Journal of Applied Research in Memory and Cognition

Investigating the Intensity and Integration of Active Learning and Lecture

Amedee Marchand Martella, Darryl W. Schneider, Garrett M. O’Day, and Jeffrey D. Karpicke

Online First Publication, February 1, 2024. https://dx.doi.org/10.1037/mac0000160

Journal of Applied Research in Memory and Cognition

ISSN: 2211-3681

https://doi.org/10.1037/mac0000160

EMPIRICAL ARTICLE

Investigating the Intensity and Integration of Active Learning and Lecture

Amedee Marchand Martella[1], Darryl W. Schneider[2], Garrett M. O'Day[2], and Jeffrey D. Karpicke[2]
[1] Department of Psychological and Brain Sciences, University of California, Santa Barbara, United States
[2] Department of Psychological Sciences, Purdue University, United States

Given that the active learning literature lacks systematic investigations on how the intensity and integration of lecture and active learning affect learning, we conducted two experiments to examine the impact of these variables. The first experiment involved 146 participants who learned about biological taxonomies through pure lecture or pure active learning. Participants in the pure lecture condition scored significantly higher on a posttest than those in the pure active learning condition. The second experiment involved 219 participants who learned about biological taxonomies through pure lecture, a lecture and active learning activity that were interspersed, or a lecture and active learning activity that were blocked. Participants in the interspersed condition scored significantly higher than participants in the blocked and pure lecture conditions (which did not significantly differ). Based on these experiments, it may not be a question of either/or but rather a question of how to integrate lecture and active learning.

General Audience Summary

Should we teach college courses using lecture or active learning? Perhaps it is a matter of both. In two experiments, we systematically studied the effects of different intensities of active learning and ways to integrate active learning and lecture to determine which maximized student learning of science content. Our findings suggest that (a) a pure lecture intervention may encourage greater student learning of science content than a pure active learning intervention, and (b) interspersing equal amounts of lecture and active learning may encourage greater student learning of science content than blocking the instructional modes or having lecture alone. Therefore, it is important to consider how to integrate lecture and active learning to improve student learning in Science, Technology, Engineering, and Mathematics courses.

Keywords: active learning; lecture; Science, Technology, Engineering, and Mathematics education; effective instruction

Supplemental materials: https://doi.org/10.1037/mac0000160.supp

The prominent mode of instruction in higher education courses continues to be the traditional lecture method (Stains et al., 2018), which has been criticized for promoting passive learning (Deslauriers et al., 2019). Numerous university- and national-level efforts have encouraged moving courses from passive learning to active learning in an attempt to improve Science, Technology, Engineering, and Mathematics (STEM) education and reduce STEM disparities (e.g., Association of American Universities, 2017; Center for STEM

Amedee Marchand Martella https://orcid.org/0000-0002-9207-4300
Darryl W. Schneider https://orcid.org/0000-0002-0916-4269

Amedee Marchand Martella acknowledges support from the Division of Graduate Education, National Science Foundation Graduate Research Fellowship Program, under Grant DGE-1842166, and the Directorate for STEM Education, National Science Foundation Postdoctoral Research Fellowship Program, under Grant 2222208. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation. The authors have no conflicts of interest to disclose.

All data and materials for this experiment are available on the Open Science Framework at https://osf.io/sk8p4/files/osfstorage. This study has been preregistered on the Open Science Framework website at https://doi.org/10.17605/OSF.IO/EDBJU. This research was based on Amedee Marchand Martella's doctoral dissertation, supervised by Darryl W. Schneider and Jeffrey D. Karpicke.

Amedee Marchand Martella played a lead role in conceptualization, data curation, formal analysis, funding acquisition, investigation, methodology, project administration, visualization, and writing–original draft and an equal role in writing–review and editing. Darryl W. Schneider played a lead role in supervision, a supporting role in conceptualization, formal analysis, and methodology, and an equal role in writing–review and editing. Garrett M. O'Day played a supporting role in data curation, formal analysis, visualization, and writing–review and editing. Jeffrey D. Karpicke played a supporting role in conceptualization, formal analysis, methodology, supervision, and writing–review and editing.

Correspondence concerning this article should be addressed to Amedee Marchand Martella, Department of Psychological and Brain Sciences, University of California, Santa Barbara, Building 251, Psychology Office 3235, Santa Barbara, CA 93106, United States. Email: amedeemartella@ucsb.edu

Learning, 2016). In the active learning literature, *passive* and *active* typically refer to students' overt behaviors, often with the assumption that participating in class activities (high behavioral activity) leads to high cognitive activity and listening to a lecture or other instructor-led demonstration (low behavioral activity) leads to low cognitive activity. Researchers have investigated a broad collection of instructional methods that focus on overt learner activity. For example, the interactive, active, constructive, and passive model (Chi, 2009; Chi et al., 2018; Chi & Wylie, 2014) distinguishes among different levels of learner activity: passive, which involves receiving information with no overt learner activity; active, which involves low-level activity such as highlighting information; constructive, which involves generating new information or inferences such as writing a summary; and interactive, which involves interacting with another person such as participating in partner discussions.

Although educational scholars may agree that active learning involving high cognitive engagement should be cultivated in the classroom (Fiorella & Mayer, 2015; Mayer, 2002, 2011), they do not necessarily agree on what it looks like as an instructional approach (Martella et al., 2021; Martella & Schneider, in press; Zakrajsek, 2018). As such, active learning is an umbrella term that encompasses an array of classroom implementations and intensities (Freeman et al., 2014; Lombardi et al., 2021; Martella et al., 2021), making it difficult to determine how to effectively design an active learning course. In particular, it is unclear how much lecture—if any—should be included in active learning courses. In Freeman et al.'s (2014) meta-analysis on active learning, active learning courses ranged from having 0%–90% of class time dedicated to lecture, with no evaluation as to whether the effectiveness of active learning was moderated by the proportion of class time dedicated to active learning activities. It is important to note that some courses devoted 100% of class time to activities, but they were not treated differently than courses that devoted just 10% of class time to activities.

A long-standing controversy in the field involves finding the appropriate balance between teacher-led instruction and learner activity (or student-centered instruction; Mayer, 2004, 2011, 2021). On the one hand, claims have been made that lecture is the "pedagogical equivalent of bloodletting" (Wieman, 2014, p. 8320), and many departments and institutions have been working to reduce or phase out lecture (see Dawson, 2015; Massachusetts Institute of Technology, 2021; University of Georgia Office of Instruction, 2021). On the other hand, some researchers say that "there are still times when lectures will be needed" (Noah Finkelstein quoted in Bajak, 2014, para. 7) and that whether one should promote student activity or lecture is "a matter of *both*, not one or the other" (Opdal, 2022, p. 16). The literature on active learning provides limited insight into how instructor-led instruction and student-centered activities should be balanced. Therefore, it is important to systematically study the intensity of the instructional modes, which refers to the percentage of class time dedicated to active learning activities and lectures. The intensity of active learning can range between the extremes of 0% time spent on activities (100% time spent on lecture) and 100% time spent on activities (0% time spent on lecture), or fall somewhere in between (e.g., 50% lecture and 50% activities). Although it is unclear what the optimal amounts of lecture and active learning are within the classroom, both lecture and active learning activities afford certain benefits.

There are several theoretical reasons why instructor-led instruction is beneficial. First, direct instruction can remove the discovery process that can lead to errors that place high demands on working memory (Sweller, 2004). Second, direct instruction can help reduce cue overload—which is the idea that as more items become associated with a retrieval cue, the more difficult it is for that cue to bring to mind any particular associated item (Surprenant & Neath, 2009; Watkins & Watkins, 1975)—by making direct connections between concepts for students. Third, direct instruction can help connect new information to students' prior knowledge because instructors can reference past information or sequence information such that each new concept builds upon the previous one. Moreover, direct instruction can support the learning of new information by novice learners who lack existing knowledge structures for encoding information in long-term memory (Sweller et al., 1998).

Active learning in the form of student activities can also confer many benefits. First, active learning can provide expert learners with experiences that eliminate the redundancy a full lecture might provide, thereby avoiding having working memory overloaded by redundant information (see Sweller et al., 2011). Moreover, engaging expert learners in minimally guided problem solving can support their learning and retention (Sweller et al., 2011), in part because activities with added difficulty (e.g., generating solutions) can be desirable for them (see Bjork & Linn, 2006). Second, active learning affords students the opportunity to engage in practice activities that allow information to be processed more deeply (see Fiorella & Mayer, 2015, 2016) and strengthened in long-term memory through repeated exposure and retrieval (Carpenter, 2017; Karpicke & Roediger, 2007). Third, active learning affords the opportunity to provide feedback on students' understanding and performance, which is important because feedback can help students avoid acquiring misinformation and promote the encoding of correct information. Research shows that feedback conditions result in better retention than no-feedback conditions (e.g., Butler & Roediger, 2008; Karpicke & Roediger, 2010; Lhyle & Kulhavy, 1987).

Given that both lecture and active learning activities can afford students important learning experiences, it is critical to systematically study how the intensity of lecture and active learning affects student learning. An important second factor to consider besides intensity is integration schedule. When integrating lectures and active learning activities, is it more effective to intersperse or block them? In an *interspersed schedule*, students receive a minilecture and a minilearning activity over the first concept followed by a minilecture and minilearning activity over the second concept, and so forth. In a *blocked schedule*, students are presented with a full lecture over multiple concepts and then a full active learning activity over these concepts.

Research on the placement of practice tests and other retrieval-based activities offers initial insight into whether lecture, and active learning should be interspersed or blocked. Weinstein et al. (2016) compared the effects of providing students with short-answer practice questions throughout lectures or providing them with these practice questions at the end of the lectures, finding an advantage for interspersing practice questions on initial performance but not on final performance. Uner and Roediger (2018) compared the effects of taking practice tests after each section of a chapter, after the whole chapter, or both, and found the placement of the initial tests to produce comparable benefits. Further, Healy et al. (2017) examined how interspersing or blocking quizzes affected student performance and found a benefit for interrupting learning with quiz questions (i.e., interspersing). Other researchers have examined interspersed (also called *interpolated*) practice test conditions as compared to no-practice test conditions or restudy conditions and have found interspersing to reduce mind wandering (Szpunar et al., 2013), encourage more focused attention (Jing et al., 2016), and improve learning from online/video lectures (Jing et al., 2016; Szpunar et al., 2013). These results provide initial insight into the effects of interspersing or blocking activities albeit the results tend to be mixed on whether one is more effective than the other. Additionally, these studies typically use practice tests, whereas our interest was in other forms of active learning activities that can be used to help with encoding and retrieval. Therefore, the present study was designed to specifically address open questions within the active learning literature surrounding how to best integrate lecture and active learning.

