Background Topic Knowledge as Possible Moderator of the Testing Effect: an Experimental Investigation

by

Jessica Allison Macaluso

B.S., University of California, Davis, 2020

Submitted to the Graduate Faculty of the

Dietrich School of Arts and Sciences in partial fulfillment

of the requirements for the degree of

Master of Science

University of Pittsburgh

2023

UNIVERSITY OF PITTSBURGH

DIETRICH SCHOOL OF ARTS AND SCIENCES

This thesis was presented

by

Jessica Allison Macaluso

It was defended on

August 23, 2023

and approved by

Benjamin M. Rottman, Associate Professor, Psychology

Timothy J. Nokes-Malach, Professor, Psychology

Thesis Advisor/Dissertation Director: Scott H. Fraundorf, Associate Professor, Psychology

Background Topic Knowledge as a Possible Moderator of the Testing Effect: an Experimental Investigation

Jessica Allison Macaluso, B.S., M.S.

University of Pittsburgh, 2023

A large body of research has established a testing effect: testing of information results in better long-term retention than restudying. Given the relevance of such effects for education, there is interest in the conditions and learner differences that may moderate the utility of testing, like pre-existing knowledge. It is possible that *(a)* the testing effect is stronger for those who are more novice because testing is so effective that it helps novice learners catch up to expert learners or *(b)* the testing effect works equally well for everyone because it is so robust that all learners can benefit from it. In two experiments, college students read texts and were tested on them one week later. I orthogonally manipulated study strategy (retrieval practice versus restudying) and availability of background material for a given topic. In Experiment 2 only, feedback availability was orthogonally manipulated during testing. I found that participants had significantly better retention when they studied via retrieval practice and received feedback (Experiment 2). This demonstrates that immediate feedback is beneficial for learners, at least over a one-week retention interval. Further, when collapsing across categories of feedback, I found that participants had better retention when they received background topic material, regardless of which study strategy was used (Experiment 2) or performed significantly better when studying via restudying (Experiment 1). These results suggest that the learning benefits of testing do not depend on having high levels of existing domain knowledge.

Table of Contents

List of Tables

List of Figures

1.0 Introduction

In recent decades, a large body of research has been conducted concerning the testing effect. The *testing effect*, also known as the *retrieval-practice effect*, refers to the fact that intermittent testing of information (e.g., quizzing oneself on previously studied material) enhances learning and long-term retention of that material (Bjork & Bjork, 1992; Roediger & Karpicke, 2006; Rowland, 2014). Typically, experiments regarding the testing effect compare two study strategies: restudying and retrieval practice (Carpenter, Pashler, Wixted, & Vul, 2008; Karpicke & Roediger, 2008). *Restudying* is when one studies the provided material and then studies that same material again, such as rereading the same passage. *Retrieval practice* is when one studies the necessary material and then is required to remember the information in some way (e.g., writing down as much as one can remember of the studied material). Prior research has revealed that deeper processing of information, such as via retrieval practice, results in superior retention of that material, when compared to restudying (Adesope, Trevisan, & Sundararajan, 2017; Rowland, 2014; Yang, Luo, Vadillo, Yu, & Shanks, 2021).

In the present study, I examined whether the testing effect is moderated by learners' existing knowledge about a topic, which I experimentally manipulated in each of two studies. I will first review some basic research on the testing effect before introducing the need to identify key moderators of this effect, and why background topic knowledge may be one of them. By understanding possible moderators of the testing effect, we can better understand what study strategies are best for individuals and in what conditions certain study strategies are most helpful.

1.1 The Testing Effect

As of 2023, there are at least three published meta-analyses focusing on the testing effect (Adesope et al., 2017; Rowland, 2014; Yang et al., 2021). The results of these meta-analyses establish that the testing effect is a convincing phenomenon with medium-to-large effect sizes according to the criteria established by Gignac & Szodorai (2016b), in both lab studies (g = 0.51, Adesope et al., 2017; g = 0.50 Rowland, 2014) and in authentic educational settings (g = 0.33, Yang et al., 2021).

1.1.1 Cognitive Mechanisms Underlying the Testing Effect

Practice testing may benefit learning through both direct and indirect pathways, as I detail below. The present studies are not intended to discriminate between these mechanisms — it is possible that either or both are occurring — but to instead explore background topic knowledge as a possible moderator. It is feasible that direct testing could be strengthened by additional background topic knowledge. It is also possible that via an indirect pathway, background topic knowledge aids a learner in their understanding of the presented material, boosting the effectiveness of testing.

1.1.2 The Direct Testing Effect

The *direct testing effect* occurs when practice tests directly improve future retrieval of information from long-term memory. Several mechanistic accounts of direct testing effects have been proposed. The *elaborative retrieval hypothesis* proposes that testing enhances learning because retrieving the answer for a question in a practice test results in activation of similar information and thus strengthens the integration of the information being tested (Carpenter, 2009). This proposal aligns with the broader *levels-of-processing framework,* which suggests that the strength and duration of a memory trace depends on how deeply that information in memory is processed (Craik & Lockhart, 1972). For example, rather than just memorizing terms regarding animal biology, one could try to relate these terms and concepts to their own pet. This elaboration of information during testing makes learners rely on their previous knowledge and increase their overall memory retention. Similarly, the *mediator effectiveness hypothesis* (Pyc & Rawson, 2010) suggests that practice testing creates improved mediators that link information to cues. Mediators are mental processes that occur between the onset of a stimulus and of a corresponding response. These cues, when recalled, allow the target information to be activated and then recalled. More broadly, the testing effect is consistent with the principle of *transfer appropriate processing* (Morris, Bransford, & Franks, 1977) — that the conditions that are most effective for learning are those that most resemble the conditions under which knowledge will eventually need to be deployed. Practice testing, compared to restudying, creates conditions during learning that are more comparable to a later test. Lastly, Lehman, Smith, & Karpicke (2014) have argued that retrieval practice pushes learners to learn items and facts as being distinct from each other, rather than generalize items into one context.

In sum, practice testing is not just memorization; practice testing activates elaborative processes during learning and, as a result, creates memory structures that aid in the retrieval of information in a later test of memory retention. Thus, retrieval practice requires more active engagement because effortful retrieval of the material (e.g., remembering the right answer when responding to a question) is necessary. In the same way, these theories support the idea that retrieval itself modifies one's knowledge; it is not simply a check on what information one already knows.

1.1.3 The Indirect Testing Effect

Practice testing can also support learning through an indirect route, especially when feedback is given (as I discuss below). Through this *indirect testing effect*, practice tests allow learners to assess their current performance level and prevent the *illusion of knowing* (Glenberg, Wilkinson, & Epstein, 1982) – when an individual feels confident in their understanding even though their understanding and comprehension of the material has actually failed. Practice testing wards off the illusion of knowing because if one takes an exam and struggles, the learner can realize they actually know less than they had thought. Consequently, practice testing may improve metacognitive processes, such as the ability to identify what one knows and does not know (Efklides, 2014). Practice testing could also improve the process of *self-regulated learning* (e.g., Barenberg & Dutke, 2019; Soderstrom & Bjork, 2015) – the process whereby learners make their own decisions about how to study, like what study strategies to use. Specifically, once learners are familiar with the testing effect and understand the benefits of testing, this could lead them to adopt it in the future (Tullis, Finley, & Benjamin, 2013).

1.1.4 Feedback and the Testing Effect

In the traditional testing-effect experimental paradigm, participants do not get feedback whether they got the answer correct or not when practicing retrieval. This is typically done to equate the number of exposures to the to-be-learned material: All participants practice the material once with a single study strategy, either by retrieval or restudying. By comparison, when feedback is provided in a retrieval practice condition, material in that condition is encountered twice: once via retrieval and once via getting feedback about their response.

Nevertheless, while testing has been found to be beneficial even without feedback, testing with feedback can lead to even better learning outcomes (Rowland, 2014; Yang et al., 2021; c.f., Adesope et al., 2017). Prior research has shown that feedback, both immediate and delayed, can lead to better memory retention of material when learning via a multiple-choice test (Butler, Karpicke, & Roediger, 2007). Corrective feedback (i.e., communicating to the learner why their answer was wrong and why the correct answer is right) is especially helpful with multiple-choice questions, compared to short answer testing, for long-term retention of material (Kang, McDermott, & Roediger, 2007). Further, both immediate and delayed feedback both can lead to greater memory gains on a cued memory recall test, compared to no feedback (Butler & Roediger, 2008).

Feedback is advantageous because it is helpful with alleviating intrusions (i.e., the false "recollection" of incorrect information; Butler & Roediger, 2008). Feedback also increases the likelihood a correct response will be retained and recalled when a learner is correct but has low confidence in their response (Agarwal, Bain, & Chamberlain, 2012; Butler, Karpicke, & Roediger, 2008; Fazio, Huelser, Johnson, & Marsh, 2010; c.f., Pashler, Cepeda, Wixted, & Rohrer, 2005).

Given the known relevance of feedback as a moderator of the testing effect, I here examine my effects of interest both with and without feedback. In Experiment 1, I focus on the moderating effect of background topic knowledge on testing exclusively in the case where feedback was not provided. In Experiment 2, I manipulate feedback orthogonally of background topic knowledge to explore the role of feedback on testing.

1.1.5 Summary

Overall, prior work has established the testing effect as a potent learning tool; in fact, the use of retrieval practice has been suggested as a great example of how cognitive science can inform education (Roediger, Putnam, & Smith, 2011; Rohrer & Pashler, 2010). For example, Roediger, Putnam, and Smith (2011) proposed ten benefits of testing with respect to education, such as filling in gaps in knowledge, improving metacognitive monitoring, and helping students study more.

While the benefits of retrieval practice have been well-established and are very educationally relevant, additional research is needed to answer more practical questions, such as how best to use it (i.e., under what conditions) and who would benefit the most from it (i.e., what works for whom). For example, educational activities generally need to reach a broad range of students with varying levels of expertise, so it is critical to establish whether retrieval practice is beneficial for all students or just for some. Thus, it is critical to explore possible moderators of the testing effect that may impact its effectiveness, like background topic knowledge.

1.2 Background Topic Knowledge and Learning

1.2.1 Metacognition and Background Topic Knowledge

Background topic knowledge is one highly plausible moderator of the testing effect because there is broad evidence that, across a variety of tasks and domains, prior topic knowledge exerts a powerful influence on learning (Chase & Simon, 1973; Chi, Feltovich, & Glaser, 1981; Huet & Mariné, 2005; Kalakoski & Saariluoma, 2001; for a review, see Vicente & Wang, 1998).

In sum, experienced ("expert") individuals tend to produce representations that go beyond what is explicitly shown to the learner, resulting in deeper and more elaborative understanding. In contrast, less-experienced ("novice") individuals concentrate on surface-level features of a problem that are clearly seen and presented. For instance, Chi and colleagues (1981) found that experts in physics (i.e., physics Ph.D. students) focused mainly on rules and equations governing a problem while novices (i.e., undergraduate physics students) relied on surface-level aspects of a problem (e.g., relying just the information shown in a given problem rather than tying it into a bigger concept).

Prior knowledge also matters because it has been shown to predict learning of new but related information. Witherby and Carpenter (2022) had participants complete a previous knowledge test for the domains of cooking and football, then had them learn new information regarding the two topics, and lastly presented a final test to assess their learning. Witherby and Carpenter found evidence for a *rich-get-richer* effect where prior domain knowledge helped participants learn new information about the same topic (e.g., having prior knowledge of football helped participants learn more new information about football).

Learners' previous knowledge can also affect their metacognitive processes and help them to effectively monitor their learning (cf. Löffler, von der Linden, & Schneider, 2016; Shanks & Serra, 2014; Toth, Daniels, & Solinger, 2011; Witherby & Carpenter, 2022). For example, Witherby and Carpenter found that, with new domain-relevant information, learners often make judgments based on their own learning that are consistent with their prior knowledge (e.g., they found that previous knowledge about football was positively correlated with the learning of novel football information). Additionally, when deciding what to study, learners often focus on information for which they judge their prior knowledge to be low (Ariel, Dunlosky, & Bailey, 2009; Thiede, 1999). On the other hand, prior knowledge can also lead to memory errors, where learners overestimate their knowledge on a given topic or have false memory related to the topic at hand (Castel, McCabe, Roediger, & Heitman, 2007; van Loon, de Bruin, van Gog, & van Merriënboer, 2013). If one's prior knowledge on a topic is false, false memories can arise about the presented topic.

It's also important to consider *cognitive load theory* and how prior knowledge may impact one's learning (Chen, Castro-Alonso, Paas, & Sweller, 2018; Hanham, Leahy, & Sweller, 2017; Sweller, van Merriënboer, & Paas, 2019). Cognitive load theory proposes that humans have a limited working memory capacity, and information overload can occur when information goes beyond this finite capacity. If learners already have more prior information stored in their long-term memory regarding a given topic, they are less reliant on their working memory to answer questions and can therefore better encode new information.

1.2.2 The Bifurcation Model

The relevance of prior knowledge to the testing effect is implied by the *bifurcation model*. The bifurcation model of Kornell, Bjork, and Garcia (2011) proposes that testing produces one of two differential outcomes for each item. If a learner recalls an item correctly during retrieval practice, this item gets a strong boost to memory strength, and testing clearly outperforms restudy. On the other hand, if a learner does not recall an item during retrieval practice, this item will not get a memory boost and it is more prone to being forgotten. As a result, there is a "bifurcation" where some items do get a benefit from retrieval practice, but others do not. By contrast, restudying, such as through rereading, provides a weaker boost, but it applies to all items, regardless of how strong initial knowledge is.

