M., & Malik, A. S. (2017). The influences of emotion on learning and memory. *Frontiers in Psychology*, *8*, 1454. https://doi.org/10.3389/fpsyg.2017.01454

VoiceThread LLC. (n.d.). VoiceThread [Computer software]. https://voicethread.com/