The purpose of the present study was to systematically compare the effects of lecture and active learning on novice student learning in a science domain, both as individual approaches and as integrated approaches. The first experiment investigated intensity by comparing a pure lecture condition (i.e., 100% lecture/0% active learning activities) to a pure active learning condition (i.e., 100% active learning activities/0% lecture), thereby allowing for a cleaner comparison of the effectiveness of the two approaches than what typically occurs in active learning research. The second experiment investigated how integrating lecture and active learning in equal intensities but with different integration schedules (interspersed or blocked) affected student learning, particularly in comparison to the pure lecture condition from Experiment 1. This experiment thereby allowed for a systematic study of the integration of the two approaches, which is lacking within the active learning literature.

Experiment 1

Experiment 1 included two conditions in which undergraduate students learned biology content. In the pure lecture condition, participants learned taxonomic associations (i.e., classifications of organisms) through a lesson involving a video lecture. In the pure active learning condition, participants learned taxonomic relationships through a lesson that involved a matching activity. Participants indicated how much they thought they had learned from their lesson and how much they enjoyed the instructional method they received. They then took a posttest 5 min after the lesson to assess their learning of the lesson content; the posttest contained questions that assessed directly and indirectly learned associative pairs. There was one primary research question and two exploratory research questions that were investigated in the present experiment:

1. Primary research question: Does the intensity of active learning impact participants' learning of science content?
2. Exploratory Research Question 1: Does the intensity of active learning impact how much participants feel they learned from the lesson?
3. Exploratory Research Question 2: Does the intensity of active learning impact how much participants enjoyed the lesson?

Method

Transparency and Openness

We report how we determined our sample size, all data exclusions (if any), all manipulations, and all measures in the experiment.

Participants and Design

One hundred seventy introductory psychology undergraduate students from a large public Midwestern university participated for partial course credit (research ethics approval was obtained for our study through the institutional review board); however, 24 participants were excluded (11 in the pure lecture condition and 13 in the pure active learning condition) due to being under 18 years of age, failing the lecture-viewing check(s) during the experiment, not completing Part 1 of the activity, and/or admitting to using notes or other aids on the posttest. The final analytic sample included 146 participants; see Table 1, for demographic information. These participants were randomly assigned to one of the two conditions (pure lecture or pure active learning) as part of a between-participants design; 73 participants were in each condition.

Including 73 participants in each condition was based on an a priori power analysis conducted using G*Power 3.1.9.6 (Faul et al., 2007) that indicated 73 participants were required to achieve 80% power ($\alpha = .05$, two-tailed test) to detect a between-participants difference with an effect size of $d = 0.47$, which was the effect size for a similar comparison (i.e., active learning vs. lecture) reported in the meta-analysis on active learning by Freeman et al. (2014). It is important to note that the effect size obtained from a similar pilot study was .88 (see Supplemental Pilot Study); therefore, using the effect size of .47 for this power analysis provided a more conservative sample size estimate.

Materials and Measures

Lesson Content. Participants in both conditions learned how to categorize five organisms (roundworm, red kangaroo, water flea,

Table 1
Participant Demographics

Demographic	Experiment 1	Experiment 2
Age	18.79	19.02
Sex		
Female	67.12	47.49
Male	32.88	52.05
Other	0.00	0.46
Race/ethnicity		
Caucasian	78.08	74.43
Asian	15.07	14.16
Black	2.74	4.57
Hispanic	2.05	2.28
Multiracial	1.37	2.28
Other	0.69	2.28
First language		
English	93.15	94.52
Other	6.85	5.48

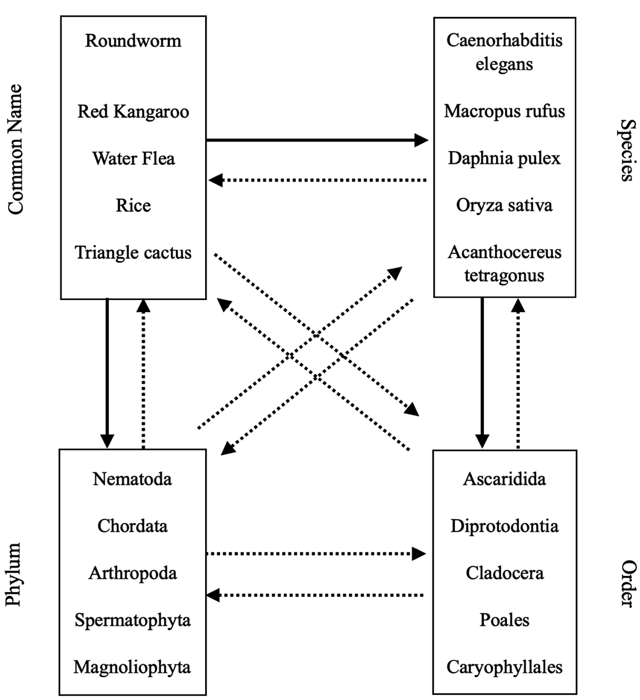
Note. Age is reported as the mean age in years. All other values are percentages.

rice, and triangle cactus) according to four different taxonomic ranks (common name, species, order, and phylum). The lesson content was intentionally chosen and designed to achieve three aims. First, having separate taxonomic relationships to learn allowed for the precise manipulation of intensity and integration schedule (which was particularly important for Experiment 2). Second, consistency in the content in both conditions was ensured by having the exact same relationships learned through explicit connections made by the instructor or identified through matching exercises. Third, the lesson content reflects a major concept (i.e., biological taxonomies) found within science curricula. Therefore, the content was within a STEM domain and helped to minimize potential confounds that might undermine comparisons of different instructional approaches or integrations of such.

There were three primary taxonomic relationships learned in the lesson (common name → species, species → order, and common name → phylum), and each taxonomic relationship had five to-be-learned associative pairs based on the five organisms (e.g., roundworm → *Caenorhabditis elegans*; red kangaroo → *Macropus rufus*; water flea → *Daphnia pulex*; rice → *Oryza sativa*; and triangle cactus → *Acanthocereus tetragonus* [common name → species]). In total, participants learned 15 associative pairs (Five Pairs per Taxonomic Relationship × Three Taxonomic Relationships; see Figure 1, for an overview). The first taxonomic relationship was common name → species, the second was species → order, and the third was common name → phylum. Rather than designate the third taxonomic relationship as order → phylum, which would occur as part of the typical taxonomic sequence, the third relationship was chosen to be common name → phylum to add complexity to the lesson by requiring the inference of order → phylum based on learning the three other relationships. By omitting the teaching of order → phylum, we could see how well-integrated participants' knowledge was of the different organisms and taxonomic categorizations on the posttest and if they could generate this knowledge under different instructional conditions. Therefore, these materials allowed us to examine the retention of the content that was directly learned as well as the inferences that were made about the directly learned content.

Figure 1

Direct and Indirect Taxonomic Associations for Five Organisms



Note. Solid arrows represent directly learned associations. Dotted arrows represent generated (i.e., indirectly learned) associations.

Overview Video. Participants in both conditions received an overview video that discussed what a taxonomy is and what each taxonomic rank represents, with representative examples given for each ranking. The purpose of this video was to embed the lesson content into a larger context. This video lasted approximately 5.5 min and was developed by the lead author who has a background in biology. Note that despite the initial presentation of the overview video, the active learning condition still reflected "pure" active learning because the overview video was a general/high-level introduction to provide context for participants and did not address any of the specific taxonomic relationships covered in the lecture or matching activity. Moreover, the information in the overview video was not on the posttest nor would specifically prove helpful on the posttest.

Recorded Lecture Presentation. Participants in the pure lecture condition received an 18-min lecture that completely explained how each of the five organisms is categorized according to species, order, and phylum. More specifically, there were three primary taxonomic relationships discussed in the lecture in the following order: common name → species, species → order, and common name → phylum. This phase of the lecture totaled 9 min. Within each relationship, the lecturer taught five associative pairs (one for each organism), and Greek and Latin roots were discussed to help connect the different scientific names to common knowledge. For example, for common name → species, the lecturer explained how red kangaroo is the common name for the picture shown of the red kangaroo, and *Macropus rufus* is its species name. The lecturer then discussed Greek and Latin roots that would help participants remember that the red kangaroo is of the species *Macropus rufus* (*Macropus* means "long foot" and *rufus* means "red-haired"). Roots typically did not overlap among the organisms—for two of the phyla names, the ending roots overlapped but the beginning roots were distinct. A review (totaling 9 min of the 18-min lecture) of each associative pair was presented after the lecture on each taxonomic relationship and at the end of the lecture. This review was presented in a matching-activity format where the instructor connected terms on the left side of the slide (e.g., common names) to terms on the right side of the slide (e.g., species names) with an arrow that designated the association. During this passive matching review, the instructor first asked participants to think about which terms should be connected to form an associative pair and then provided the answer to participants by matching the correct terms on the slide.