Thus, how much retrieval practice could be expected to benefit one's memory, relative to restudying, is dependent upon how well a learner can perform on the initial test, which is highly likely to depend on one's level of existing knowledge about a topic. This is further substantiated in the meta-analysis by Rowland (2014). For example, a true novice who does not know anything about a topic might not be able to do meaningful retrieval practice. In contrast, if someone already has some baseline knowledge of a topic, retrieval practice will likely benefit this individual since they know enough to be successfully tested on it. The more initial items correct, the more protected these items are from being forgotten, so more expert learners will potentially benefit more from retrieval practice, compared to more novice learners who may get fewer initial items correct.

1.2.3 Background Topic Knowledge and the Testing Effect

While the robustness of the testing effect is well-established, and the literature review above suggests background topic knowledge could plausibly moderate the testing effect, there is not a large body of work that empirically explores how background topic knowledge moderates the effectiveness of retrieval practice.

Some studies that have explored this relationship by examining the correlation between learners' prior topic knowledge and the size of the testing magnitude. For example, in one study with undergraduates in an educational psychology class, retrieval practice benefitted all students regardless of their previous topic knowledge, and the benefits were the greatest when students were studying unfamiliar content (Cogliano, Kardash, & Bernacki, 2019). In contrast, Carpenter and colleagues (2016) found that retrieval practice is more beneficial for learners with high knowledge (Carpenter, Lund, Coffman, Armstrong, Lamm, & Reason, 2016). In their study, undergraduate students in a biology course performed in-class exercises where they either retrieved information (e.g., remembering definitions for different vocabulary terms) and received feedback or just copied the necessary information down without retrieval. Carpenter and colleagues found that higher-performing learners benefited the most from retrieving and compared to middle-performing and lower-performing learners that benefited more from copying. Moreover, another study found that undergraduates in a psychology course performed better when they studied via a practice test (i.e., retrieval practice) with corrective feedback compared to restudying, regardless of prior knowledge (Glaser & Richter, 2023; for null results using a pre-experimental measure of prior knowledge, see also Xiaofeng, Xiao-e, Yanru, & AiBao, 2016). The mentioned studies have very different results – one finds testing most helpful for high-performing learners, another finds a compensatory benefit for low-performing learners, and a third finds it is equally helpful for everyone. Consequently, more research is necessary to truly understand the role of prior knowledge in the testing effect. Further, a limitation common to all of the studies is that they are correlational, so of course one cannot infer causation from these findings. This being said, prior knowledge was not controlled or manipulated by the experimenters in these studies; there could be correlations with many possible confounding variables (e.g., motivation to study the presented topic, working memory, and general intelligence). Confounding variables such as these could be the actual cause of the given learning differences.

I know of only one study where background topic knowledge was manipulated experimentally to study its moderating effect on the testing effect (Buchin & Mulligan, 2023). Buchin and Mulligan trained participants on either historical geology (e.g., geologic time, minerals, rocks, and isotopic dating) or sensation and perception (e.g., color perception, auditory perception, cutaneous senses, and chemical senses). In the training phase of three days, participants were randomly assigned to learn about three of the four sub-topics for either the domain of historical geology or of sensation and perception, one sub-topic per day for about forty-five minutes each day. Next, participants immediately entered the learning phase where they read four short passages; three topics participants were already trained on, and one topic was novel to the participant. Two days later participants, regardless of condition assignment, took a final test on both historical geology and sensation and perception. The final test had questions nearly identical to the information shown in the learning phase ("retention"), questions that were similar to what was studied in the learning phase ("near-transfer"), and questions about information that was initially studied in the training phase, but not retrieved in the learning phase ("far-transfer"). Buchin and Mulligan also found that participants performed significantly better on the test questions when they studied via retrieval practice, compared to restudying. Critically, retrieval practice was equally effective for both levels of prior knowledge on a topic. However, given that only one known study has explored this phenomenon, more research is necessary.

1.2.4 The Present Study

In the present study, I present two experiments examining background topic knowledge as a potential moderator of the testing effect among college students reading expository science texts. For both Experiments 1 and 2, both study strategy (testing versus restudying) and availability of background material for a given topic were orthogonally manipulated. For Experiment 2 only, I additionally manipulated feedback, such that half of the participants received immediate feedback when studying via retrieval practice, and the other half did not receive feedback.

I present three competing hypotheses for these studies. First, it is possible that participants who receive background material and study via retrieval practice will perform better than participants who do not receive background material and study via restudying. Given that retrieval practice has proven to be a superior study strategy compared to restudying and background knowledge provides additional expertise, it seems likely that learners in these seemingly optimal conditions will have the best memory retention. In sum, retrieval practice would be beneficial for expert learners, but not for novice learners.

Second, it is feasible that retrieval practice is beneficial particularly for those who are more novice on a given topic. For the novice learners, retrieval practice could be compensatory, where testing is so effective that it helps novice learners catch up to expert learners. This would align with the analogous finding that individuals with lower working memory capacity showed a greater, compensatory benefit from retrieval practice compared to those with higher working memory capacities (Agarwal, Finley, Rose, & Roediger, 2017) as well as one of the correlational studies mentioned above (Cogliano, Kardash, & Bernacki, 2019).

A third possibility is that the testing effect works equally well for everyone – as suggested by the correlational results of Glaser and Richter (2023) and the experiment by Buchin and Mulligan (2023). Perhaps the testing effect is so robust that all learners can benefit from it, regardless of one's level of background knowledge about a topic.

2.0 Experiment 1: Materials and Methods

2.1 Design

Experiment 1 used a 2 x 2 mixed design. First, within-subjects, each participant reviewed two topics using testing and two topics using restudying, as is typical in retrieval-practice study. Second, each participant was assigned to one of two between-subjects conditions, which will also be referred to as the between-subjects participant condition. In the *retrieve high knowledge* condition, participants first received additional background knowledge about the topics they would later use retrieval practice with and did not receive background knowledge about the topics they would restudy. Conversely, in the *retrieve low knowledge* condition, participants first received additional background knowledge about the topics they would later restudy and did not receive background knowledge about the topics they would use retrieval practice with.

If background topic knowledge strengthens the testing effect, I would expect the testing effect to be larger (or only exist) for the participants in the retrieve high knowledge condition than those in the retrieve low knowledge condition. Conversely, if retrieval practice has a compensatory effect in the face of low domain knowledge, I would expect the testing effect to be larger or only present for participants in the retrieve low knowledge condition than those in the retrieve high knowledge condition.

The experiment consisted of two sessions (a learning session and a final test session) separated by one week.

2.2 Participants

A power analysis conducted in G*Power revealed a target sample size of 129 participants to detect a small-to-medium effect size (R^2 = .02, power = .80) for a main effect of retrieval practice. All participants were University of Pittsburgh undergraduate students enrolled in an introductory psychology course and were recruited via Sona Systems in exchange for partial course credit for full completion of both sessions of the study. Participants were required to speak fluent English, be eighteen years of age or older, and have access to a computer. There was approximately 23% attrition between session one (N = 223) and session two (N = 172).

2.3 Materials

2.3.1 Session One Background and Main Passages

Learners read about four different science topics: dinosaurs, comets, the Great Barrier Reef, and acupuncture. The dinosaurs, comets, and acupuncture passages were from Norberg (2022), and the Great Barrier Reef texts were created from texts from Readworks (Readworks, 2013b), the Great Barrier Reef Foundation (Great Barrier Reef Foundation, 2021), and the Coral Reef Alliance (Coral Reef Alliance, 2021). Within each topic, there were three main passages and three background passages, one per subtopic. Each main passage was approximately 200 words, and each supplementary background reading was about 400 words long (see Appendix A for main passages and Appendix B for background passages).

The dinosaurs, comets, and acupuncture passages – both the background and main passages – had been previously constructed by Norberg (2022) to be divided into three subtopics of difficulty: one easy, one medium, and one hard. I constructed the new Great Barrier Reef passages – both the background and main passages – to be composed of three subtopics to mirror the design of the other three topics, but these three subtopics were not normed to vary by difficulty. The difficulty of Norberg (2022) texts was constructed in this way simply due to the research questions of Norberg (2022); I did not have any hypotheses about how difficulty would interact with my research questions of interest[1].

2.3.2 Session One Retrieval Practice Questions and Restudying Sentence Facts

During session one, participants studied two topics via retrieval practice and two via restudying, depending on condition assignment. For each topic, I created twelve items for participants to learn (forty-eight questions total). Three items corresponded to the easy subsection, three to the medium subsection, three to the hard subsection, and three to the passage as a whole. The difficulty of questions for dinosaurs, comets, and acupuncture topics was normed in Norberg (2022). The Great Barrier Reef items were coded by difficulty to mimic the mean accuracies of the difficulty divisions seen in Norberg (2022) and were thus divided in the same way: three for easy difficulty, three for medium difficulty, three hard difficulty, and three full passage questions.

[1] Norberg (2022) explored learners' ability to identify and overcome comprehension failure regarding well-learned, partially learned, or completely unlearned material. Norberg (2022) also had learners identify the difficulty of the material to further explore learners' beliefs about learning.

This was done on a post-hoc exploratory basis to explore the possible impact of difficulty on the testing effect and background topic knowledge (see Appendices E and F).

Across the retrieval practice and restudying conditions, I sought to control the presented content as closely as possible and vary only how participants interacted with it. For the retrieval practice conditions, each item was encountered in the form of a three-option multiple-choice practice question (e.g., How has the perception of dinosaurs changed? (a) It was once believed that they were slow and not likely to survive, but we now believe they could move more quickly and easily. (b) It was once believed that they only existed for 300,000 years, but now we know they lasted 160 million years. (c) It was once believed that dinosaurs could swim, but now we know they were too big to swim.; (a) indicates the correct answer). For the restudy condition, participants read a sentence that mimicked the information shown in the multiple-choice question and answers but was instead in a sentence form. The "answer" portion of the sentence was bolded to emphasize the important part of the sentence, as one would focus on the answer options in a retrieval practice question. (e.g., It was once believed that they were slow and not likely to survive, but we now believe they could move quickly and easily) (see Appendix C for the full set of retrieval practice and see Appendix D for the full set of restudying sentence facts).

2.3.3 Session Two Materials

Session two consisted of a final test and a few exploratory measures. The final test contained the same forty-eight questions as the session one retrieval practice questions. Along with the final test, participants also completed the 50-question version of the Big 5 Personality Test (John, Donahue, & Kentle, 1991), the 81-question version of the Motivated Strategies for Learning

Questionnaire (MSLQ) (Pintrich & De Groot, 1990), and self-reported their undergraduate GPA based on a 4.0 scale, which were collected on an exploratory basis.

2.4 Procedure

All study methodology was approved by the University of Pittsburgh's Institutional Review Board (IRB). All sessions of the presented experiments took place online on Qualtrics (https://www.qualtrics.com/).

Learners participated in session one at any time; session two was then made available one week later and available for up to twenty-four hours. Session one consisted of the background topic knowledge intervention, reading all of the main texts for each of the four topics, and the study strategy manipulation (see Figure 1). Session two consisted of a final retrieval practice test for all four topics and a few exploratory measures (see Figure 2).

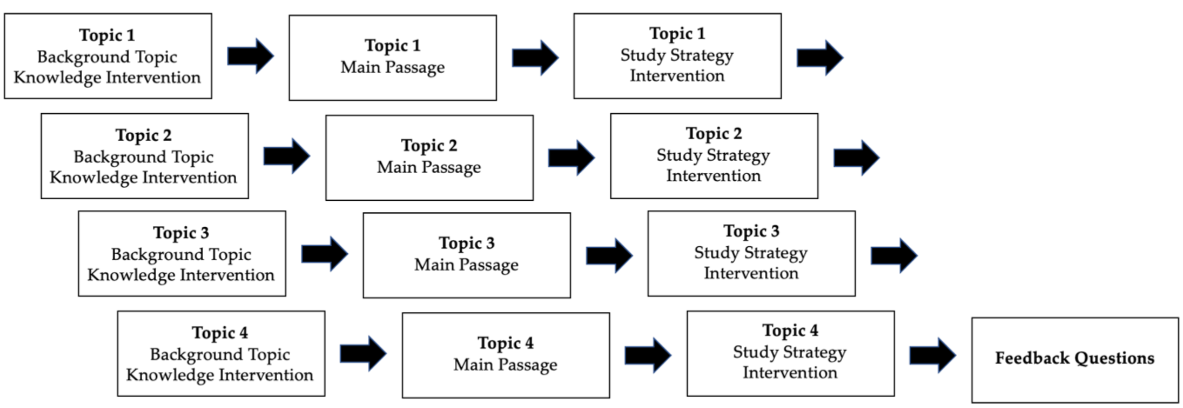


Figure 1

Flow of activities for session one in Experiment 1.

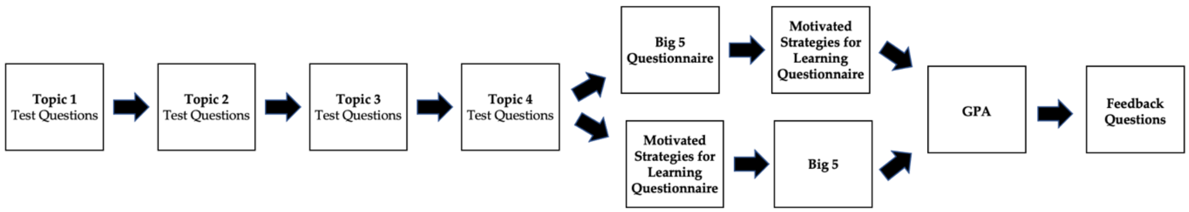


Figure 2

Flow of activities for session two in Experiments 1 and 2.

2.4.1 Session One

First, participants encountered the background topic knowledge intervention. Participants covered each of the four topics in sequence. For each topic, participants received background material (potentially), then the main text, and then one of the study activities (restudy or testing). Then, the procedure repeated for the next topic. Participants were randomly assigned to read a background knowledge text for two of the four topics. Participants had 115 seconds to read each background passage before Qualtrics would auto advance. This timing was determined based on prior norming conducted by Norberg (2022) as well as participant feedback during my initial piloting. (In my debriefing survey, 92% of participants indicated this provided enough time to read the background passage). For the remaining two topics, participants did not receive background material.

Following the background topic knowledge intervention, all participants received a main passage for each of the four topics. Participants had 60 seconds to read each main passage before Qualtrics would auto advance. As with the background passages, this timing was determined due to the timing used in Norberg (2022) as well as participant feedback during an initial norming study. (94% of participants stated that they had ample time to read the main passage).

The study strategy manipulation followed the main passages. For two of the topics, learners were randomly assigned to use retrieval practice to study, and for the other two topics, learners used restudying. The retrieval practice condition was self-paced, but participants needed to answer all twelve of the topic questions to proceed[2]. The order of the twelve practice questions, as well as the order of the three response options within each question, were randomized for each participant. Participants in the restudy condition had 60 seconds to read each of the twelve sentence facts affiliated with that topic before Qualtrics automatically advanced to the next fact. This timing was determined by fact that the number of words read during the restudying condition was comparable to the number of words contained in a main passage, so participants were allocated the same amount of time. Further, I also wanted participants to not click through the restudy items as quickly as possible. The presentation of these twelve sentence facts was also randomized.

Critically, prior norming by Norberg (2022) established that none of the information provided during the study strategy intervention phase of the experiment required the background passage[3]. The information in the background passage was given to provide additional support for the participants' learning of the main passage, and participants were tested only on information provided in that main passage.

---

[2] There was no feedback following each retrieval practice question. As discussed above, this controls the number activities and exposures to the material and is standard in retrieval practice literature. Experiment 2 will explore the role of feedback on accuracy with retrieval practice questions.

[3] Participants could not answer the retrieval practice questions with the background passage alone; norming by Norberg (2022) established that participants needed the main passage in order to be able to effectively answer the presented retrieval practice questions.

Participants experienced this flow of activities for all four topics. The topics and condition assignments were counterbalanced across eight presentation lists so that, across participants, each topic appeared in each serial position and in each condition of the background-topic-knowledge and study-strategy manipulations.

At the end of session one, participants answered a series of 1-to-5 Likert-scale questions (1 indicating a low response and 5 indicating a high response) and some free-response questions. The Likert-scale questions asked participants about their knowledge of each of the four presented topics (e.g., *How familiar were you with the material on Dinosaurs before the experiment*?) and how they felt about the task (e.g., *How much did you enjoy the task?*) The free-response questions asked participants to discuss their strategy (*Please describe your overall strategy during the task.*) and describe their experience during the study (*We just wanted to ask you about your overall experience with the task. Also, let us know if you have any additional comments and if you have any suggestions for improvement.*). This exploratory analysis of these questions is presented in the Appendices E and F.

2.4.2 Session Two

Session two took place one week after session one, and participants had twenty-four hours to complete it. In session two, participants took a final multiple-choice test regarding the passages all the participants read (i.e., the main passages). The final test included twelve questions per topic (ranging in difficulty, as described above), resulting in forty-eight total questions. Participants were tested on each topic one-at-a-time, with all twelve questions for that topic appearing simultaneously. The final quiz was self-paced, but participants needed to answer all the presented questions in order to proceed. These questions were identical to the questions that would be presented in the session one retrieval practice condition. The order of presentation of topics was counterbalanced across stimulus lists. The question order and question answers were also randomized, in the same way as they were in session one.

Following the forty-eight questions, participants completed the Big 5 Questionnaire, the MSLQ, and entered their GPA. These questionnaires were self-paced, but participants needed to answer each question in each questionnaire in order to proceed. The order of the questionnaire items for both questionnaires was randomized. The GPA measure was also self-paced; participants just needed to enter a value in order to proceed. Participants ended session two after completing a free-response question regarding their overall strategy during the task (*We just wanted to ask you about your overall experience with the task. Also, let us know if you have any additional comments and if you have any suggestions for improvement.*).

2.5 Summary

Experiment 1 employed a 2 x 2 mixed design resulting from the crossing of a within-subjects study strategy intervention (i.e., retrieval practice or restudying) and a between-subjects participant condition (i.e., retrieve high knowledge and retrieve low knowledge). Each participant received background materials for some topics but not for others, and each participant received retrieval quizzes for some topics and restudying for others. These variables (i.e., retrieval practice versus restudying and having background topic material versus not having background topic material) were counterbalanced to ensure that all the effects were de-confounded with the topic themselves. This resulted in a total of eight different presentation lists for both session one and session two.

3.0 Experiment 1: Results

All mixed-effects logit models were run using the R package *lme4* (Bates, Mächler, Bolker, & Walker, 2015). The dependent variable was the accuracy of each item on the session two final test, scored on a binary basis as either 1 or 0, where 1 indicated a correct response and 0 indicated an incorrect response. There were forty-eight questions total, twelve per topic.

Fixed effects reflected experimental condition assignment and were contrast-coded in analyses to obtain main effects analogous to that of an ANOVA. For Experiment 1, there were two fixed effects: study strategy intervention and the between-subjects participant condition variable (i.e., retrieval high knowledge and retrieval low knowledge). The study strategy intervention was contrast-coded with retrieval practice coded as 0.5 and restudying as -0.5, and the between-subjects participant condition variable was contrast-coded with the retrieve high knowledge condition as 0.5 and the retrieve low knowledge condition as -0.5.

The random effects of the model initially included random intercepts for participants, the forty-eight final test questions, the broader passage topic, and the subsection difficulty. Following model comparisons via the likelihood ratio test, I eliminated random intercepts that did not contribute significantly to the model, and the final model contained only the random intercepts for participant and test question.

Table 1 presents results from the main model. There were not significant main effects of study strategy (p = .18) or the between-subjects participant condition variable (p = .81)[4]. There

---

[4] Given the design of Experiment 1, it would be incorrect to treat the background topic material as a within-subjects variable. However, when this main effect is included in analyses as a within-subjects variable, there is a significant main effect of background topic material (p < .01) where

was a significant two-way interaction between study strategy and the between-subjects participant condition variable ($p < .01$). Follow-up Tukey tests found a significant effect of study strategy in the retrieval-low knowledge condition ($p < .01$) but not in the retrieval high knowledge condition ($p = .66$). In the retrieve low knowledge condition, I found a reverse testing effect such that participants performed better on the session two test when they had studied via restudying (and had access to background topic material) then they studied via retrieval practice (and did not have access to background topic material). Figure 3 shows a graphical representation of the mean accuracies. (see Appendix E for additional analyses exploring the role of difficulty on session two test performance.)

Table 1

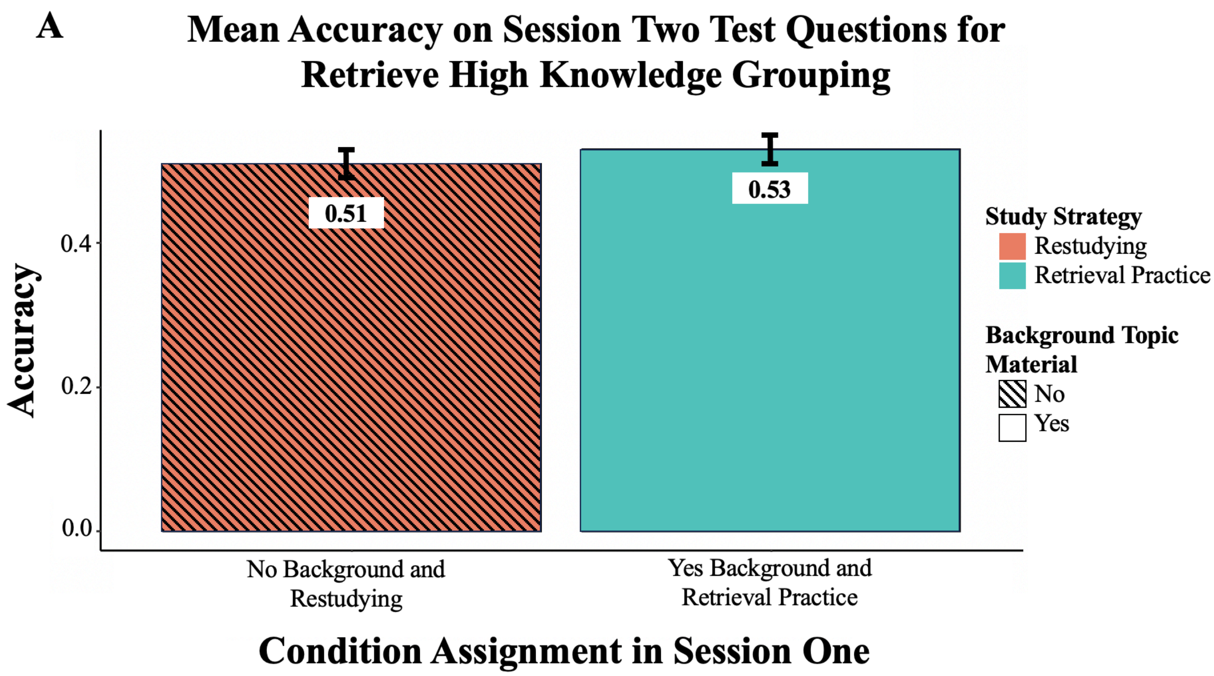
Results from mixed-effects model for Experiment 1.

***Note*. Coefficients = the estimate of each effect of each variable on accuracy on session two final test, as reported in log odds; SE = standard error of the estimate; ** *p* < .01**

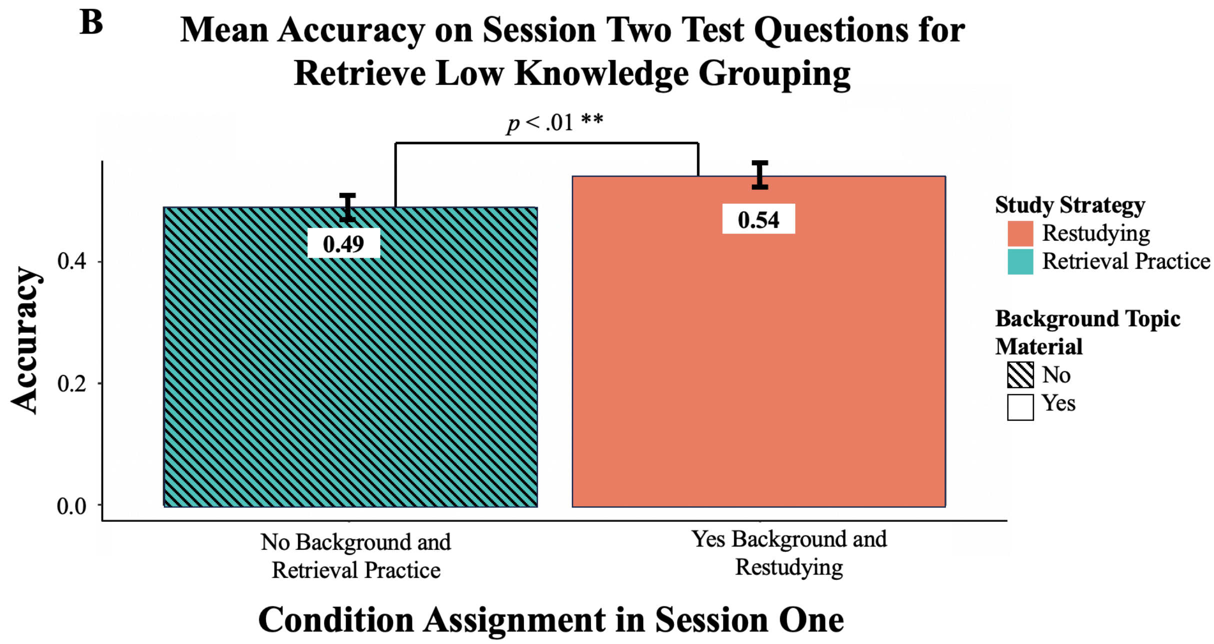
	Estimate	SE	*z*-value	*p*-value
Study Strategy Intervention	-0.07	.05	-1.34	.182
Between-subjects Participant Condition	0.02	.07	0.24	.812
Study Strategy Intervention X Between-subjects Participant Condition	0.32	.10	3.09	.002 **

participants performed significantly better on the session two test when they receive background topic material compared to not receiving background topic material. The main effect of study strategy and the interaction between study strategy and background material were not significant.

A



B



Figures 3A & 3B

Figure 3A shows the mean accuracy on the session two test of Experiment 1 based on session one condition assignments for retrieve high knowledge and Figure 3B shows the mean accuracy on the session two test of Experiment 1 based on session one condition assignments for retrieve low knowledge. Error bars signify the 95% confidence interval.

All exploratory measures (free-response questions, Likert-scaled questions, Big 5, MSLQ, and GPA) were also analyzed (see Appendix E).

4.0 Experiment 1: Discussion

In Experiment 1 both study strategy (testing versus restudying) and availability of background material for a given topic were orthogonally manipulated to explore the possible moderating effect of background topic knowledge on the testing effect. I found that participants who studied via restudying and had access to background topic material performed significantly better on the session two test compared to learners who studied via retrieval practice and did not have access to background topic material. While not significant, we also see that participants who received background topic material and studied via retrieval practice did perform numerically better on the session two test compared to learners to studied via restudying and did not receive background topic material.

These results confirm the relevance of my manipulation of topic knowledge: participants who received the background text about a topic outperformed those who did not, indicating the background was relevant to learning about the topics. Nevertheless, even when participants did not receive the background topic material, they performed above chance (i.e., above 33%)[5], demonstrating that the background topic material was not *required* to understand the texts (consistent with the norming by Norberg, 2022).