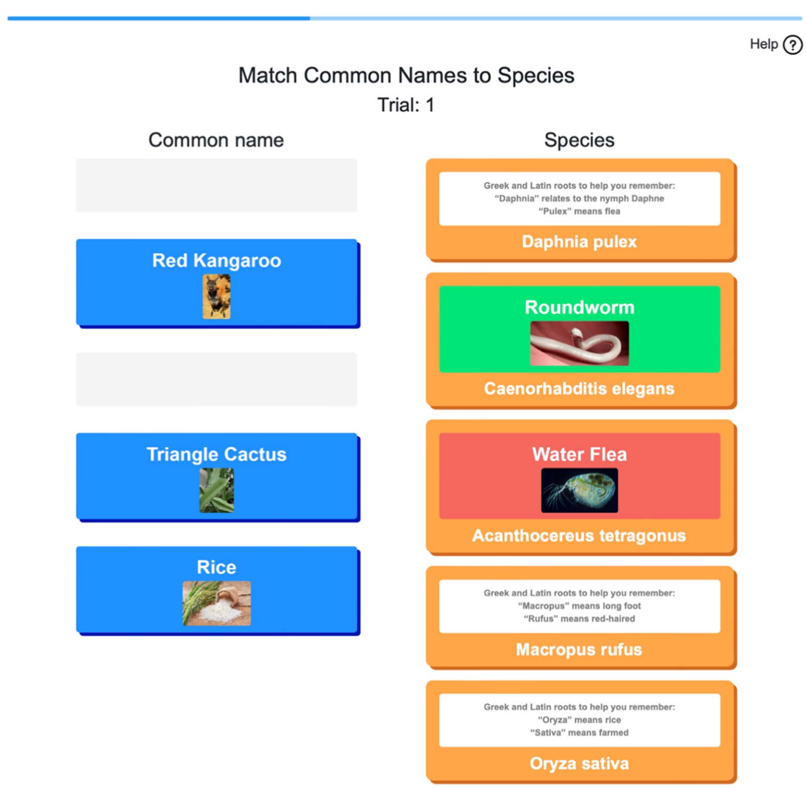
This condition reflected pure lecture in that the lecturer provided a complete explanation of the 15 associative pairs for the full 18 min. Participants did not receive any explicit opportunities for active behavioral participation. The lecture video was a recorded PowerPoint presentation that had the lecturer (the first author) present alongside the slides to make the experience more educationally authentic. Further, there were three "are you watching?" questions embedded throughout the lecture (after each taxonomic relationship had been discussed) to help determine if participants had viewed the full 18-min presentation; participants had 15 s to respond to these questions and had to select "yes."

Active Learning Activity. Participants in the pure active learning condition received an 18-min active learning activity (a matching activity; see Figure 2) to learn the same 15 associative pairs in the same order as in the recorded lecture presentation (i.e., common name → species, species → order, common name → phylum). The difference was that they learned these associative pairs through direct interaction in the matching activity and without any direct instruction. For example, for common name → species, participants dragged and dropped common names into bins labeled with their associated species names. More specifically, participants completed matching trials for each taxonomic relationship whereby they created an association between two taxonomic pairs for each organism (e.g., dragging "red kangaroo" into the bin for "*Macropus rufus*"). The same Greek and Latin roots provided in the lecture were provided below each species, order, and phylum name in the activity. In the activity, each taxonomic relationship (and its five associative pairs) had to be correctly matched before advancing to the next taxonomic relationship (i.e., the associative pairs for common name → species had to be correctly matched before participants could practice matching pairs for species → order). If an association was incorrect, participants received immediate feedback in the form of an unpleasant "ding" noise and a red light. If an association was correct, they received immediate feedback in the form of a pleasant "ding" noise and a green light. Participants needed to engage in rematching until all five associative pairs were correct before they could move on to matching the associative pairs for the next taxonomic relationship. If participants were struggling with technical aspects of how to complete the matching game, they could click on a "help button" that provided instructions on how to drag and drop terms to complete a match.

Participants were challenged to complete two parts of the activity: Part 1 was to get each associative pair correct on the first try, four times in a row for it to be removed from the activity, and Part 2 was to get all five associative pairs correct as one unit (i.e., all five associative pairs had to be correctly matched on the first try or the matching trial would be considered a failed trial) for each taxonomic relationship on the first try, four times in a row for all of the associative pairs to be removed from the activity. It is important to note that in Part 1 of the activity, individual pairs that had been correctly matched four times in a row were removed from the game, and individual pairs that had not been correctly matched four times in a row remained in the game until this criterion had been met.

The activity was presented as a challenge to see how many matching trials it would take participants to complete Parts 1 and 2. It is important to note, however, that Part 1 was considered the main

Figure 2
Matching Game Example



Note. See the online article for the color version of this figure.

gamified element in that participants who succeeded in the challenge of completing Part 1 would have reached a high success rate before moving on to the posttest (i.e., they would have correctly matched each associative pair correctly four times in a row). Part 2 was created to keep participants engaged if they finished Part 1 before the activity time had elapsed. If participants completed both parts of the activity before the 18 min had elapsed, they reentered Part 1. Participants' trial-by-trial accuracy was shown for each associative pair during the activity to help them keep track of their performance (see Figure 3).

This matching activity falls under the active learning umbrella for the following reasons. First, *active learning* may be best characterized as an event(s) in which participants interact with class content through participatory activities (see Driessen et al., 2020; Freeman et al., 2014; Lombardi et al., 2021). The matching activity requires behavioral participation for the full 18 min in order for participants to move on to the posttest. Participants interacted with the content by forming associations through the dragging and dropping of different taxonomic terms into different bins. For example, they would drag the common name red kangaroo over to the species bin labeled *Macropus rufus*.

Second, active learning is often said to be cognitively engaging in that participants are constructing their own knowledge and doing more than passively listening. To ensure participants did not mindlessly drag and drop terms for the duration of the 18-min activity, they were challenged to complete the activity at least once in the 18 min before moving on to the next phase of the experiment. To complete the activity successfully, participants needed to learn 15 associative pairs between 30 scientific names and get each associative pair correct on the first try, four times in a row; therefore, they needed to actively engage with the content and acquire this knowledge to be successful.

Third, games are often used in active learning courses to promote active learning. In fact, games are an identified active learning activity category (Driessen et al., 2020; Martella et al., 2021). Various types of matching games/activities have been implemented in active learning classrooms (e.g., McCarroll et al., 2009; Nuetzman & Abdullaev, 2012).

Prior Knowledge Questions. Although the lesson content was chosen because it is not typical common knowledge, participants received five prior knowledge questions to gauge their familiarity with the content. Four of these questions were asked before they listened to the lecture presentation or participated in the active learning activity; these questions represented "general prior knowledge." The fifth question was asked after the lecture or activity; this question represented "specific prior knowledge" in that it was directly related to the specific content within the lecture and active learning activity. The questions and their order were:

1. "On a scale of 0%–100%, how well do you know what a taxonomy is?"
2. "On a scale of 0%–100%, how well do you know the different levels of categorization of a taxonomy?"
3. "On a scale of 0%–100%, how strong is your background in biology?"
4. "On a scale of 0%–100%, how well do you know Greek and Latin roots?"
5. "How many of the five organisms could you categorize based on species, order, and phylum before this experiment?"

Judgment of Learning Question. Participants received one judgment of learning (JOL) question to gauge their beliefs about how well they learned the lesson content. They were asked this question after they listened to the lecture or participated in the active learning activity. The question was: "On a scale of 0%–100%, how well do you think you learned the phylum, order, and species name for the five organisms from this experiment?"

Instructional-Mode Enjoyment Question. Participants received one "instructional-mode enjoyment" question to gauge how enjoyable they believed the lecture or activity was. They were asked this question after they listened to the lecture or participated in the active learning activity. Depending on the condition, the question was: "On a scale of 0%–100%, how much did you enjoy the activity?" or "On a scale of 0%–100%, how much did you enjoy the lecture?"

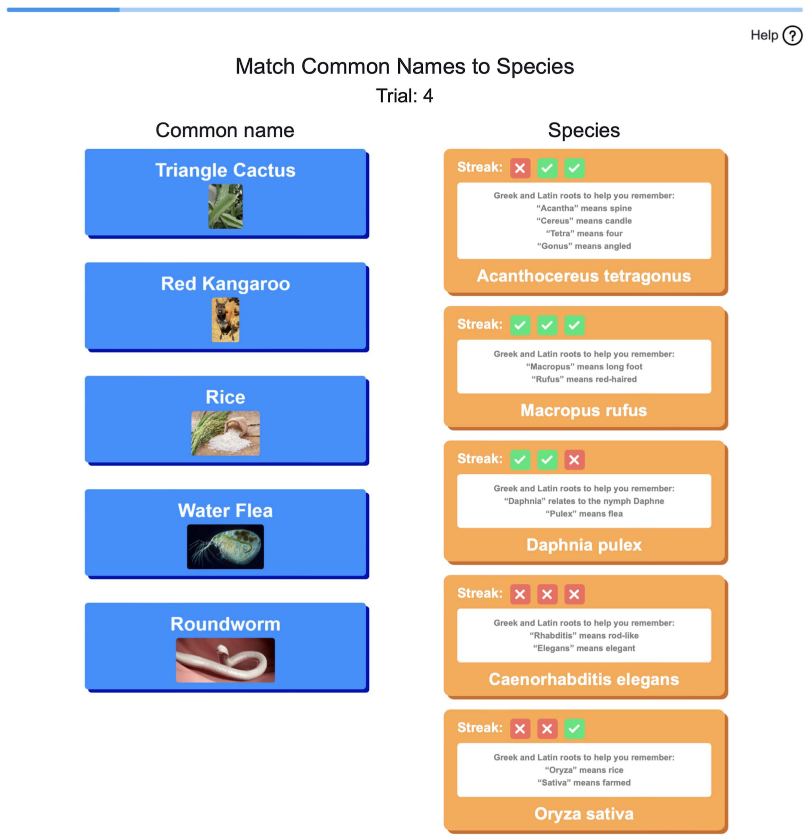
Multiple-Choice Posttest. Participants in both conditions received a 60-item multiple-choice posttest assessing their learning of the three taxonomic relationships.[1] The order of the questions was randomized for each participant. Each item had five answer options. An example question was: "*Macropus rufus* is the species name for which organism?" To completely measure participants' learning of the content, the posttest included questions designed to assess all of the possible associations that could be formed from learning the three taxonomic relationships. More specifically, it contained 15 verbatim questions and 45 inference questions. Verbatim questions provided (a) the common name of one of the five organisms and asked for the species name (common name → species); (b) the species name of one of the five organisms and asked for the order name (species → order); or (c) the common name of one of the five organisms and asked for the phylum name (common name → phylum). For example, one of the posttest questions was: "Triangle cactus is the common name for which species?" and there were five answer options: "*Caenorhabditis elegans*," "*Acanthocereus tetragonus*," "*Macropus rufus*," "*Daphnia pulex*," and "*Oryza sativa*." Participants learned these specific associative pairs during the lecture or activity (see solid arrows in Figure 1). Inference questions followed the same format but represented indirectly learned associations: order → phylum, phylum → order, order → common name, common name → order, phylum → species, species → phylum, species → common name, order → species, and phylum → common name (see dotted arrows in Figure 1). For example, one of the posttest questions was: "The scientific order Cladocera contains which species?" and there were five answer options: "*Caenorhabditis elegans*," "*Acanthocereus tetragonus*," "*Macropus rufus*," "*Daphnia pulex*," and "*Oryza sativa*." For these test questions, the relationships were not directly learned but could be generated based on knowing the associative pairs directly taught in the lesson. Therefore, these questions assessed generative learning.