On the other hand, my primary question was whether background topic material would moderate the testing effect. However, I did not in fact find a traditional testing effect — where

---

[5] Participants in both Experiment 1 and Experiment 2 who were performing at or below chance on the session two test (i.e., having an overall mean accuracy on the final text of 33% or lower) were removed from analyses. The removal of these participants did not change the significance of any of the main effects or interactions for either experiment.

retrieval practice enhances later test performance relative to restudy — at all, so the Experiment 1 data were uninformative for this question. To better produce a testing effect and thereby test for its potential moderators, in Experiment 2 I included another manipulation known to strengthen the testing effect: feedback.

5.0 Experiment 2: Materials and Methods

Experiment 2 differs from Experiment 1 because it also explores the role of feedback. Two versions of Experiment 2 were run concurrently during the academic school year. The two experiments were combined to conduct a between-subjects comparison of feedback (i.e., receiving or not receiving feedback on session one retrieval practice questions). This between-subject variable will also be referred to as "experiment."

5.1 Design

Experiment 2 used a 2 x 2 x 2 mixed design. First, as in Experiment 1, within-subjects, each participant reviewed two topics using testing and two topics using restudying. Second, within-subjects, each participant received a background passage for two topics and no background passage for the remaining two topics; unlike in Experiment 1, this variable was manipulated orthogonally of study condition, fully within-subjects. Third, each participant was randomly assigned to one of two between-subjects conditions. For about half of the participants, they did not receive feedback on the retrieval practice questions in session one. In contrast, approximately half of the participants received immediate feedback after each retrieval practice question during session one. For the participants who received feedback, the questions in session one were presented in the same order to all participants, but question answer options were still randomized. In session two, the question order and answer order were still randomized as they were in Experiments 1. The between-subjects variable (i.e., receiving or not receiving feedback) was originally two separate studies (Experiment 2A and Experiment 2B). Given that these two studies were run concurrently and varied only by the immediate feedback manipulation, the two experiments were combined to serve as Experiment 2.

As with Experiment 1, participants would participate in session one of the study at any time and then participate in session two one week later. Session two was available for twenty-four hours.

5.2 Participants

A power analysis conducted in G*Power revealed a target sample size of 105 participants per version of the study (i.e., receiving immediate feedback on the session one retrieval practice questions, or not) to detect a medium effect size (R^2 = .04, power = .80) for a main effect of retrieval practice. Participants were recruited via the same methods as Experiment 1 with the same inclusion criteria. With respect to the no-feedback version, there was approximately 14% attrition between session one (*N* = 138) and session two (*N* = 119). Regarding the immediate feedback version, there was approximately 21% attrition between session one (*N* = 148) and session two (*N* = 117). Thus, Experiment 2 had a total sample size of *N* = 236.

5.3 Materials

For the no-feedback version of Experiment 2, the materials were the same as with Experiment 1. The materials used in the immediate feedback version of Experiment 2 were nearly the same as Experiment 1. The only difference is that for the session one retrieval practice questions, participants received immediate feedback following each question. The feedback would restate the question, state "Correct answer:", and then state the correct answer to the question. This feedback block was shown for 10 seconds before Qualtrics auto advanced to the next question.

5.4 Procedure

As with Experiment 1, both sessions of Experiment 2 took place online via Qualtrics. The procedure of Experiment 2 was nearly identical to Experiment 1 except that, in Experiment 2, before beginning the experimental tasks, participants were asked Likert-scale questions regarding their self-perceived knowledge of the four topics (e.g., *How familiar are you with the topic of Dinosaurs*?) (see Figure 4 for the flow of activities for session one and see Figure 2 for the flow of activities for session two). Participants then began the main experiment, and it proceeded the same way as Experiment 1.

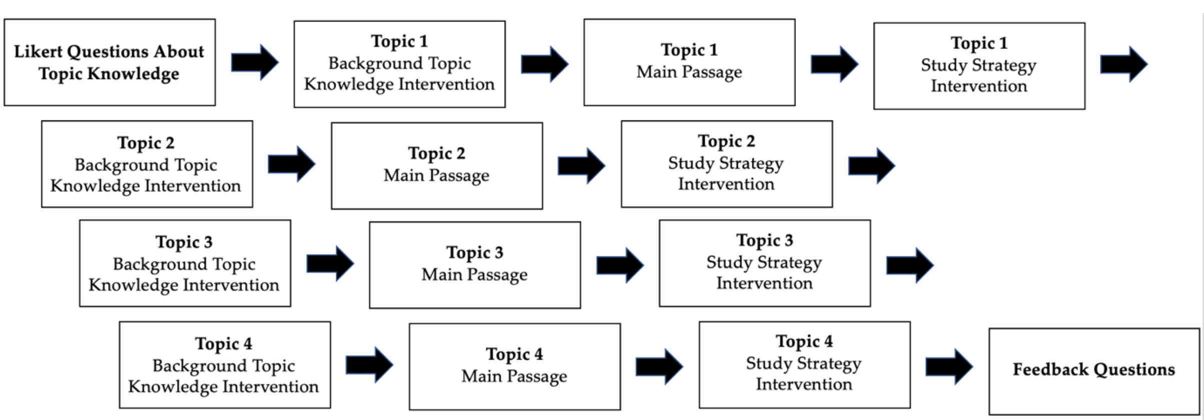


Figure 4

Flow of activities for session one in Experiment 1.

5.5 Summary

Experiment 2 had a 2 x 2 x 2 mixed design: a within-subjects study strategy intervention (i.e., retrieval practice or restudying), a within-subjects background topic knowledge intervention (i.e., receiving a background passage for a given topic, or not), and a between-subjects feedback condition (i.e., receiving immediate feedback in the session one retrieval practice condition, or not). All variables were counterbalanced, resulting in four different presentation lists for both session one (as compared to the eight presentation lists in Experiment 1 session one) and eight different presentation lists for session two (as with Experiment 1).

6.0 Experiment 2: Results

For Experiment 2, there were three fixed effects: study strategy intervention, background topic knowledge intervention, and experiment (feedback or no feedback). Study strategy was coded the same as in Experiment 1, and the background topic knowledge intervention was contrast-coded for receiving background topic knowledge (0.5) or not receiving background topic knowledge (-0.5). The between-subjects experimental variable regarding feedback was contrast-coded to compare feedback (0.5) to not having feedback (-0.5).

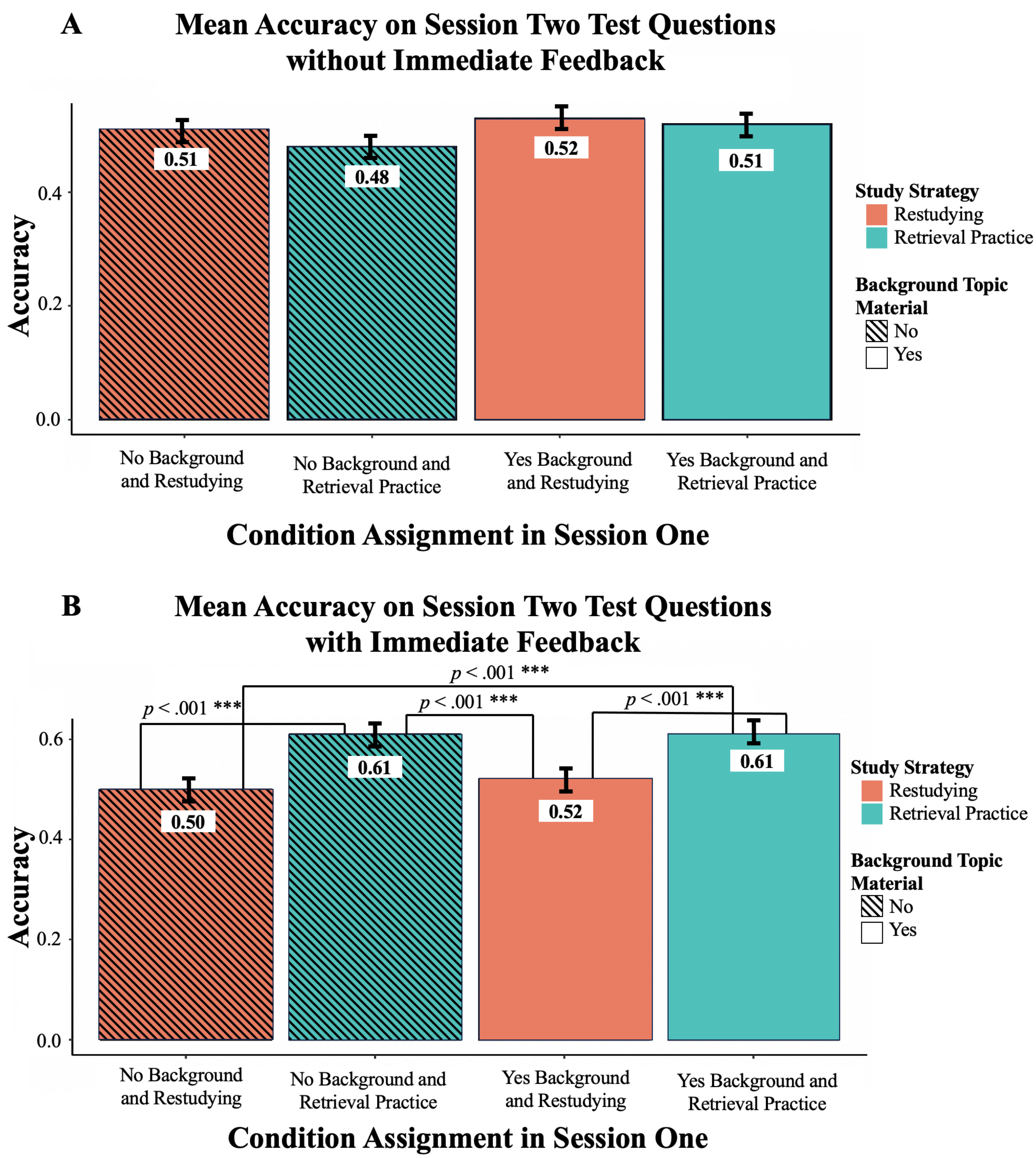
Table 2 presents results from the main model. There was a significant main effect of study strategy ($p < .001$) such that participants performed significantly better on the session two test if they studied via retrieval practice than restudying, replicating the classic testing effect. There was also a main effect for the between-subjects experimental variable of feedback ($p < .001$). Participants performed better if they received feedback on their session one retrieval quiz questions. Further, there was a significant two-way interaction between study strategy and feedback ($p < .001$) such that the testing effect was amplified if participants received feedback. There was a main effect of background topic knowledge ($p < .01$), such that participants' learning of the main text was improved if they had access to background material. However, background topic knowledge did not significantly interact with any of the other variables. Figure 5 shows a graphical representation of the mean accuracies (see Appendix F for additional analyses exploring the role of difficulty on session two test performance).

Table 2

Results from mixed-effects model for Experiment 2.

***Note*. Coefficients = the estimate of each effect of each variable on accuracy on session two final test, as reported in log odds; SE = standard error of the estimate; * *p* < .05, *** *p* < .001**

	Estimate	SE	*z*-value	*p*-value
Study Strategy Intervention	0.29	.04	4.68	< .001 ***
Background Topic Knowledge Intervention	0.07	.04	1.97	.049 *
Experiment Feedback Manipulation	0.22	.06	3.50	< .001 ***
Study Strategy Intervention X Background Topic Knowledge Intervention	0.02	.08	0.27	.790
Study Strategy Intervention X Experiment Feedback Manipulation	0.57	.08	6.66	< .001 ***
Background Topic Knowledge Intervention X Experiment Feedback Manipulation	-0.07	.08	-0.84	.400
Study Strategy Intervention X Background Topic Knowledge Intervention X Experiment Feedback Manipulation	-0.23	.17	-1.35	.176



Figures 5A & 5B

Figure 5A shows the mean accuracy on the session two test of Experiment 1 based on session one condition assignments for not receiving immediate feedback during retrieval practice and Figure 5B shows the mean accuracy on the session two test of Experiment 1 based on session one condition assignments for receiving immediate feedback during retrieval practice. Error bars signify the 95% confidence interval.

All exploratory measures (free-response questions, Likert-scaled questions, Big 5, MSLQ, and GPA) were also analyzed (see Appendix F).

7.0 Experiment 2: Discussion

In contrast to my findings from Experiment 1, I found a significant main effect of study strategy where participants performed significantly better on the session two test if they studied via retrieval practice than restudying, replicating the classic testing effect. I also found a main effect of "experiment," such that performance was higher overall in Experiment 2B (in which feedback was provided on retrieval practice) than in Experiment 2A (in which no feedback was provided). This main effect was presumably driven by the retrieval-practice condition, which was the only condition that differed across experiments, and indeed, I found a significant two-way interaction where the testing effect was amplified if participants received feedback (Experiment 2B).

Critically, there was no interaction of the testing effect with background topic knowledge, consistent with the findings of Buchin and Mulligan (2023). Is this simply because the background material was irrelevant to learners' understanding of the next? No — I did find a significant main effect of background topic knowledge where participants' learning of the main text was improved if they had access to background material, but, rather, background topic knowledge did not significantly interact with any of the other variables.

8.0 General Discussion

In two experiments testing college students' learning of expository science texts and retention one week later, I examined how the testing effect may be moderated by background topic knowledge, experimentally manipulated via additional expository science texts.

In Experiment 1, I did not find an overall testing effect at all; indeed, participants who studied via retrieval practice without access to the background material did significantly *worse* than those who had access to background material and studied via restudy. The absence of a general testing effect in this experiment meant it was not fruitful grounds for testing background topic knowledge as a potential moderator.

Thus, in Experiment 2, I added a manipulation of feedback, a variable that has been shown in prior work to enhance the effect of testing. This had a significant effect on session two performance: learners performed significantly better on the session two test if they received feedback on their session one retrieval practice quiz questions compared to no feedback. (An additional benefit of Experiment 2 was that it used a within-subjects manipulation of background topic knowledge and study strategy, unlike the between-subjects design of Experiment 1.) Although I found a significant overall testing effect in Experiment 2, and that feedback boosts the testing effect (as in past work), I found that this was *not* moderated by background topic knowledge; testing with feedback was beneficial regardless of whether or not participants had received prior background on the topic.