Procedure

We implemented a between-participants experimental design with two levels: 100% lecture and 100% active learning.

[1] See Supplemental Material for the multiple-choice question set.

Figure 3
Trial Accuracy Feedback Example



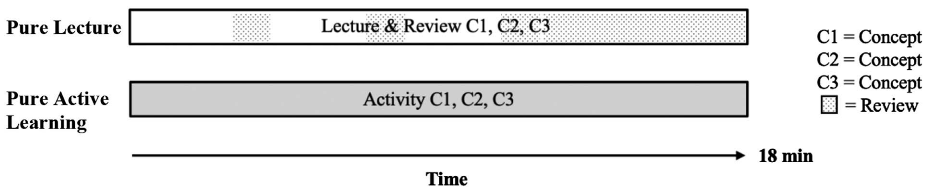
Note. See the online article for the color version of this figure.

Participants were randomly assigned to one of the two conditions and were initially blind as to what instructional mode would occur. The experiment took place online (using custom software hosted on a lab-based server for conducting experiments), and participants were able to participate at a time of their choosing. Participants completed an online consent form at the start of the experiment.

Each condition consisted of six phases completed within 1.5 hr, with relevant instructions provided when appropriate at the start of each phase to provide participants with information on what to expect during the forthcoming learning phase. In Phase 1, participants answered four demographic and four prior knowledge questions. In Phase 2, they watched an overview video for the content they would be learning about. In Phase 3, participants either watched a recorded lecture presentation or participated in an active learning activity, depending on their randomly assigned experimental condition. In Phase 4, they answered the final prior knowledge question, one JOL question, and one "instructional-mode enjoyment" question. They were also asked if they took notes during the lecture and/or activity. In Phase 5, they played Pacman for 5 min to serve as a short distractor task in-between the lesson and the posttest. In Phase 6, which was preceded by several reminders not to use notes or other aids, participants completed a 60-item multiple-choice posttest assessing knowledge of the content covered in the lecture or activity. Participants had 30 min to complete this test. Afterward, they were asked if they used any notes or other aids on the test. Finally, they viewed a debriefing form before exiting the experiment.

Phase 3 differed for the pure lecture and pure active learning conditions, but its duration was the same for both conditions (see Figure 4). In the pure lecture condition, participants watched a lecture presentation that lasted 18 min. Three "are you watching?" questions were asked—one after each taxonomic relationship was discussed during the lecture presentation—to determine if participants watched the entire lecture video. To be categorized as "engaged" for data analysis purposes, participants had to respond to two of the three questions within 15 s. In the pure active learning condition, participants participated in a matching activity that lasted 18 min. To complete the activity, participants had to correctly match each associative pair on the first try, four times in a row (Part 1) for each taxonomic relationship and correctly match all five associative pairs (as one unit) for each taxonomic relationship on the first try, four times in a row (Part 2). If they completed the activity before the 18-min period was over, they would reenter the activity and participate again. To be categorized as "engaged" for data analysis purposes, participants only had to complete Part 1 of the activity but did not have to complete the full activity (i.e., Part 2). Participants took an average of 8.40 min to complete Part 1 of the activity in a pilot study under an easier criterion of "correct on the first try, three

Figure 4
Phase 3 Differences Between Conditions in Experiment 1



times in a row," and therefore 18 min was deemed sufficient time for participants to complete at least Part 1 if they were actively engaging with it. Completing both Parts 1 and 2 would be more of a challenge in the 18 min.

Results and Discussion

Did Participants Differ in Their Prior Knowledge of the Content?

Although participants were randomly assigned to conditions to help reduce the chances of systematic differences between conditions at the start of the experiment, and although they would likely be novice learners due to the obscurity of the lesson content, their beliefs about their prior knowledge were analyzed to investigate these assumptions. Ratings for the four general prior knowledge questions were significantly correlated with one another (*r*s ranged from .323 to .756, all *p*s < .05), and had acceptable internal consistency (Cronbach's $\alpha = .78$). Therefore, data for the four questions were averaged for an overall general prior knowledge rating. The pure lecture and pure active learning conditions had average general prior knowledge ratings of 32.98% ($SD = 21.37$) and 30.58% ($SD = 20.87$), respectively, which were not significantly different, $t(144) = .69$, $p = .494$, 95% CI [−4.51, 9.31], $d = .11$.

With regard to specific prior knowledge, participants in the pure lecture and pure active learning conditions reported knowing an average of 1.04 (out of five possible; $SD = 1.65$) and .71 ($SD = 1.37$) organism categorizations, respectively, which were not significantly different, $t(144) = 1.31$, $p = .193$, 95% CI [−.17, .83], $d = .22$. Given that in both conditions, participants' ratings of their specific (and general) prior knowledge were similar and on the lower end of the scale, they could be categorized as novice learners. Therefore, any differences in posttest performance between conditions cannot be explained by differences in prior knowledge.

Did the Intensity of Active Learning Impact Participants' Learning of the Science Content?

To investigate our primary research question and determine whether pure lecture or pure active learning leads to better learning of science content, a 2 (condition: pure lecture, pure active learning) × 2 (question type: verbatim, inference) mixed factorial analysis of variance (ANOVA) was carried out with percentage of questions answered correctly on the posttest as the dependent measure.[2] We also report composite posttest scores that appropriately weight the accuracy scores for the verbatim and inference questions (i.e., 15/60 questions [25%] were verbatim questions, and 45/60 questions [75%] were inference questions). Finally, we calculated the internal consistency for both the verbatim and inference items for each condition separately by computing Cronbach's α. Values ranged from .80 to .96 (see Table 2).

The data are summarized in Figure 5. There was a main effect of question type, $F(1, 144) = 35.50$, $p < .001$, $\eta_p^2 = .20$, with participants performing higher on the verbatim questions (M = 54.95%, $SD = 29.37$) as compared to the inference questions (M = 48.70%, $SD = 27.47$). There was also a main effect of condition, $F(1, 144) = 24.02$, $p < .001$, $\eta_p^2 = .14$. This result held when a one-way ANOVA was conducted with the weighted composite score as the dependent measure, $F(1, 144) = 23.33$, $p < .001$, $\eta_p^2 = .14$, wherein participants scored higher on the posttest in the pure lecture condition (M = 60.47%, $SD = 29.74$) than in the pure active learning condition (M = 40.07%, $SD = 20.45$). These effects were not qualified by an interaction, $F(1, 144) = .88$, $p = .35$, $\eta_p^2 = .01$; as can be seen in Figure 5, participants in the pure lecture condition scored higher on both question types than participants in the pure active learning condition.

Did the Intensity of Active Learning Impact How Much Participants Felt They Learned or Enjoyed the Lesson?

JOL Ratings. For the judgments of learning after the lesson, participants in the pure lecture and pure active learning conditions had average ratings of 63.15% ($SD = 22.29$) and 60.68% (SD = 23.65), respectively, which were not significantly different, $t(144)$ = .65, $p = .518$, 95% CI [−5.05, 9.98], $d = .11$.

Instructional-Mode Enjoyment Ratings. For the assessment of instructional-mode enjoyment after the lesson, participants in the pure lecture and pure active learning conditions had average ratings of 49.04% ($SD = 28.49$) and 54.38% ($SD = 26.14$), respectively, which were not significantly different, $t(144) = 1.18$, $p = .240$, 95% CI [−3.60, 14.29], $d = .20$. Therefore, although lecture is often criticized for being boring and for putting "kids to sleep" (Strauss, 2017, p. 1), participants who viewed the lecture did not rate it less favorably than their peers who viewed the active learning activity.

How Much Practice Did Participants Receive in the Active Learning Condition?

As an additional exploratory analysis, participants' activity performance was examined to provide insight into the degree of practice participants received with the lesson content in the pure

[2] Due to a computer error, one participant in the lecture condition saw 54 test questions rather than the full set of 60 questions. Percent correct for this participant was therefore calculated based on the number of questions presented.

Table 2
Reliability Results for the Posttest Items

Experiment	Condition	Question type	Cronbach's α
1	Pure lecture	Verbatim	.901
		Inference	.963
	Pure active learning	Verbatim	.798
		Inference	.899
2	Pure lecture	Verbatim	.919
		Inference	.966
	Interspersed	Verbatim	.869
		Inference	.957
	Blocked	Verbatim	.890
		Inference	.954

active learning condition. During both parts of the activity, participants completed matching trials whereby they created an association between two taxonomic ranks for each organism (e.g., dragging "red kangaroo" into the bin for "*Macropus rufus*"). See Table 3 for the average number of matching trials participants went through in the activity and their average accuracy across these trials. The high activity accuracy (average of 91.0% for all three associative concepts combined) indicates that participants were engaged and successful in the activity, and not mindlessly dragging and dropping items during the activity time period. Moreover, there was a moderate correlation between the proportion of trials successfully completed and participant performance on the posttest, $t(71) = 3.69$, $p < .001$, 95% CI [.19, .58], $r = .40$. Although one cannot infer causation from it, this correlation may provide insight into the importance of successful repeated practice whereby the more successful practice participants had, the higher they scored on the posttest.