Given these findings, I can see that background topic material, or perhaps expertise more broadly, helps learners' memory retention regardless of study strategy even if immediate feedback is not provided. Further, feedback during retrieval practice proves to be effective beyond the boost from additional background information on a given topic. Critically, however, I found that these effects are independent of one another, such that the background material was beneficial with either restudy or retrieval practice, and the testing effect was beneficial with or without background material.

8.1 Potential Counter-explanations for Findings

I found that the testing effect was not moderated by background topic knowledge. Could this be because participants simply ignored the background material? This could be due to these experiments take place in a low-stakes context where it does not really matter if participants learn the material or not; introductory psychology students participating in this study to fulfill a course requirement likely are not particularly invested in learning the material. I argue this is very unlikely because the background material *did* have a significant main effect whereby participants performed significantly better on the session two test if they had access to background material during session one. Further, one could be concerned that the material contained in the background topic material is simply not relevant or helpful for understanding the main texts, but again, the significant main effect of background topic material mitigates this worry.

8.2 Comparison to Prior Work

These findings extend the existing literature exploring the moderating role of background topic knowledge on the testing effect. As previously mentioned in the introduction, there are some correlational studies that have explored this relationship, resulting in mixed findings. For instance, Cogliano and colleagues (2019) found that the testing effect was compensatory for learners with low knowledge. Instead, I found that background topic knowledge did not moderate the testing effect. Additionally, Carpenter and colleagues (2016) found that retrieval practice is more beneficial for learners with high knowledge, which again I did not find. Further, Glaser and Richter (2023) found that their participants performed better when they studied via a practice test with corrective feedback (i.e., retrieval practice) compared to restudying, regardless of prior knowledge. These results align with my Experiment 2 findings where I found that retrieval practice with feedback was most beneficial, and this was not affected by background material accessibility.

Buchin and Mulligan (2023) produced the only study I know of in which prior knowledge was manipulated experimentally to study its moderating effect on the testing effect. In contrast with the Buchin and Mulligan (2023), I only found a significant testing effect when participants received immediate feedback following their session one retrieval questions (Experiment 2) whereas Buchin and Mulligan found that participants always had greater performance in the retrieval practice condition compared to the restudying condition. Nevertheless, an important commonality with my study and theirs is that, regardless of my variation in methodology, we both found that background topic knowledge does not moderate the testing effect.

There were some other methodological differences between studies. In my study, all participants proceeded to session two regardless of their performance on their single encounter with the session one retrieval practice questions. By contrast, Buchin and Mulligan trained participants to criterion by requiring their participants to score above 80% during their training phase topic tests or else they were required to redo their training and test sessions. While this ensures that all participants had some level of background topic knowledge, my study may be more ecologically valid in that participants in formal education settings are not typically trained or tested in this way (i.e., in most high school or college courses, all students get a fixed amount of class time before taking a test). Further, the texts in Buchin and Mulligan were much longer than those used in my study; in Buchin and Mulligan's study, training texts were about 4000 to 4500 words (comparable to my background texts that were approximately 400 words each) and domain texts were about 3100 words (comparable to my main texts, which were approximately 200 words each). My study also took place over a one-week period while Buchin and Mulligan's study was over a three-day period. It is possible that, given my longer retention interval, there was more forgetting and thus the retrieval practice was found to be less effective. On the other hand, it is usually the case that the relative benefit of retrieval practice compared to restudying is larger at longer retention intervals (Agarwal et al., 2012; Chan, 2009; Roediger & Karpicke, 2006; Rowland, 2014; Toppino & Cohen, 2009; Wheeler, Ewers, & Buonanno, 2003; Yeo & Fazio, 2019). What is most noteworthy is that I reach the same overall conclusion as Buchin and Mulligan — that the testing effect is independent of background topic knowledge — despite these differences in procedure.

8.3 Theories Regarding Prior Knowledge and Testing

8.3.1 Revisiting the Elaborative Retrieval Hypothesis, Cognitive Load Theory, and the Bifurcation Model

My findings are somewhat in contrast with the *elaborative retrieval account* (Carpenter, 2009) or *cognitive load theory* (Chen, Castro-Alonso, Paas, & Sweller, 2018; Hanham, Leahy, &

Sweller, 2017; Sweller, van Merriënboer, & Paas, 2019). Both of these accounts predict greater benefits from retrieval practice than restudying for learners high in prior knowledge compared to the lower-knowledge learners. The *elaborative retrieval account* proposes that retrieval of information increases a learner's retention of material due to the fact that the retrieval process produces semantic mediators that connect cues to targeted information. This elaboration of information during retrieval practice requires learners rely on their prior knowledge and boost their overall retention. *Cognitive load theory* proposes that humans have a limited amount of working memory space, and information overload can occur when information exceeds this limited capacity. If learners already have information stored in their long-term memory regarding a topic, they are less reliant on their working memory to answer questions; and retrieval practice leads to better encoding of information compared to restudying.

In contrast to the key claim of the theoretical accounts that there are greater benefits from retrieval practice than restudying for learners high in prior compared to the low prior knowledge learners, I did not find a significant interaction between retrieval practice and background topic knowledge.

On the other hand, when feedback is provided during retrieval practice in session one, I see better retention in session two, compared to not receiving feedback. Perhaps the feedback is necessary in order to result in more elaborative activation of material, resulting in the superior performance on the session two test in Experiment 2. It seems possible that feedback leads to deeper and more focused processing of retrieval practice information, and prior research has shown that deeper processing of material results in better retention of the information, compared to more shallow processing (Adesope et al., 2017; Rowland, 2014; Yang, et al., 2021). This rationale is also in line with the *bifurcation model*, where testing such as via retrieval practice can help boost one's memory for correctly recalled information (Kornell, Bjork, and Garcia, 2011). Accordingly, how much testing could be expected to boost one's memory, compared to restudying, is dependent upon how well a learner does on the initial test, which is quite likely to rely on one's level of prior knowledge about the presented topic.

8.3.2 Episodic Context Account

My results also provide mixed support for the *episodic context account* (Karpicke et al., 2014), given that I did not find that retrieval practice was always more beneficial, regardless of accessibility to background information. The episodic context account suggests that the testing effect is due to the contextual reinstatement of information found on the initial test. Under this account, the reinstatement of information (i.e., to my session one retrieval practice quizzes) should aid retrieval on the final (session two) test. With the immediate feedback following the retrieval practice quiz questions in session one of Experiment 2, I did find that participants performed better on the session two test. Thus, the feedback could be the additional reinstatement necessary to boost memory retention.

8.4 Individual Differences and the Testing Effect

Researchers have also considered how the testing effect is moderated by other individual differences beyond prior expertise, such as working memory or general intelligence. Agarwal and colleagues (2017) found that retrieval practice can be more beneficial for students who have lower working memory capacities (Agarwal, Finley, Rose, & Roediger, 2017). In contrast, others have found that the testing effect is equally beneficial regardless of working memory capacity (Jonsson, Wiklund-Hörnqvist, Stenlund, Andersson, & Nyberg, 2021; Bertilsson, Stenlund, Wiklund-Hörnqvist, & Jonsson, 2021; Wiklund-Hörnqvist, Jonsson, & Nyberg, 2014).

Moreover, it's important to consider testing for one's intelligence rather than just their competence (McClelland, 1973). For example, being able to perform well on a standardized exam such as the GRE establishes competence to successfully complete that particular exam, but this does not establish innate intelligence more broadly and does not account for possible confounding variables like socioeconomic status. Similar hypotheses can be formed for the role of general intelligence as for the role of topic knowledge. For instance, it could be possible that people who have more general intelligence at large will benefit the most from testing, supporting the "*rich get richer*" idea. In contrast, it is feasible that the retrieval practice effect is compensatory where novice learners can catch up to experts. For example, Brewer and Unsworth (2012) found that there was a larger testing effect for lower general intelligence individuals compared to higher general intelligence individuals. It is also possible that the testing effect is equally beneficial regardless of general intelligence, as found in Jonsson et al., 2021.

Overall, prior literature does not clearly suggest that the testing effect is moderated by either working memory or general intelligence. This broadly aligns with my studies, where I also found that the testing effect was not moderated by prior topic knowledge; rather, testing with feedback was broadly effective in Experiment 2 regardless of whether participants had received the background topic knowledge.

Additional individual differences that have been considered regarding learning performance include achievement goals (Dweck, 1986; Harackiewicz, Barron, & Elliot, 1998), learners' need for cognition (i.e., the desire to engage with and enjoy more effortful cognitive processing, such as via generation of information; Cacioppo, Petty, & Kao, 1984; Schindler, Pfattheicher, & Reinhard, 2019), and emotional dispositions like test anxiety (Weissgerber and Reinhard; 2018), to name a few. More research is necessary to know if these factors do in fact moderate the testing effect.

8.5 Limitations and Future Directions

One limitation of my study is that I did not test for transfer (i.e., testing via questions that are similar to those found in session one, but not quite the same). It is important to note that transfer is often difficult to achieve (Gick & Holyoak, 1987; Haskell, 2001; Singley & Anderson, 1989), and its existence has been debated (e.g., Barnett & Ceci, 2002; Detterman, 1993; Singley & Anderson, 1989). In a meta-analysis, Pan and Rickard (2018) found that positive transfer from the testing effect Is dependent on *(a)* response congruency between initial and final testing (i.e., having the same or similar responses at the first and last tests), *(b)* if there was elaborative retrieval practice (e.g., elaborative feedback that explains why a selected retrieval practice answer is right or wrong), and *(c)* initial test performance. Given that I did not include more elaborative feedback in either of my presented experiments and that evidence for transfer is mixed, I decided to not test for transfer in this set of studies.

Another limitation is that I deliberately picked topics for which my participants were likely to have little or no background topic knowledge (e.g., acupuncture). This may not be fully realistic because people often come into a given learning task with at least some baseline of knowledge. For instance, when taking a class on genetics, the majority of learners will already have some prior knowledge about biology (e.g., rudimentary knowledge about DNA replication or protein synthesis based on introductory biology coursework that is typically required before taking higher-level coursework).

In the present experiments, I experimentally manipulated access to background topic knowledge of a given topic. A clear advantage of this design is that random assignment allowed me to make causal inferences about the effect of background knowledge. On the other hand, the level of background topic knowledge that could be created in these laboratory manipulations is small relative to some people's out-of-laboratory expertise on a particular topic. In my ongoing work, I am examining whether the testing effect is moderated by pre-existing expertise in topics some people tend to know a lot about while others less so (e.g., sports, Marvel superheroes, and Harry Potter). Indeed, prior research has found that undergraduate students are knowledgeable about many of these topics (Bruett, Fang, Kamaraj, Haley, & Coutanche, 2018; Troyer, 2017). This study is more ecologically valid compared to Experiments 1 and 2 because it explores pre-existing and untrained background topic expertise.

8.6 Educational Relevance

By better understanding the effect of background materials on the testing effect, educators can enhance students' educational success and memory retention of necessary materials. Educational activities usually need to reach a large range of students with varying levels of expertise, so it is crucial to determine whether retrieval practice is helpful for everyone or just for some. Thus, it is critical to explore possible moderators of the testing effect that may impact its effectiveness, like background topic knowledge. Given my findings, it seems that background knowledge does lead to better learning of information regardless of study strategy used. It is possible that participants in these studies were able to draw on background material and draw connections with the information in the main text, leading to better comprehension and retention. Perhaps additional readings should be encouraged for students who have less prior expertise on a given topic.

Further, I found that providing feedback following retrieval practice questions is particularly helpful. Consequently, it could be beneficial to also have students take intermittent smaller practice quizzes to test their knowledge throughout a term or school year, but to always provide feedback on their responses.

8.7 Conclusion

Across two experiments, I found no evidence that an experimental manipulation of background topic knowledge moderates the testing effect when applied to reading expository science texts. Further, this effect cannot be attributed to the irrelevancy of the background topic knowledge because simply having more background information on a given topic *did* lead to better learning and memory retention overall. In conclusion, the testing effect is robust across even meaningful differences in background knowledge – and that may be one reason why testing is a powerful effect in learning.

9.0 Appendix A: Session One Main Passages

9.1 Appendix A. 1. Dinosaur Main Passages.

9.1.1 Appendix A. 1. 1. Subtopic 1: We Dinosaurs Dumb?: Changes in Beliefs (Easy)

The discovery of dinosaurs in the 1800s provided, or so it appeared, proof that big bodies meant less smarts. With their pea brains and giant bodies, dinosaurs became a symbol of stupidity. Their extinction seemed only to confirm their bad design. For some time after their discovery, dinosaurs were thought of as slow and clumsy. For example, a typical image from the past involves a giant dinosaur called a Brontosaurus wading in a murky pond because he cannot hold up his own weight on land.

Modern scientists do not agree with this image. Modern imaginings of dinosaur bodies show them as strong, fast, and agile. Today, most paleontologists view dinosaurs as active and capable animals. Now, the Brontosaurus is imagined running on land. They even believe pairs of males could wrap their long necks around each other in combat (much like the neck wrestling of giraffes).

But the best example of dinosaur ability may well be the fact most often used against them—their extinction. What's remarkable about dinosaurs is not that they became extinct, but that they lasted on Earth for so long. Dinosaurs dominated the Earth for 160 million years before they become extinct. Meanwhile, humans have only been around ~300,000 years.

9.1.2 Appendix A. 1. 2. Subtopic 2: Were Dinosaurs Dumb?: Dinosaur Behavior (Medium)

Beyond brain size, behavior is another way of determining dinosaur intelligence. If dinosaurs were intellectually capable, we should find evidence of behavior that demands social and mental coordination. Indeed, we do. Multiple trackways have been uncovered, with evidence for more than twenty animals and multiple species traveling together. At the Davenport Ranch sauropod trackway, small footprints lie in the center and larger ones at the edge. As a further indication of herd life, upwards of thirty juveniles have been found next to a single adult dinosaur as well. That is among plant-eaters, but similar signs are present among meat-eaters, too. A group of Velociraptors were found in a quicksand pit next to an Iguanodon.

Further, few reptiles today are involved in the lives of their young, although the crocodile and pythons are notable exceptions. A finding of dinosaur bones next to unhatched eggs was once believed to be evidence of one dinosaur eating another's eggs. Multiple similar findings have changed that belief. Care for young may have been particularly important among dinosaurs. The Tyrannosaurus Rex hatched from an egg the size of a pigeon. It would take some time to grow from that small size to the 40-foot-long and eight-ton beast of the adult.

9.1.3 Appendix A. 1. 3. Subtopic 3: Were Dinosaurs Dumb?: Dinosaur Brains (Hard)

Beliefs about dinosaur intelligence have changed over the years primarily because of a change in understanding of stupidity and its correlation with size. Brain mass (weight) relative to body size, known as the encephalization quotient (EQ), correlates with intelligence. A higher ratio generally means greater intelligence. To put this in perspective, the Brontosaurus is 12x larger than a human, but its skull is 4x larger.

Part of the discrepancy may be that there wasn't evolutionary growth in brain size among dinosaurs over time, unlike in mammalian and bird groups. Instead, dinosaurs evolved a "second brain" (a bundle of neurons in their tails) to help speed up processing. Also, only the correlation of brain size with body size among similar animals (all reptiles, all mammals, for example) is reliable because brain size increases less than body size and at different rates among different types of animals, so when calculating EQ, an adjustment for brain to body growth rates has to be made based on animal type. Further, we must conclude that large animals require relatively less brain to do as well as smaller animals. The current view doesn't claim that dinosaurs are highly intelligent, only that they had the right brains for their bodies.

9.2 Appendix A. 2. Comet Main Passages

9.2.1 Appendix A. 2. 1. Subtopic 1: A Comment on Comets: Comet Orbits (Easy)

A comet is a small chunk of dust and ice that orbits (travels around) the Sun in an irregular but mostly oval shape. It is sometimes described as a "dirty snowball." The main part of a comet is called the nucleus. The nucleus is usually a few miles wide and has many holes in the surface which give it a spongy appearance, but it is not actually soft.

The most famous comet is called Halley's Comet. It can be seen from Earth without a telescope about every 76 years. Comets come from two areas at the farthest edges of the solar system. These areas are called the Kuiper Belt and Oort Cloud. Comets that can only be seen from Earth every several hundred years are from the Oort Cloud. Any comet that passes by Earth more frequently comes from the Kuiper Belt.

There are billions of comets in the solar system. Most never come close to the Earth. The comets that are seen from Earth have been pushed out of their normal orbits by the gravity of passing stars from other solar systems. The change in orbit can put a comet on a path closer to the Sun and Earth.

9.2.2 Appendix A. 2. 2. Subtopic 2: A Comment on Comets: Comet Research (Medium)

After the explosion that created our solar system about four billion years ago, some of the materials that were pushed farthest from the Sun froze together. Comets are believed to be made-up of these materials. Because comets spend most of their time in the outer reaches of space, they have remained relatively unchanged and are thought of as a “fossil record” of the solar system. Comets may even carry the secret to life. Water and some organic materials may have been brought to Earth by comets hitting our planet during its earliest days.

Scientists are unlocking these answers by studying comets directly rather than through a telescope. For example, a collection of tiny dust particles left behind by a comet led to the discovery of a previously unknown mineral. Even more recently, the Rosetta probe caught up with a comet beyond the asteroid belt after a ten-year flight. It sent back data from water vapor surrounding the comet that was fundamentally different from water on Earth. The probe also found organic compounds that could be the building blocks for DNA. Unfortunately, the solar battery died two days after landing in a crater, and no additional data was collected.

9.2.3 Appendix A. 2. 3. Subtopic 3: A Comment on Comets: Comet Light (Hard)

Most of the time, a comet only has a dark nucleus. The bright portions, called the coma and tail, are temporary and depend on the distance from the Sun and Earth. The Sun's heat causes frozen material to evaporate, and the resulting cloud formation around the nucleus is called the coma and can be larger than Earth. As the comet moves towards its closest point to the Sun, the perihelion, the momentum of solar photons creates radiation pressure as it meets dust in coma. The speed of each dust particle as it meets the radiation pressure varies according to its size which creates a tail of dust. Gas particles break away because the magnetic field of plasma of the outward bound solar winds attracts magnetized ions in the gas.

If a comet reaches its nearest point to Earth after its perihelion, it will be much brighter than if it reaches its nearest point to Earth while it is still relatively cold. However, the tails, which are sometimes longer than the Earth's distance to the Sun, and coma last only while the comet is fairly close to the Sun. After each pass, the nucleus of the comet is smaller and will eventually evaporate.

9.3 Appendix A. 3. The Great Barrier Reef Main Passages

9.3.1 Appendix A. 3. 1. Subtopic 1: The Great Barrier Reef: Humpback Whales

Humpback Whales are one of the largest species of baleen whale, averaging a length of 1 to 16 metres and weighing 25 to 30 metric tons. Named for the small hump on their dorsal fin, humpback whales are known for their complex breaching behaviours that make them popular with whale watchers.

Humpback Whales are intelligent, with brains weighing up to a ton. They can travel over 25,000km and remember locations, can recognize hunting group techniques, and have been known to play with and protect other species such as dolphins, seals, smaller whales, and humans. There is some scientific evidence to suggest that these whales may be able to recognize and remember certain humans and other species.

Humpback Whales have elaborate courtship rituals, which take place in the winter. Males form competing groups around females they want to mate with. They compete by song, breaching, charging, sparring, tail- and fin-slapping, and elaborately moving. As males are unsuccessful, they leave the group. Female humpback whales breed every 2 to 3 years and have more than one male partner over their lifetimes. Humpback whales have hybridized with other whale species, such as blue whales.

Humpback Whales exhibit a range of behaviours. Singing, breaching and jumping, slapping their tails, congregating, group hunting, and playing with other animals are just some of the playful activities these unique creatures engage in.

9.3.2 Appendix A. 3. 2. Subtopic 2: The Great Barrier Reef: Clownfish and Sea Anemones

The Great Barrier Reef is home to a number of species that have special, interdependent relationships. One such example is the unique, mutually beneficial partnership that exists between the clownfish and the sea anemone. The clownfish and sea anemone each benefit the other. In science, this type of relationship is called symbiotic and mutualistic.

Clownfish are small fish, typically about three to seven inches long. The name comes from their bright coloring, which can be orange, red, or yellow, interspersed with stripes of black and white. Clownfish use sea anemones as their homes. The sea anemone is a polyp, a cousin to the jellyfish. Sea anemones have long tentacles and look like exotic underwater flowers. But the sea anemone has a hidden power-its tentacles have venom that paralyzes fish and crabs. Once a fish or crab is paralyzed, the sea anemone eats it.

How does the clownfish survive living in such a dangerous home? The body of the clownfish is covered in a particular type of mucus. This mucus protects the clownfish from the anemone, making it immune to the poison. Because clownfish live in anemones, the poison tentacles protect them from other predators. The clownfish is also able to eat some of the food the anemone can't digest. The sea anemone benefits from having clownfish live inside it, as well. The clownfish plays a crucial role defending the sea anemone from fish and parasites that might otherwise harm it.

9.3.3 Appendix A. 3. 3. Subtopic 3: The Great Barrier Reef: Threats

The Great Barrier Reef, home to so many diverse species, is now in danger due to several threats. These threats include pollution, human interference, and changing ocean temperatures. Pollution and declining water quality endanger both the coral reef and the species that live within it. Rivers coming from northern Australia can bring pollution from farm runoff when there are floods. Farm runoff pollution includes animal waste, fertilizer, and pesticides. In recent years, pollution from these rivers has become worse because there are fewer coastal wetlands. In the past, coastal wetlands between the rivers and the Great Barrier Reef would serve as filters, keeping the worst of the pollution from reaching the ocean.

Human interference that harms the Great Barrier Reef includes shipping accidents and overfishing. Many ships pass through the Great Barrier Reef when they are bringing cargo to and from Australia. It can be tricky for captains to navigate through these waters, and, as of 2013, there were over 1,600 known shipwrecks in the Great Barrier Reef. Shipwrecks not only damage the physical structure of the reef; they can also spill oil into the water, killing local species.

Though pollution and human interference are both problems, many scientists consider climate change the greatest threat to the Great Barrier Reef. Ocean temperatures are rising, making coral reefs weaker and more susceptible to disease. Rising ocean temperatures also affect the ecosystems in the coral reef, throwing off the delicate balance that allows so many species to coexist. The Great Barrier Reef is one of the planet's treasure troves of biodiversity-but it may disappear within 100 years.

9.4 Appendix A. 4. Acupuncture Main Passages

9.4.1 Appendix A. 4. 1. Subtopic 1: Needles and Nerves: Acupuncture Origins (Easy)

Acupuncture has been practiced in China for over 2,000 years. It is based on the belief that the body contains energy called Qi (pronounced "chee"). Qi is energy that flows through the body on pathways called meridians. When you're healthy, the energy flows freely, but during illness, the energy may be weak or blocked. The goal of acupuncture is to improve the energy flow. In fact, those receiving acupuncture sometimes report feeling a small, slightly painful pinch, followed by a tug in the body. They believe the tug is related to the movement of energy.

According to acupuncturists, the flow of Qi through the meridians is greater in certain areas—these are the acupuncture points. Over 1,500 acupoints have been found. However, most of the points have no obvious relationship to the parts of the body they are intended to treat. For example, a point on the second toe is used to treat headaches.

Acupuncture is also used to promote general health. Those who practice acupuncture believe it can keep Qi in balance. By keeping Qi in balance, they believe it can stop the body from getting sick. Acupuncture is popular in the United States, but the explanation for how acupuncture actually works has long been a mystery for most Western doctors.

9.4.2 Appendix A. 4. 2. Subtopic 2: Needles and Nerves: Acupuncture and Vision (Medium)

Acupuncture is the practice of inserting tiny, hair-thin needles into the skin at specific points to treat pain and illness. Doctors and acupuncturists give millions of treatments each year in the U.S., usually for pain control. But studies show that acupuncture is also extremely useful for the type of nausea caused by chemotherapy and pregnancy. It can even reverse effects of eye degeneration which typically cannot be helped by Western medicine. Acupuncturists believe eye degeneration is caused by problems with Qi flowing through the spleen, liver, and kidney. However, the area they apply the needles to treat the problem is in the outside of the foot.

To understand if a point in the foot could affect the eyes, physicist Zang-Hee Cho strapped volunteers into an fMRI (functional magnetic resonance imaging) machine to get a photograph of their brain activity. Cho flashed a light in front of the volunteers' eyes so the fMRI image would show him what regions of their brain were involved in vision. Then, Cho had an acupuncturist stimulate the side of the foot. The very same areas of the brain lit up on the fMRI. To remove the possibility of a placebo effect, Cho also stimulated a nonacupoint in the big toe. This time, there was no response in the areas of the brain related to vision.

9.4.3 Appendix A. 4. 3. Subtopic 3: Needles and Nerves: Acupuncture and Pain Management (Hard)

Although a medical reason for all of acupuncture's benefits has not been found, scientists agree about how it reduces pain. The points at which acupuncture needles are inserted are likely the spots where nerves are gathered together. According to neuroscientist Bruce Pomeranz, many studies have shown that acupuncture stimulates nerves in the muscles. Researchers believe the stimulated nerves send signals up the spinal cord to the pituitary gland which produces and stores chemicals called endorphins. With a strong enough signal, the pituitary gland will begin releasing endorphins.

Endorphins are a well-understood chemical primarily involved in blocking pain signals from reaching the brain. Pain is a chemical message which travels from the source of a nerve through multiple cells on its way to the brain. Endorphins bind to opiate receptors which triggers the release of additional chemicals that block the reception of chemicals created by distressed nerves. Because of this, endorphins also trigger a positive feeling throughout the body and are responsible for the feeling of a "runner's high." However, unlike a runner's high, the brain keeps releasing endorphins up to 24 hours after acupuncture. This can improve blood flow, reduce inflammation, and allow the body to heal more rapidly.

10.0 Appendix B: Session One Background Passages

10.1 Appendix B. 1. Dinosaur Background Passages.

10.1.1 Appendix B. 1. 1. Subtopic 1: We Dinosaurs Dumb?: Changes in Beliefs (Easy)

Dinosaurs were the main animals on Earth for more than 160 million years. Some of them were the largest animals that ever walked on land. The last dinosaurs went extinct, or died out, about 65 million years ago. Dinosaur bones were around for a long time before people knew what they were. In fact, people thought one dinosaur bone belonged to a giant human. The term dinosaur was used for the first time in 1842 and literally translates to "terrible lizard." The study of dinosaurs is less than 200 years old, and early beliefs about dinosaurs have now been proven wrong.

One of the biggest mistakes scientists made was assuming that dinosaurs went extinct because they were too dumb and too big to survive. In fact, scientists thought that the Brontosaurus, a dinosaur with a really long neck, could not have held up its own neck on land. As a result, they drew pictures of the animal in water and believed that the water must have helped support the animal's neck. In reality, the Brontosaurus had extra tissue to connect the muscles in the neck to the bone and could support its neck on land. They may have used their necks similar to a giraffe, but scientists are not certain.

We also now know that dinosaurs did not go extinct because they were too dumb or too big. This should have been obvious to scientists early on because we knew dinosaurs as a group survived for over 165 million years. That is far longer than most groups of animals survive. For example, humans belong to a group of animals called primates, which include apes and monkeys.