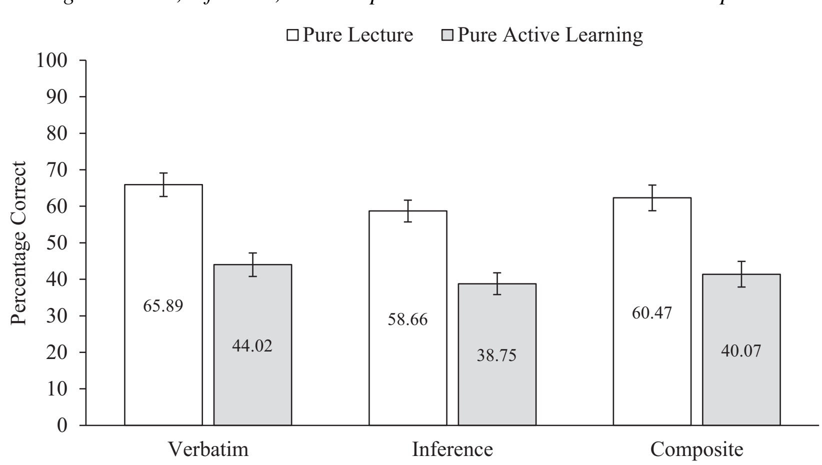
Experiment 2

The results of Experiment 1 indicated that pure lecture was more effective for participant learning of taxonomic relationships than pure active learning. This finding is important for research on active learning versus lecture because it reflects an unconfounded comparison (100% active learning vs. 100% lecture), showing how the pure forms of both approaches influenced participant performance. To some readers, our results might appear to be at odds with the results reported in the Freeman et al. (2014) meta-analysis, where active learning was found to be better than traditional lecture. However, previous studies—such as those included in that meta-analysis—typically compared lecture-only conditions to active learning conditions that often involved a lecture component. In contrast, our Experiment 1 compared a lecture-only condition to an active learning condition where lecture was absent. Thus, Experiment 1 filled an empirical gap by providing a clearer test than what is usually presented in the literature.

That said, given that the active learning conditions in research and in practice typically involve both activity and lecture, and there is meta-analytic evidence on the effectiveness of their combination (Freeman et al., 2014), Experiment 2 was designed to more systematically investigate the effectiveness of different ways of combining the two instructional approaches. Based on the main outcome of Experiment 1 (better posttest performance after 100% lecture than after 100% active learning), the goal of Experiment 2 was to determine whether a combination of lecture and active learning might result in better learning than lecture alone (which would be consistent with the findings from the meta-analysis by Freeman et al., 2014). Experiment 2 also investigated whether the way in which lecture and active learning were combined impacts participant learning.

Experiment 2 was designed to assess the effects of integrating lecture and active learning on participant performance by comparing the pure lecture condition from Experiment 1 to equal intensities of lecture and active learning (i.e., 50% lecture and 50% active learning) in interspersed and blocked schedules. Experiment 2 included three conditions. In the pure lecture condition, participants learned taxonomic relationships through a lesson involving a video lecture. In the interspersed condition, participants learned taxonomic relationships through a lesson that involved three minivideo lectures

Figure 5
Average Verbatim, Inference, and Composite Scores Across Conditions in Experiment 1



Note. Error bars represent standard errors of the means.

Table 3
Average Number of Activity Trials and Activity Accuracy Across Associative Concepts in Experiment 1

Associative concept	Average number of trials	Average accuracy (%)
Common name → species	55.49	97.6
Species → order	75.90	88.1
Common name → phylum	67.58	90.3
Overall (all three combined)	198.97	91.0

interspersed with three minimatching activities. In the blocked condition, participants learned taxonomic relationships through a lesson that involved a short video lecture followed by a short matching activity. Participants in the three conditions indicated how much they thought they had learned from their lesson and how much they enjoyed the instructional method they received. They then took a posttest after the lesson to assess their learning of the lesson content; the posttest contained questions that assessed directly and indirectly learned associative pairs. The materials used in Experiment 1 were identical to those used in Experiment 2. There was one primary research question and two exploratory research questions that were investigated in the present experiment:

1. Primary research question: Does the intensity and integration of active learning and lecture impact participants' learning of science content?
2. Exploratory Research Question 1: Does the intensity and integration of active learning and lecture impact how much participants feel they learned from the lesson?
3. Exploratory Research Question 2: Does the intensity and integration of active learning and lecture impact how much participants enjoyed the lesson?

Method

Participants and Design

Two hundred fifty-six introductory psychology undergraduate students from a large public Midwestern university participated for partial course credit (research ethics approval was obtained for our study through the institutional review board); however, 37 participants were excluded (10 in the pure lecture condition, eight in the interspersed condition, and 19 in the blocked condition) due to being under 18 years of age, failing the lecture-viewing check(s) during the experiment, not completing Part 1 of the activities, and/or using notes or other aids on the posttest. The final analytic sample included 219 participants; see Table 1, for demographic information. These participants were randomly assigned to one of the three conditions (pure lecture, interspersed lecture and active learning, or blocked lecture and active learning) as part of a between-participants design; 73 participants were in each condition.

Including 73 participants in each condition was based on an a priori power analysis conducted using G*Power 3.1.9.6 (Faul et al., 2007). By considering the effect size for the posttest performance difference in Experiment 1 and examining various scenarios for how the combined conditions (i.e., interspersed and blocked) might impact participant learning as compared to the pure lecture condition, we determined that a sample size of 73 participants per condition would achieve at least 80% power ($\alpha = .05$, two-tailed test) to detect a moderate effect size of .50.

Materials and Measures

The lesson content, overview video, prior knowledge questions, JOL question, and multiple-choice posttest were identical to what was used in Experiment 1. We describe differences in other materials and measures below.

Recorded Lecture Presentation. Participants received the same recorded lecture presentation as in Experiment 1. However, in the conditions that were reduced to 50% lecture, the 9-min review of each associative pair during the full 18-min lecture was removed. Therefore, the lecture presentation was a total length of 9 min and contained the same explanations of the three taxonomic relationships (and 15 associative pairs) as in the full lecture, but it did not contain any redundant information that occurred during the review in the full 18-min lecture.

Active Learning Activity. Participants in conditions that contained active learning received the same active learning activity (a matching activity) as in Experiment 1. However, to reduce active learning to 50% of the time during the relevant phase of the experiment, the following changes occurred. First, participants had half of the time (i.e., 9 min) to complete the activity. Second, participants had to correctly match each associative pair on the first try, two times (rather than four) in a row (Part 1), and correctly match all five associative pairs (as one unit) for each taxonomic relationship on the first try, two times (rather than four) in a row (Part 2). If they completed the activity before the 9-min period was over, they would reenter the activity and participate again.

Instructional-Mode Enjoyment Question. Participants in the pure lecture condition received the same instructional-mode enjoyment question that was asked in Experiment 1 (see Question 1 below). However, participants in conditions that included both a lecture and an active learning activity received the following three questions after the lesson:

1. On a scale of 0%–100%, how much did you enjoy the lecture?
2. On a scale of 0%–100%, how much did you enjoy the activity?
3. On a scale of 0%–100%, how much did you enjoy having both a lecture and an activity?

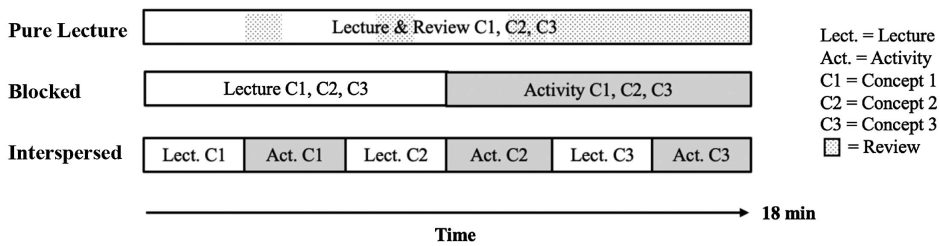
Procedure

We implemented a between-participants experimental design with three levels: pure lecture (100%), interspersed lecture (50%) and active learning (50%), and blocked lecture (50%) and active learning (50%). Participants were randomly assigned to one of the three conditions and were initially blind as to what instructional mode would occur. The experiment took place online, and participants were able to participate at a time of their choosing.

The timeline for each condition was identical to the timeline for the conditions in Experiment 1. However, Phase 3 (the lesson) differed across conditions (see Figure 6). The pure lecture condition

Figure 6

Phase 3 Differences Across Conditions and Timelines in Experiment 2



was identical to that of Experiment 1. In the interspersed condition, participants alternated between watching a lecture video and participating in the active learning activity. The lecture video lasted 9 min and was divided into three segments, with each segment (~3 min) covering a different taxonomic relationship. One "are you watching?" question was asked within 45 s of each video segment ending to ensure participants viewed the entire lecture. To be categorized as "engaged" for data analysis purposes, participants had to respond to this question within 15 s for two of the three segments. After each lecture segment, participants participated in the active learning activity over the taxonomic relationship covered in the preceding segment. Therefore, the active learning activity was also divided into three segments (each 3 min). To complete the activity, participants had to correctly match each associative pair on the first try, two times in a row (Part 1), and correctly match all five associative pairs (as one unit) for the taxonomic relationship on the first try, two times in a row (Part 2). If they completed both parts of the activity before the 3-min period was over, they would reenter the activity and participate again. To be categorized as "engaged" for data analysis purposes, participants had to complete Part 1 of each activity segment at least once.

In the blocked condition, participants watched the entire lecture video before participating in the active learning activity. The lecture video lasted 9 min, and three "are you watching?" questions were asked at the end of each taxonomic relationship to determine if participants watched the entire lecture. To be categorized as "engaged" for data analysis purposes, participants had to respond to two of the three questions within 15 s. After watching the lecture presentation, they received instructions about the active learning activity and then participated in the activity for 9 min. To complete the activity, participants had to correctly match each associative pair on the first try, two times in a row (Part 1), and correctly match all five associative pairs (as one unit) for each taxonomic relationship on the first try, two times in a row (Part 2). If they completed the activity before the 9-min period was over, they would reenter the activity and participate again. To be categorized as "engaged" for data analysis purposes, participants had to complete Part 1 of the activity at least once.

Results and Discussion

Did Participants Differ in Their Prior Knowledge of the Content?

As in Experiment 1, ratings for the four general prior knowledge questions were significantly correlated with one another (rs ranged from .363 to .793, all ps $<$ .05), and had acceptable internal consistency (Cronbach's α = .81). Therefore, data for the four questions were averaged for an overall general prior knowledge rating. Participants in the pure lecture, interspersed, and blocked conditions had average ratings of 31.68% (SD = 20.13), 30.82% (SD = 22.06), and 27.47% (SD = 19.54), respectively, which were not significantly different, $F(2, 216) = .85$ $p = .428$, $\eta_p^2 = .01$.

With regard to specific prior knowledge, participants in the pure lecture, interspersed, and blocked conditions reported knowing an average of .77 (SD = 1.42), .60 (SD = 1.27), and .88 (SD = 1.69) organism categorizations, respectively, which were not significantly different, $F(2,216) = .64$, $p = .527$, $\eta_p^2 = .01$. Given that in all conditions, participants' ratings of their specific (and general) prior knowledge were similar and on the lower end of the scale, they could be categorized as novice learners. Therefore, any differences in posttest performance between conditions cannot be explained by differences in prior knowledge.

Did the Intensity and Integration of Active Learning Impact Participants' Learning of the Science Content?

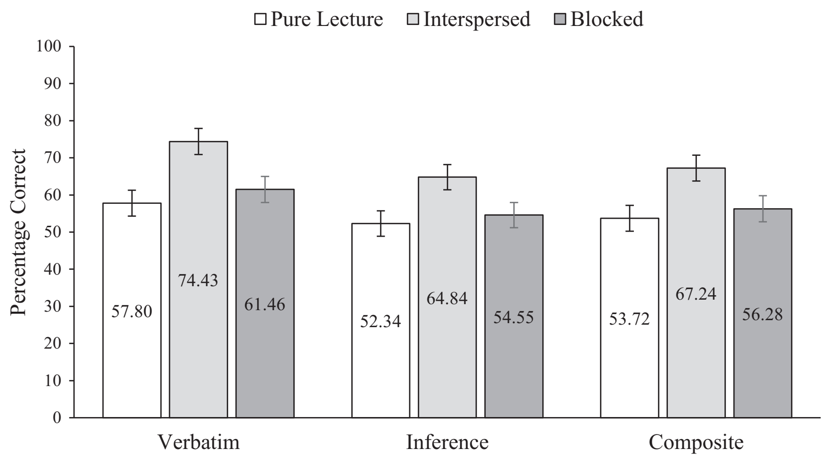
To investigate our primary research question and determine which intensity and integration schedule leads to better learning of science content, a 3 (condition: pure lecture, interspersed, blocked) × 2 (question type: verbatim, inference) mixed factorial ANOVA was carried out with percentage of questions answered correctly on the posttest as the dependent measure.[3] As before, we also report appropriately weighted composite posttest scores. Finally, we calculated the internal consistency for both the verbatim and inference items for each condition separately by computing Cronbach's α. Values ranged from .87 to .97 (see Table 2).

The data are summarized in Figure 7. There was a main effect of question type, $F(1, 216) = 100.45$, $p < .001$, $\eta_p^2 = .32$, with participants performing higher on the verbatim questions (M = 64.56%, SD = 30.26) as compared to the inference questions (M = 57.25%, SD = 29.35). There was also a main effect of condition, $F(2, 216) = 5.24$, $p = .006$, $\eta_p^2 = .05$. This result held when a one-way ANOVA was conducted with the weighted composite score as the dependent measure, $F(2, 216) = 4.56$, $p = .012$, $\eta_p^2 = .04$, wherein participants in the interspersed condition scored higher than participants in both the blocked condition (M_{diff} = 10.96%, SE = 4.76, $p = .022$, 95% CI [1.58, 20.33], $d = .40$) and the pure lecture

[3] Due to a computer error, three participants in the lecture condition did not see the full set of 60 questions (they saw 41, 43, and 59 questions, respectively). Percent correct for these participants was therefore calculated based on the number of questions presented.

Figure 7

Average Verbatim, Inference, and Composite Scores Across Conditions in Experiment 2



Note. Error bars represent standard errors of the means.

condition (M_{diff} = 13.52%, *SE* = 4.76, *p* = .005, 95% CI [4.14, 22.89], *d* = .47). The blocked condition did not differ significantly from the pure lecture condition (M_{diff} = 2.56%, *SE* = 4.76, *p* = .591, 95% CI [−6.82, 11.93], *d* = .09). These effects were not qualified by an interaction, *F*(2, 216) = 2.75, *p* = .066, η_p^2 = .03; as can be seen in Figure 7, participants in the interspersed condition generally scored higher on both question types than participants in the blocked or pure lecture conditions.

Did the Intensity of Active Learning Impact How Much Participants Felt They Learned or Enjoyed the Lesson?

JOL Ratings. For the judgments of learning after the lesson, there was a significant effect of condition, *F*(2, 216) = 8.67, $p <$ .001, η_p^2 = .07. Participants in the pure lecture, interspersed, and blocked conditions had average ratings of 62.33% (*SD* = 24.58), 76.85% (*SD* = 18.63), and 71.23% (*SD* = 20.07), respectively. The pure lecture condition's ratings were significantly lower than those in both the interspersed condition (M_{diff} = 14.52%, *SE* = 3.52, $p <$ .001, 95% CI [7.59, 21.45], *d* = .67) and the blocked condition (M_{diff} = 8.90%, *SE* = 3.52, *p* = .012, 95% CI [1.97, 15.83], *d* = .40). However, the interspersed condition and the blocked condition did not significantly differ in their ratings (M_{diff} = 5.62%, *SE* = 3.52, *p* = .112, 95% CI [−1.31, 12.55], *d* = .29).

Instructional-Mode Enjoyment Ratings. For the assessment of instructional-mode enjoyment after the lesson, three separate analyses were conducted, one for each of the three questions participants were asked. The first analysis compared lecture enjoyment ratings across the three conditions. There was a significant effect of condition, *F*(2, 216) = 3.48, *p* = .033, η_p^2 = .03. Participants in the pure lecture, interspersed, and blocked conditions had average ratings for the lecture of 50.96% (*SD* = 28.00), 60.96% (*SD* = 22.86), and 60.82% (*SD* = 27.68), respectively. The pure lecture condition's ratings were significantly lower than those in both the interspersed condition (M_{diff} = 10.00%, *SE* = 4.35, *p* = .022, 95% CI [1.43, 18.57], *d* = .39) and the blocked condition (M_{diff} = 9.86%, *SE* = 4.35, *p* = .024, 95% CI [1.29, 18.44], *d* = .35). However, the interspersed and blocked conditions did not significantly differ in their ratings (M_{diff} = 0.14%, *SE* = 4.35, *p* = .975, 95% CI [−8.44, 8.71], *d* = .01). Therefore, the reduced lecture in the combined conditions was deemed, on average, more enjoyable than the full lecture in the pure lecture condition. Interestingly, having the lecture broken into 3-min segments (in the interspersed condition) as compared to an uninterrupted 9-min lecture (in the blocked condition) did not affect how much participants enjoyed the lecture.

The second analysis compared activity enjoyment ratings between the two combined conditions. The interspersed and blocked conditions had average ratings of 67.26% (*SD* = 21.56) and 72.33% (*SD* = 23.84), respectively; the difference was not significant, *t*(144) = 1.35, *p* = .180, 95% CI [−2.37, 12.50], *d* = .22. Therefore, having the activity interspersed or blocked did not affect how much participants enjoyed the activity.

The third analysis compared activity + lecture enjoyment ratings between the two combined conditions. The interspersed and blocked conditions had average ratings of 76.58% (*SD* = 19.81) and 72.88% (*SD* = 25.03), respectively; the difference was not significant, *t*(144) = .99, *p* = .324, 95% CI [−3.69, 11.08], *d* = .16. Therefore, the scheduling of lecture and active learning did not affect how much participants enjoyed having both the activity and the lecture.

How Much Practice Did Participants Receive in the Active Learning Condition?

As an additional exploratory analysis, participants' activity performance was examined to provide insight into the degree of practice participants received with the lesson content and how this practice differed between the blocked and interspersed conditions. During both parts of the activity, participants in the interspersed and blocked conditions completed matching trials whereby they created an association between two taxonomic ranks for each organism (e.g., dragging "red kangaroo" into the bin for "*Macropus rufus*"). See Table 4 for the average number of matching trials participants went through in the activity and their average accuracy

Table 4
Average Number of Activity Trials and Activity Accuracy Across Associative Concepts in Experiment 2

Associative concept	Condition	Average number of trials	Average accuracy
Common name → species	Blocked	28.22	98.4
Common name → species	Interspersed	52.89	98.9
Species → order	Blocked	39.92	86.9
Species → order	Interspersed	40.32	94.6
Common name → phylum	Blocked	33.89	89.4
Common name → phylum	Interspersed	53.92	97.8
Overall (all three combined)	Blocked	102.03	89.6
Overall (all three combined)	Interspersed	147.12	97.4

across these trials. Participants in the interspersed condition completed more matching trials, $t(141.43) = 8.07$, $p < .001$, 95% CI [34.05, 56.14], $d = 1.34$, and had higher accuracy, $t(87.95) = 6.69$, $p < .001$, 95% CI [.05, .10], $d = 1.10$, than participants in the blocked condition. However, the high activity accuracy in both conditions (averages of 97.4% and 89.6% for all three concepts combined in the interspersed and blocked conditions, respectively) indicates that participants were engaged and successful in the activity regardless of how it was scheduled. Moreover, there was a strong correlation between the proportion of trials successfully completed and participant performance on the posttest in both the blocked condition, $t(71) = 7.24$, $p < .001$, 95% CI [.50, .77], $r = .65$, and in the interspersed condition, $t(71) = 5.24$, $p < .001$, 95% CI [.34, .68], $r = .53$. As in Experiment 1, these correlations highlight the relationship between successful practice and posttest performance.