Primates have only existed for 50 million years. That's far less than dinosaurs. There is no doubt that humans and other primates are among the most intelligent animals to ever live. There's also no doubt that primates, and many other mammals, are smarter than dinosaurs. But dinosaurs must have been doing something right to have lived for so long.

Dinosaurs went extinct because a large asteroid hit the Earth. The asteroid instantly killed off a lot of animals and plants. With so few animals left, there was not enough food for dinosaurs to eat, and they quickly went extinct. There was nothing dinosaurs could have done to stop it. Some scientists even believe that if an asteroid had not hit the Earth, dinosaurs would still be alive today.

10.1.2 Appendix B. 1. 2. Subtopic 2: Were Dinosaurs Dumb?: Dinosaur Behavior (Medium)

In 1923 scientists discovered eggs believed to belong to a dinosaur called a Protoceratops. Next to the eggs were the bones of another dinosaur, an Oviraptor, that they believed died while trying to eat the eggs. Years later, they discovered more of these same eggs, but inside one was the body of an Oviraptor. This meant that the Oviraptor was the parent and was not eating the eggs.

The new discovery shouldn't have been surprising. Exposed eggs require an adult to keep them warm. Dinosaurs at a minimum would have had to stay with their eggs until they hatched. Supporting this, the eggs were arranged in a wide circle so a large dinosaur could keep them warm without crushing them. But the evidence for dinosaur parenting does not stop there. Young dinosaurs, but not babies, have been found in nests next to left-over food presumably brought by adults. The reason dinosaurs cared for their young was likely related to their size. Bigger eggs need thicker shells for support, but if the shell is too thick, oxygen can't get in. As a result, dinosaurs had to be born proportionally much smaller than many other species and would need protection to reach their adult size.

Dinosaurs also had a social network outside their immediate family. Trackways, which are just dinosaur footprints, show that they sometimes traveled in herds. But even more important are the Davenport Ranch Trackways which show a small herd of two adults and twenty-one juveniles of the same species traveling together. This would have required attention to the position of the other adults at a minimum. Further evidence that herd behavior was important to dinosaurs comes from a nest containing more than 30 eggs. Dinosaurs did not produce this many eggs at once, so the number suggests multiple mothers were taking turns keeping them warm.

There is also some evidence of social coordination among meat-eating dinosaurs. In Utah, researchers found several Velociraptors and an Iguanodon trapped together in quicksand. It is possible that the Velociraptors were each trapped separately, but a trackway in China shows an Iguanodon running from multiple Velociraptors supporting the pack-hunting theory.

Altogether, the evidence suggests that despite their small brains, dinosaurs could engage in complex social coordination. This does not mean they were as smart as primates or even dogs, but it does mean they weren't dumb and were likely smarter on average than modern reptiles.

10.1.3 Appendix B. 1. 3. Subtopic 3: Were Dinosaurs Dumb: Dinosaur Brains (Hard)

In the 1800s, scientists started to notice that the size of an animal's brain in proportion to its body size correlated with its intelligence. At the same time, scientists realized that dinosaur skulls were small relative to their body. The theory went that the larger the brain is relative to the body, the more brain mass available for complex cognitive tasks. The fact that an animal weighing over 5 tons could have a brain of no more than 2.8 oz led to the idea that dinosaurs were unintelligent.

The ratio between brain size and body mass is a generally reliable predictor of intelligence, as long as you know how to apply it. The encephalization quotient (EQ) formula varies but is usually Ew(brain) = 0.12w(body)2/3. From this, we get a mean EQ for mammals around 1, with meat-eaters, marine mammals, and primates above 1, and insects and planteaters below. The power sign corrects for the brain growing at 2/3 the rate of the body in mammals. For example, mice have a brain/body size ratio similar to humans (1:40), while elephants have comparatively small brain/body size (1:560), though elephants are obviously intelligent. What is likely happening is that the brain can only be so small and still function. In some ways the cost for each additional neuron in terms of overall brain volume gets smaller as the number of neurons go up. The end result is that larger animals don't need as much brain mass to support greater intelligence and as a result their brain to body ratios reduce.

The formula for EQ is based on data from mammals, so it should be applied to other animals with caution. Reptile brain to body growth is less understood, but a power of ¼ might be more appropriate. Differences in brain to body growth rates may reflect differences in evolutionary selection pressures. Only in mammals and birds has evolution favored large brains. In mammals and birds, brain size relative to body size increased over the course of evolution. In dinosaurs, evolution made the brain more efficient, though not more intelligent, by adding additional neurons to the base of the spine which could speed up processing related to movement and reflexes. This was highly adaptive but didn't require more intelligence and didn't require a larger brain. Ultimately, dinosaurs had small brains but given corrections for EQ this does not mean they had to be unintelligent, at least compared to other reptiles.

10.2 Appendix B. 2. Comet Main Passages

10.2.1 Appendix B. 2. 1. Subtopic 1: A Comment on Comets: Comet Orbits (Easy)

Our Sun was formed 4.5 billion years ago through an explosion. The material from the explosion included gas, water, and dust. The explosion kicked many of these materials out far away from the Sun. In fact, they ended up so far away from the Sun that they froze. When some of the gas and dust froze together, it created comets. That means that comets are made of frozen gas and dust. The frozen ball of gas and dust is called a nucleus. You may have heard of a nucleus in biology or physics. They form the center of cells and atoms. Just like with cells and atoms, the nucleus of a comet forms the center of a comet.

The combination of frozen gas and dust gives comets an unusual appearance. They look like a sponge because there are many holes in the comet. Of course, the frozen material that makes up a comet does not feel like a sponge. It is not soft. Instead, the material is hard, like a rock. In fact, it's hard enough that probes sent by NASA have been able to successfully land on comets.

Comets orbit the Sun. An orbit is just an object's path as it moves around the Sun. The Earth orbits the Sun too. Because comets are so far away, it takes them a long time to orbit around the Sun, and the farther away they are, the longer it takes. Objects that are close to the Sun, like Earth and Mars, orbit the Sun more quickly than objects in the Kuiper Belt and Oort Cloud. And objects in the Kuiper Belt are closer to the Sun than objects in the Oort Cloud. Comets in the Oort Cloud and Kuiper Belt are so far way that we cannot see them from Earth, even with a high-powered telescope. In order for comets to be visible from Earth, something has to force them out of their normal orbit. And then the comet's new orbit needs to bring it close to the Earth. Only then, will we have a chance of seeing the comet.

The chance of a comet's orbit bringing it close to the Earth is very small. But, when it does happen, we can sometimes see them with our naked-eye. That means that a telescope is not needed to see near-Earth comets as long as the comet is lit up.

10.2.2 Appendix B. 2. 2. Subtopic 2: A Comment on Comets: Comet Research (Medium)

Comets were created during an explosion that created the Sun and marked the beginning of our solar system. Because comets spend most, if not all, of their time in the outer solar system, far from the Sun, they are frozen, and, because frozen material does not change, comets have not changed much since the beginning of the solar system. As a result, learning about comets means learning about the origins of the solar system and everything within it.

One question scientists are trying to answer is how water got to Earth. When the Earth was forming, it was so hot that most of its water evaporated. Once the Earth cooled down, there was virtually no water left, meaning that the water that makes up the oceans had to have come from somewhere else. One theory is that the water came from the frozen ice on comets that hit the Earth.

All water contains H2O: two parts hydrogen and one part oxygen. But hydrogen comes in two types: regular hydrogen and deuterium. Deuterium is just hydrogen with an added neutron. Earth's water contains far more regular hydrogen than deuterium. If Earth's water came from comets, then the water in comets should also contain more regular hydrogen.

Answering this question was one of the reasons scientists launched the Rosetta probe to land on a comet, analyze its materials, and send back the results. Rosetta's measurements revealed far more deuterium in the water on the comet than exists in Earth's water. This makes it highly unlikely that Earth's water came from a comet.

Nevertheless, the Rosetta probe found other materials on the comet which are critical to life. The most important was glycine, a building block of DNA. Unfortunately, the Rosetta probe's solar died faster than expected. When the probe landed on the comet, it unexpectedly bounced and ended up in the shadow of cliff. It was out of view of the Sun, and its batteries could not recharge. The probe fell silent when its solar batteries ran out of power.

Thankfully, we don't have to wait for the next probe to study comets. By plotting the orbit of a comet, NASA can pinpoint the date when its dust will enter the Earth's atmosphere. During one such occasion, NASA collected dust in the stratosphere and found a brand new type of mineral. It was a type of manganese silicide which has been named "Brownleeite" after the researcher who found it.

10.2.3 Appendix B. 2. 3. Subtopic 3: A Comment on Comets: Comet Light (Hard)

In the outer Solar System, comets remain frozen and inactive and are extremely difficult or impossible to see from Earth due to their small size. Statistical detections of inactive comet nuclei in the Kuiper belt have been reported from observations by the Hubble Space Telescope but these detections have been questioned. As a comet approaches the inner Solar System, solar radiation causes the volatile materials within the comet to vaporize and stream out of the nucleus, carrying dust away with them. Some of the dust is left behind as the ice changes. It forms a dark, protective crust on the surface of the nucleus and slows the melting. In some places the protective layer is thinner, and jets of gas break through. The gas and dust form a cloud around the nucleus called a coma.

Two distinct tails develop from the coma — the plasma (gas) tail and the dust tail, and each form their own distinct tail, pointing in slightly different directions. The different shapes and angles of the tails are caused by the way different particles are affected by the Sun. The thinner, longer plasma tail forms a straight line extending from the comet. The particles in this ion tail are electrically charged and are pushed away from the Sun by solar wind. The solar wind is made-up of a constant flow of gas and particles (mostly protons and electrons) that stream outward at 220 miles per second.

The shorter dust tail is curved slightly. The larger particles in the dust tail do not have an electric charge and are not affected by the solar wind. Dust-size particles that escape from the comet experience a much weaker push from the Sun caused by the pressure of sunlight itself (called radiation pressure), rather than by the charged particles of the solar wind. Radiation pressure is the mechanical pressure exerted upon any surface due to the exchange of momentum between the object and the electromagnetic field. While the dust tail also points generally away from the Sun, it has a slight curve back in the direction the comet came from.

Comet tails get longer and more impressive as the comet gets closer to our Sun. As the comet approaches our Sun, it gets hotter and material is released more rapidly, producing a larger tail. Scientists estimate that a comet loses between 0.1 and 1 percent of its mass each time it orbits our Sun.

10.3 Appendix B. 3. The Great Barrier Reef Main Passages

10.3.1 Appendix B. 3. 1. Subtopic 1: The Great Barrier Reef: Humpback Whales

While the Great Barrier Reef is the permanent home for many animals and plants, other species only visit the area seasonally. The humpback whale comes to the Great Barrier Reef every winter to breed and give birth to its young.

Though humpback whales look similar to fish and share many characteristics, they are, in fact, mammals. Instead of scales they are covered in skin. The markings on a humpback whale's skin are unique to each whale, similar to how every human being has a fingerprint unlike any other. Humpback whales are one of the largest animals in the Great Barrier Reef, about as long as a medium school bus. On average, the humpback whale comes to the ocean's surface to breathe every seven to 15 minutes, but they can remain underwater for as long as 45 minutes. Humpback whales are famous for their singing. Male humpback whales vocalize, making noises that last up to 20 minutes and sound eerily similar to songs.

Even though humpbacks are enormous, they only eat the tiniest of sea creatures. Favorite foods of the humpback whale include plankton, shrimp-like creatures called krill, and small fish such as herring and mackerel. Humpback whales don't have sharp teeth. Instead, their mouths are filled with large plates of baleen. Baleen is made out of keratin, the same material that our fingernails are made from. It enables the whales to strain the small creatures from the seawater. To feed, the humpback whale will gulp a mouthful of small fish, plankton, or krill and then let the water flood out.

Humpback whales use a hunting strategy called bubble net feeding. A group of whales work together to capture large schools of herring, krill, or other small sea creatures. One whale blows a wall of bubbles around the herring school, while other whales make noises. These stimuli confuse the whales' prey so that the rest of the whales can herd the small creatures together and upwards. Then the whales can easily lunge up with their mouths open and consume large quantities of sea creatures. Humpback whales eat an average of 4,500 to 5,500 pounds of plankton, krill, and fish each day during their feeding season.

The Great Barrier Reef is crucial for the humpback whales' survival. Humpback whales come from Antarctic waters to the Great Barrier Reef from May to September to give birth and build up strength over the winter before they return to the Antarctic in the summer, according to the Great Barrier Reef Marine Park Authority.

10.3.2 Appendix B. 3. 2. Subtopic 2: The Great Barrier Reef: Clownfish and Sea Anemones

The symbiotic relationship between an anemone (Heteractis magnifica) and the clownfish (Amhiron ocellaris) is a classic example of two organisms benefiting each other; the anemone provides the clownfish with protection and shelter, while the clownfish provides the anemone nutrients in the form of waste while also scaring off potential predator fish.

In ocean reefs, corals share a symbiotic relationship with fish; the corals provide fish with shelter and protection from predators, and the fish help the corals thrive by eating plants such as seaweed, which can kill corals by taking up space and light that they need to survive. Clownfish eat various small invertebrates and algae, as well as food scraps the anemone leaves behind. The clownfish also fertilizes the anemone with its feces.

Let's consider a natural ecosystem such as the ocean. Oceanic environments are known for their species diversity. Imagine you are on a diving expedition to explore the worlds beneath the waves. If we were in the warm waters of the Pacific or Indian Oceans, we'd likely spot an excellent example of mutualism: the relationship between clownfish and sea anemones. In a mutualistic relationship, both species benefit. Sea anemones live attached to the surface of coral reefs. They trap their prey with stinging cells called nematocysts, which are located on their tentacles. Nematocysts release toxins when a small animal contacts an anemone's tentacle. This paralyzes the stung animal, allowing the anemone to easily bring the animal into its mouth for ingestion. While other fish succumb to these toxic stings, clownfish secrete a substance in the mucus covering their bodies that suppresses the firing of nematocysts. This allows clownfish to swim comfortably between the tentacles of anemones, creating a protected environment in which potential predators are killed off by anemone stings. This clearly benefits the clownfish, but how about the sea anemones? The brightly-colored clownfish attract other fish looking for a meal. These unsuspecting would-be predators are then caught and eaten by the anemones.