General Discussion

The purpose of the present experiments was to investigate how the intensity and integration of lecture and active learning affected participant learning of science content. In both experiments, participants learned about taxonomic relationships—how to categorize five organisms according to their species, order, and phylum levels—and their learning was assessed on a posttest. In Experiment 1, pure lecture was compared to pure active learning to provide an uncontaminated comparison of lecture and active learning interventions. In Experiment 2, pure lecture was compared to an integration of lecture and active learning scheduled for equal time periods in either a blocked or an interspersed manner.

Comparing Lecture and Active Learning: Experiment 1

In Experiment 1, the lecture condition outperformed the active learning condition by 20.40 percentage points on the posttest. Interestingly, participants in the lecture condition not only learned the directly taught content at a higher level (assessed through the verbatim test questions) but also demonstrated generative learning at higher levels (assessed through the more difficult inference test questions). In other words, there was greater knowledge integration in the lecture condition because participants were able to use the directly learned associative pairs to make inferences about other, albeit related, taxonomic relationships. Our results might appear to be at odds with the results reported in the Freeman et al. (2014) meta-analysis. However, our Experiment 1 compared a lecture-only condition to an active learning condition where lecture was absent, providing a cleaner comparison between active learning and lecture. Our conditions were also designed to provide a fair comparison. More specifically, although lecture is behaviorally passive and active learning is behaviorally active, this does not imply that lecture is always cognitively passive and that active learning is always cognitively active (Mayer, 2004; Opdal, 2022). The pure lecture condition in Experiment 1 was designed to encourage cognitive engagement by connecting new information to prior knowledge (via Greek/Latin roots) and by asking participants to think about what they had learned during the matching review phase of the lecture.

As for the active learning condition in Experiment 1, we ensured participants received accuracy feedback and allowed them to try again until they correctly matched the terms four times in a row. Therefore, participants entered the testing phase having achieved success (i.e., average accuracy in the activity was 91% in Experiment 1) in the learning phase. The active learning activity was also designed to encourage cognitive engagement by requiring participants to correctly match each associative pair four times in a row in the 18-min period. Participants had to focus on the task at hand in order to successfully complete the first part of the matching activity in the time allotted and were given the same Greek/Latin roots during the activity to help them connect the new information to prior knowledge. Importantly, the Experiment 2 results suggest that the superiority of the pure lecture condition over the pure active learning condition in Experiment 1 was not because the matching activity in the latter condition was ineffective or low quality. When the matching activity was interspersed with lecture in Experiment 2, it resulted in superior learning than lecture alone, indicating that the activity was indeed helpful, but only when it was integrated with lecture.

In addition to examining posttest performance, we investigated participant JOLs and enjoyment in exploratory analyses. The conditions in Experiment 1 did not significantly differ in how much participants believed they learned nor in how much they enjoyed the instructional mode they received, despite having learning differences. These results counter a recent study by Deslauriers et al. (2019) that found participants who received active learning reported that they learned less than those in the lecture condition when they actually learned more. However, there were implementation differences between studies, with the results of Experiment 1 providing insight into how pure active learning is viewed by participants as compared to pure lecture.

Importance of Examining How to Integrate Lecture and Active Learning: Experiment 2

In Experiment 2, the interspersed condition outperformed the blocked and pure lecture conditions on the posttest by 10.96 and 13.52 percentage points, respectively. This finding highlights that the intensity of active learning and lecture cannot be considered in isolation from how the two are integrated. More specifically, one cannot simply conclude that 100% lecture is better or worse than 50% lecture/50% active learning because it depends on the integration schedule of the latter. Interspersing lecture and active learning resulted in better posttest performance than blocking them did. In fact, blocking lecture and active learning led to roughly equivalent performance as pure lecture. As demonstrated in studies examining the effects of interpolated memory tests during lecture, students may be better able to extract lecture content through a reduction of mind wandering (Szpunar et al., 2013) and an increase in lecture-related thoughts (Jing et al., 2016). Further, as found by Healy et al. (2017), interspersing quizzes during learning may improve test performance by enhancing learner engagement. As such, in the interspersed condition in Experiment 2, the short lecture segments may have encouraged greater learning of the lecture content due to less mind wandering, more lecture-related thoughts, and enhanced engagement.

Further, the interspersed condition may have reduced proactive interference relative to the blocked and pure lecture conditions. In the interspersed condition, each taxonomic relationship (and its five associative pairs) was learned individually and then practiced individually, whereas in the other conditions, all three taxonomic relationships (and their five associative pairs) were learned via lecture and then practiced or reviewed as a group. Szpunar et al. (2008) found that interpolating tests helped protect against proactive interference in that testing participants on previously learned content helped to reduce the negative effects of that content on learning later content. Further, the higher degree of practice and accuracy that arose from interspersing the instructional methods may have diminished any proactive interference effects, consistent with work by Underwood (1949) showing that learning a previous list (A → B) at a high level could facilitate the learning of a similar second list (A → C).

In addition to examining posttest performance, we investigated participant JOLs and enjoyment in exploratory analyses. The pure lecture condition had a significantly lower average JOL than the interspersed and blocked conditions; however, the interspersed and blocked conditions did not significantly differ on this measure. The JOL differences across conditions counter the results of Deslauriers et al. (2019) for actual learning versus feeling of learning in passive and active classrooms. The differing results between the present study and the study by Deslauriers et al., which might be partly attributable to implementation differences, suggest that it is premature to conclude that active learning leads to higher levels of actual learning but lower levels of feelings of learning.

When examining how much participants enjoyed the lecture, there was an advantage for the interspersed and blocked conditions over the pure lecture condition (by 10% and 9.86%, respectively). Therefore, by reducing the lecture, participants viewed the presentation more favorably, perhaps because they had increased confidence in their learning and enjoyment through the activity. Given that the interspersed and blocked conditions did not significantly differ in their lecture enjoyment, it appears that the scheduling of the lecture did not significantly influence participants' enjoyment of it. The interspersed and blocked conditions also did not significantly differ in terms of enjoying the activity and enjoying having both a lecture and an activity, perhaps given that participants in both conditions experienced success during the activity, which allowed them the opportunity to assess and improve their knowledge.

Limitations and Future Directions

We will discuss five primary limitations of the present study. First, it is difficult to generalize the results of the two experiments to other implementations of lecture and active learning interventions. There are numerous ways to structure and implement a lecture, and there are numerous ways to design and implement active learning activities. Future research could explore whether our results hold when the intensity and integration of the approaches are manipulated using different lectures and activities across different topics and course disciplines. Second, given that the experiments occurred online, participants could have used notes or other aids on the posttest, although this would have impacted all conditions. However, this issue was mitigated by (a) instructions telling participants not to use notes or other aids, (b) instructions telling participants they would not be adversely affected in any way if they admitted to using notes, (c) the exclusion of data from participants who indicated that they did not comply with those instructions, and (d) the observation that the excluded participants did not score unusually high relative to peers who had not taken notes. Third, each experiment was a single session lasting a maximum of 1.5 hr, and the lesson was conducted over three taxonomic relationships (each with five associative pairs). It would be important to examine if the results hold for longer interventions (e.g., 1 week or 1 semester) that cover more concepts. Fourth, the posttest in both experiments occurred 5 min after the lesson ended. Therefore, the present study examined learning as compared to retention. The results of the present study cannot speak to whether the conditions that led to the highest performance on the 5-min delayed posttest would also lead to greater retention as assessed on a 1-week delayed posttest, for example. Therefore, future research should examine if the results hold for the retention of information after longer delays. Fifth, the engagement criterion for the different conditions was set such that participants needed to engage with the content for at least 12 min. However, it is difficult to determine in the present study precisely how long participants engaged with the materials beyond meeting the engagement criterion.

Building off the present study, future experiments could be conducted to investigate how different types of activities and the match between the activity and the test impact the effectiveness of an active learning condition. Further, future research experiments could be designed to assess whether interspersing lecture and active learning is more effective when lecture is presented first or second (i.e., order of approaches). As found by Ashman et al. (2020), explicit instruction followed by problem solving was more effective than when problem solving was followed by explicit instruction. Similarly, based on the cognitive load hypothesis (e.g., Sweller et al., 2019), lecture followed by learning activities is particularly effective for students with low prior knowledge. However, researchers who subscribe to the productive failure hypothesis (e.g., Kapur, 2008) would argue that learning activities that are presented before more explicit forms of instruction are more effective for learning. Considering these different views, the order in which activities should be presented is an important variable to study when it comes to optimizing the integration of activities into a class lecture.

Regarding intensity, one could investigate how much lecture can be reduced before it no longer enhances learning beyond the pure active learning condition (e.g., by including a condition with 25% lecture and 75% active learning). This variable has not yet been systematically studied in the active learning literature. Martella and Schneider (in press) provide a detailed discussion of future research directions related to the ordering, sequencing, and intensity of lecture and activities. The materials from the present study allow for such systematic manipulation of both intensity and integration schedule. However, it is important to continue to study these variables in a variety of contexts to determine if the results shift based on course topic, lesson difficulty, students' prior knowledge level, and other factors.

References

Ashman, G., Kalyuga, S., & Sweller, J. (2020). Problem-solving or explicit instruction: Which should go first when element interactivity is high? *Educational Psychology Review*, *32*(1), 229–247. https://doi.org/10.1007/s10648-019-09500-5

Association of American Universities. (2017). *Progress toward achieving systemic change: A five-year status report on the AAU undergraduate STEM education initiative*. https://www.aau.edu/sites/default/files/AAU-Files/STEM-Education-Initiative/STEM-Status-Report.pdf

Bajak, A. (2014, May 12). Lectures aren't just boring, they're ineffective, too, study finds. *Science*. https://www.science.org/content/article/lectures-arent-just-boring-theyre-ineffective-too-study-finds

Bjork, R. A., & Linn, M. C. (2006). The science of learning and the learning of science introducing desirable difficulties. *APS Observer*, *19*(3), 1–2. https://www.psychologicalscience.org/observer/the-science-of-learning-and-the-learning-of-science

Butler, A. C., & Roediger, H. L., III. (2008). Feedback enhances the positive effects and reduces the negative effects of multiple-choice testing. *Memory & Cognition*, *36*(3), 604–616. https://doi.org/10.3758/MC.36.3.604

Carpenter, S. K. (2017). Spacing effects on learning and memory. In J. T. Wixted (Ed.), *Cognitive psychology of memory, Vol. 2 of Learning and memory: A comprehensive reference* (J. H. Byrne, Series Ed.; pp. 465–485). Academic Press. https://doi.org/10.1016/B978-0-12-809324-5.21054-7

Center for STEM Learning. (2016). *TRESTLE mini seed grant proposals: Transforming education, supporting teaching and learning excellence*. University of Colorado Boulder. https://www.colorado.edu/csl/sites/default/files/attached-files/trestle_rfp-_minigrant_0.pdf

Chi, M. T. H. (2009). Active-constructive-interactive: A conceptual framework for differentiating learning activities. *Topics in Cognitive Science*, *1*(1), 73–105. https://doi.org/10.1111/j.1756-8765.2008.01005.x

Chi, M. T. H., Adams, J., Bogusch, E. B., Bruchok, C., Kang, S., Lancaster, M., Levy, R., Li, N., McEldoon, K. L., Stump, G. S., Wylie, R., Xu, D., & Yaghmourian, D. L. (2018). Translating the ICAP theory of cognitive engagement into practice. *Cognitive Science*, *42*(6), 1777–1832. https://doi.org/10.1111/cogs.12626

Chi, M. T. H., & Wylie, R. (2014). The ICAP framework: Linking cognitive engagement to active learning outcomes. *Educational Psychologist*, *49*(4), 219–243. https://doi.org/10.1080/00461520.2014.965823

Dawson, P. (2015, July 2). Will the University of Adelaide's lecture phase-out be a flop? *The Conversation*. https://theconversation.com/will-the-university-of-adelaides-lecture-phase-out-be-a-flop-44074

Deslauriers, L., McCarty, L. S., Miller, K., Callaghan, K., & Kestin, G. (2019). Measuring actual learning versus feeling of learning in response to being actively engaged in the classroom. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(39), 19251–19257. https://doi.org/10.1073/pnas.1821936116

Driessen, E. P., Knight, J. K., Smith, M. K., & Ballen, C. J. (2020). Demystifying the meaning of active learning in postsecondary biology education. *CBE Life Sciences Education*, *19*(4), Article ar52. https://doi.org/10.1187/cbe.20-04-0068

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191. https://doi.org/10.3758/BF03193146

Fiorella, L., & Mayer, R. E. (2015). *Learning as a generative activity: Eight learning strategies that promote understanding*. Cambridge University Press. https://doi.org/10.1017/CBO9781107707085

Fiorella, L., & Mayer, R. E. (2016). Eight ways to promote generative learning. *Educational Psychology Review*, *28*(4), 717–741. https://doi.org/10.1007/s10648-015-9348-9

Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., & Wenderoth, M. P. (2014). Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(23), 8410–8415. https://doi.org/10.1073/pnas.1319030111

Healy, A. F., Jones, M., Lalchandani, L. A., & Tack, L. A. (2017). Timing of quizzes during learning: Effects on motivation and retention. *Journal of Experimental Psychology*, *23*(2), 128–137. https://doi.org/10.1037/xap0000123

Jing, H. G., Szpunar, K. K., & Schacter, D. L. (2016). Interpolated testing influences focused attention and improves integration of information during a video-recorded lecture. *Journal of Experimental Psychology: Applied*, *22*(3), 305–318. https://doi.org/10.1037/xap0000087

Kapur, M. (2008). Productive failure. *Cognition and Instruction*, *26*(3), 379–424. https://doi.org/10.1080/07370000802212669

Karpicke, J. D., & Roediger, H. L., III. (2007). Repeated retrieval during learning is the key to long-term retention. *Journal of Memory and Language*, *57*(2), 151–162. https://doi.org/10.1016/j.jml.2006.09.004

Karpicke, J. D., & Roediger, H. L., III. (2010). Is expanding retrieval a superior method for learning text materials? *Memory & Cognition*, *38*(1), 116–124. https://doi.org/10.3758/MC.38.1.116

Lhyle, K. G., & Kulhavy, R. W. (1987). Feedback processing and error correction. *Journal of Educational Psychology*, *79*(3), 320–322. https://doi.org/10.1037/0022-0663.79.3.320

Lombardi, D., Shipley, T. F., & the Astronomy Team, Biology Team, Chemistry Team, Engineering Team, Geography Team, Geoscience Team, and Physics Team. (2021). The curious construct of active learning. *Psychological Science in the Public Interest*, *22*(1), 8–43. https://doi.org/10.1177/1529100620973974

Martella, A. M., Lovett, M. C., & Ramsay, L. (2021). Implementing active learning: A critical examination of sources of variation in active learning science courses. *Journal on Excellence in College*, *32*(1), 67–96.

Martella, A. M., & Schneider, D. W. (in press). A reflection on the current state of active learning research. *The Journal of Scholarship of Teaching and Learning*.

Massachusetts Institute of Technology. (2021). *Technology-Enhanced Active Learning*. https://web.mit.edu/edtech/casestudies/teal.html

Mayer, R. E. (2002). Rote versus meaningful learning. *Theory Into Practice*, *41*(4), 226–232. https://doi.org/10.1207/s15430421tip4104_4

Mayer, R. E. (2004). Should there be a three-strikes rule against pure discovery learning? The case for guided methods of instruction. *American Psychologist*, *59*(1), 14–19. https://doi.org/10.1037/0003-066X.59.1.14

Mayer, R. E. (2011). *Applying the science of learning*. Pearson/Merrill/Prentice Hall.

Mayer, R. E. (2021). *Multimedia learning* (3rd ed.). Cambridge University Press.

McCarroll, M. L., Pohle-Krauza, R. J., & Martin, J. L. (2009). Active learning in the classroom: A muscle identification game in a kinesiology course. *Advances in Physiology Education*, *33*(4), 319–322. https://doi.org/10.1152/advan.00013.2009

Nuetzman, A. L., & Abdullaev, Y. (2012). Teaching medical terminology using word-matching games. *Journal of Continuing Education in Nursing*, *43*(7), 297–298. https://doi.org/10.3928/00220124-20120621-04

Opdal, P. A. (2022). To do or to listen? Student active learning vs. the lecture. *Studies in Philosophy and Education*, *41*, 71–89. https://doi.org/10.1007/s11217-021-09796-3

Stains, M., Harshman, J., Barker, M. K., Chasteen, S. V., Cole, R., DeChenne-Peters, S. E., Eagan, M. K., Jr., Esson, J. M., Knight, J. K., Laski, F. A., Levis-Fitzgerald, M., Lee, C. J., Lo, S. M., McDonnell, L. M., McKay, T. A., Michelotti, N., Musgrove, A., Palmer, M. S., Plank, K. M., … Young, A. M. (2018). Anatomy of STEM teaching in North American universities. *Science*, *359*(6383), 1468–1470. https://doi.org/10.1126/science.aap8892

Strauss, V. (2017, July). It puts kids to sleep—But teachers keep lecturing anyway. Here's what to do about it. *The Washington Post*. https://www.washingtonpost.com/news/answer-sheet/wp/2017/07/11/it-puts-kids-to-sleep-but-teachers-keep-lecturing-anyway-heres-what-to-do-about-it/

Surprenant, A. M., & Neath, I. (2009). *Principles of memory*. Psychology Press.

Sweller, J. (2004). Instructional design consequences of an analogy between evolution by natural selection and human cognitive architecture. *Instructional Science*, *32*(1–2), 9–31. https://doi.org/10.1023/B:TRUC.0000021808.72598.4d

Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive load theory*. Springer. https://doi.org/10.1007/978-1-4419-8126-4

Sweller, J., van Merriënboer, J. J. G., & Paas, F. (1998). Cognitive architecture and instructional design. *Educational Psychology Review*, *10*(3), 251–296. https://doi.org/10.1023/A:1022193728205

Sweller, J., van Merriënboer, J. J. G., & Paas, F. (2019). Cognitive architecture and instructional design: 20 years later. *Educational Psychology Review*, *31*(2), 261–292. https://doi.org/10.1007/s10648-019-09465-5

Szpunar, K. K., Khan, N. Y., & Schacter, D. L. (2013). Interpolated memory tests reduce mind wandering and improve learning of online lectures. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(16), 6313–6317. https://doi.org/10.1073/pnas.1221764110

Szpunar, K. K., McDermott, K. B., & Roediger, H. L., III. (2008). Testing during study insulates against the buildup of proactive interference. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(6), 1392–1399. https://doi.org/10.1037/a0013082

Underwood, B. J. (1949). Proactive inhibition as a function of time and degree of prior learning. *Journal of Experimental Psychology*, *39*(1), 24–34. https://doi.org/10.1037/h0059550

Uner, O., & Roediger, H. L. (2018). The effect of question placement on learning from textbook chapters. *Journal of Applied Research in Memory and Cognition*, *7*(1), 116–122. https://doi.org/10.1016/j.jarmac.2017.09.002

University of Georgia Office of Instruction. (2021). *Active learning*. https://ovpi.uga.edu/initiatives/active-learning/

Watkins, O. C., & Watkins, M. J. (1975). Buildup of proactive inhibition as a cue-overload effect. *Journal of Experimental Psychology: Human Learning and Memory*, *104*(4), 442–452. https://doi.org/10.1037/0278-7393.1.4.442

Weinstein, Y., Nunes, L. D., & Karpicke, J. D. (2016). On the placement of practice questions during study. *Journal of Experimental Psychology: Applied*, *22*(1), 72–84. https://doi.org/10.1037/xap0000071

Wieman, C. E. (2014). Large-scale comparison of science teaching methods sends clear message. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(23), 8319–8320. https://doi.org/10.1073/pnas.1407304111

Zakrajsek, T. (2018). Reframing the lecture versus active learning debate: Suggestions for a new way forward. *Education in the Health Professions*, *1*(1), 1–3. https://doi.org/10.4103/EHP.EHP_14_18

Received August 17, 2023
Revision received November 15, 2023
Accepted December 12, 2023 ■