Clownfish eggs can be hatched anytime during the year. Male clownfish are the primary caretakers for their young, with females only helping on occasion. After the clownfish eggs are laid, the male clownfish guard them until they hatch. Clownfish are all born males. Once they become female, they cannot be male again.

10.3.3 Appendix B. 3. 3. Subtopic 3: The Great Barrier Reef: Threats

Temperatures are rising on our land and in our oceans, caused primarily by an increase of carbon dioxide (CO2) and other greenhouse gases. CO2 levels have been rising steadily for more than 100 years due mainly to the burning of fossil fuels, trapping more heat in our atmosphere and contributing to climate change. The Great Barrier Reef is one of the richest and most complex natural ecosystems in the world, but climate change is the biggest threat to the future of coral reefs around the world. It impacts our Reef in a number of ways.

When corals suffer heat stress, they expel the microscopic algae that live inside their tissues, revealing their white skeletons. Bleached corals are not dead, but are more at risk of starvation and disease. Already marine heatwaves have triggered three mass coral bleaching events on the Great Barrier Reef in just five years, reducing shallow water coral reefs by as much as 50%. Coral reefs can recover from bleaching over time, but only if temperatures drop and conditions return to normal.

The ocean absorbs carbon dioxide (CO2) from the atmosphere, making it more acidic. This process is known as ocean acidification. Since the late 18th century, the ocean has absorbed about 30% of the CO2 humans have generated, decreasing its pH level. A more acidic ocean means corals are less able to build skeletons and form coral reefs, which help protect coastlines from storms and provide habitats for thousands of species of marine life. Climate change is increasing the frequency and intensity of severe weather events. Coastal regions like the Great Barrier Reef are particularly exposed to damaging cyclones, flooding_ and storms. Between 2004 and 2018, 10 cyclones of category three or more crossed the Great Barrier Reef, causing significant damage to coral reefs.

As water temperatures rise, many marine species are being forced to move south to cooler habitats. This shift creates increased competition for food and shelter in cooler waters, threatening the entire ecosystem. For Reef communities, the loss of marine life can have a devastating impact on local ecosystems, food sources and other industries such as tourism. To protect the Reef from climate change we must reduce emissions and help coral reefs adapt. Insufficient global action on climate change is taking a serious toll on the health of our Great Barrier Reef. Urgent global action to drastically reduce greenhouse gas emissions is needed now if we are going to have any chance of saving coral reefs.

10.4 Appendix B. 4. Acupuncture Main Passages

10.4.1 Appendix B. 4. 1. Subtopic 1: Needles and Nerves: Acupuncture Origins (Easy)

Acupuncture is an ancient Chinese form of healing. It involves a patient, the person receiving the acupuncture, and an acupuncturist, the person giving the acupuncture. The patient lies on a table, and the acupuncturist sticks special needles into points on the body. The needles are made of metal and are about as thick as a human hair. They normally go less than 0.5 inches into the skin.

When a needle is pushed into the skin, the patient may feel a slight pinch or tug and then a tingling sensation that spreads out from where the needle pierced the skin. The pinch can be a little painful, but the tug is believed to be the feeling of Qi moving through the body. Qi is like breath. According to acupuncturists, all of the parts of the body are connected by lines called meridians. The meridian lines are like a giant web that links different parts of the body together. Every organ has its own meridians that connect a specific area of the body to the larger web. Qi moves from one area to the body to another by traveling along the meridians.

Sometimes these lines cross. These are the acupuncture points. When energy in the body flows easily, we don't feel Qi, and we are balanced and healthy. But when energy gets blocked at the acupoints, it causes pain and disease. Acupuncture gets Qi unstuck so that energy can flow through the body again. This helps the body heal and stay healthy.

Acupuncture requires exact placement of needles at spots on the body called acupoints. Placing the needles requires in-depth knowledge of the body. There are over a thousand possible points where the needles can be stuck, each with a different effect. One of the interesting things about acupuncture is that acupoints do not have a clear relationship to the parts of the body they affect. For example, putting a needle into an acupoint on the wrist does not help with wrist injuries. Instead, it helps with heart problems. It is not clear to modern doctors why this works.

Doctors today do not believe that acupuncture is related to Qi. But they do believe that acupuncture can help the body heal and reduce pain. They do not know how it works, but they do see that it works in their patients. As a result, it is not unusual for doctors to tell their patients to try acupuncture.

10.4.2 Appendix B. 4. 2. Subtopic 2: Needles and Nerves: Acupuncture and Vision (Medium)

Macular degeneration is the most common cause of severe vision loss in people over age 50. The disease causes a breakdown of cells in the central part of the eye, called the macula, and results in blurred vision. Eventually, it can cause a blind spot to form in the person’s central vision. According to acupuncturists, macular degeneration is caused when the body's yin is reduced in the kidney and liver. They believe that the spleen makes the yin during digestion. When the body's ability to turn food into energy decreases, the body’s ability to produce yin decreases. This could happen if Qi was blocked in the spleen. As a result, acupuncturists increase yin in the liver and kidneys by unblocking Qi in the spleen. This in turn reverses macular degeneration.

Meridians in the spleen, cross with meridians from other areas of the body in the foot. Acupuncturists therefore unblock Qi in the spleen using acupoints on the side of the foot. Do Western doctors believe this actually works? One possibility is that it works like a placebo effect. That just means that it works because people believe it will work. However, placebo effects do not usually last for very long. Researchers tested whether or not acupuncture was related to placebo effects by seeing if stimulating vision-related acupoints on the foot would produce activity in vision-related areas of the brain. They used fMRI, which shows if a specific area of the brain is being used when the picture is taken. The researchers used the acupoint on the side of the foot that is believed to be related to reversing macular degeneration and they also used a point on the big toe which is not believed to be related to vision. Acupuncture at the vision-related acupoint caused more increases in activity within the area of the brain responsible for vision than at the non-vision-related acupoint.

However, macular degeneration is related to changes in the eye, not the brain. The results of the fMRI study show that acupuncture can produce activity in seemingly unrelated brain regions, but it does not explain how this improves macular degeneration. The same is true for other disorders. Acupuncture treats nausea from pregnancy, also called morning sickness, but interestingly there is no evidence that it can treat nausea from the flu. Instead, it helps with the flu by boosting the body's immune system. Right now, the only explanations for how acupuncture works lie outside of Western Medicine.

10.4.3 Appendix B. 4. 3. Subtopic 3: Needles and Nerves: Acupuncture and Pain Management (Hard)

Acupuncture triggers the release of endorphins, which are "feel-good" chemicals that stop the brain from feeling pain. Endorphins are also released through exercise. In fact, both acupuncture and exercise cause the same series of events. The nerves within the muscles or skin receive a negative sensation, cells in the skin or muscles release a chemical called adenosine which travels to the hypothalamus (a portion of the brain which essentially routes incoming signals), and the hypothalamus then produces a separate chemical which is sent to the pituitary gland where endorphins are stored. When the pituitary gland receives the message from the hypothalamus, it releases the endorphins.

Endorphins reduce pain in a similar way to pain killers. In fact, pain killers are essentially man-made endorphins. They both bind to opiate cells involved in making the body feel good. When they bind to these cells, even more chemicals are released, some of which go back to the brain and reduce stress, and some of which compete to bind at the same receptors as the adenosine, effectively preventing it from doing its job. When adenosine or similar chemicals are not able to reach the brain as easily, pain is reduced. Of course, when these pain signals stop reaching the pituitary gland, it stops releasing endorphins and the "feel good" feelings go away.

Acupuncture works because, unlike with exercise, the adenosine is still present as endorphin levels reduce, thus prompting the pituitary gland to release endorphins over a longer timeframe. Exercise produces adenosine through temporary stress on the muscles and joints whereas acupuncture needles cause minor damage to tissue under the skin. In fact, a critical component of making acupuncture effective is twisting the needle, which increases the tissue damage. That means that adenosine will continue to be produced until the repair is complete, which can take 24 hours.

The longer period of increased endorphins following acupuncture has a number of longterm health benefits. In the short term, inflammation helps the body heal by increasing blood flow, but if it stays too high for too long it can actually cause more problems. Inflammation is caused by an increasing immune cells. At first, they help the body heal by attacking viruses and bacteria. But if they stay active too long, they will actually start attacking the body. Endorphins can attach themselves to immune cells and turn them off. This reduces inflammation and allows the body to finish healing and return to its normal state.

11.0 Appendix C. Session One Retrieval Practice Questions

a. Indicates the correct answer.

11.1 Appendix C. 1. Dinosaur Retrieval Practice Questions

1. How has the perception of dinosaurs changed?
   a. It was once believed that they were slow and not likely to survive, but we now believe they could move quickly and easily.
   b. It was once believed that they only existed for 300,000 years, but now we know they lasted 160 million years.
   c. It was once believed that dinosaurs could swim, but now we know they were too big to swim.
2. The author says, "Now, the *Brontosaurus* is imagined running on land " to make the point that:
   a. Scientists' understanding of the *Brontosaurus* has changed within the last generation.
   b. The *Brontosaurus* evolved from living in the water to living on land.
   c. The *Brontosaurus* eventually learned to hold up its weight on land.

3. Which of the following statements would the author most likely agree with regarding dinosaur extinction?
    a. They lasted far longer than most animals before going extinct.
    b. Their extinction proves their bad design.
    c. They were too big to survive in the ice age that came after the asteroid hit.
4. Which of the following is true about large animals?
    a. Larger animals typically have lower brain to body ratios than smaller animals.
    b. EQ should not be used to compare intelligence between small and large animals.
    c. Larger animals do not need as much intelligence as smaller animals to survive.
5. Which of the following best states the relationship of brain size to body size?
    a. The brain grows at two-thirds the rate of the body.
    b. Brain size is not related to body size.
    c. If an animal has a bigger body, they will have a smaller EQ.
6. Which of the following is a potential problem for judging dinosaur intelligence based on EQ?
    a. The ratio of brain size to brain mass works within animal types (e.g., mammals) but not across animal types because of variation in brain to body growth rates.
    b. The ratio of brain size to brain mass is less relevant among animal types which have not have not experienced an evolutionary increase in brain size over time.
    c. The ratio of brain size to brain mass is off among dinosaurs because they had "second brain" at the base of their spine that reduced the need for a large brain.

7. Why are the Davenport Ranch Sauropod Tracks evidence of dinosaur intelligence?
    a. They demonstrate coordinated efforts among the herd.
    b. They demonstrate that some dinosaurs lived in large herds.
    c. They demonstrate that dinosaurs had cross-species social organization.
8. What can be assumed about the relationship between parental care and dinosaurs?
    a. Dinosaurs looked after their young until the young were large enough to survive on their own.
    b. Dinosaurs looked after their young until the young had learned enough to survive on their own.
    c. Dinosaurs may have watched their eggs, but they likely did not "raise" their young.
9. Which of the following is implied about dinosaurs?
    a. Dinosaurs in herds may have cared for each other's young.
    b. Raising the large numbers of young produced by a single dinosaur would have taken a lot of mental coordination.
    c. Dinosaur parental behavior was unique among reptiles.
10. Which of the following is probably a current belief about the *Brontosaurus*?
    a. Its small brain to body ratio was likely related to its large body size.
    b. Its small brain meant that Brontosaurus herds probably did not have complex social behavior.
    c. Its size meant it did not need as much intelligence as smaller dinosaurs to survive.

11. Which of the following questions can be answered best by the passage?
    a. Why might dinosaurs have needed more intelligence than many modern reptiles?
    b. What social behaviors were characteristic of the *Brontosaurus*?
    c. Why did dinosaurs travel in herds when no modern reptiles do?
12. Which of the following best demonstrates the author's beliefs about dinosaurs?
    a. Interpretation of dinosaur fossils and behavior was influenced by false beliefs about brain size.
    b. Recent ideas about dinosaur behavior suggest that they were highly intelligent animals.
    c. New discoveries about dinosaurs have changed the way scientists view the relationship between brain and body size.

11.2 Appendix C. 2. Comet Retrieval Practice Questions

1. What is a comet made of?
    a. Dust and ice
    b. Cast off materials from when meteors hit planets
    c. Organic materials that do not exist on Earth
2. Which of the following best describes the nucleus of a comet?
    a. It's full of holes
    b. It's a dense block of solid rock
    c. It's soft, like a sponge

3. How likely is it that a comet pushed out of its orbit will come close enough to Earth to be seen by astronomers?
   a. Unlikely
   b. Likely
   c. A comet cannot be pushed out of its orbit
4. What might cause a comet near the Earth to be less visible to astronomers?
   a. If it approaches Earth before it has moved closest to the Sun
   b. If it approaches Earth after it has moved closest to the Sun
   c. If its tails have begun to separate as they are attracted by magnetic fields in the Sun
5. Which of the following are true about the tail of a comet?
   a. The tails are created by radiation pressure blowing dust off the coma and solar winds ionizing and attracting the gasses.
   b. The tail is created when material is, in essence, blown off the coma by solar winds and magnetized ions force separation of the dust and gas particles.
   c. The tails become smaller during each orbit around the sun as the nucleus loses more and more material.
6. What direction is a comet tail pointed as the comet travels around the Sun?
   a. Away from the Sun
   b. Away from the Sun as it approaches and toward the Sun as it departs
   c. The dust tail points away from the sun while the ionized gas tail is attracted towards the sun

7. Why did the Rosetta probe only collect data for 2 days after it landed on the comet?
    a. The probe happened to land in a crater, blocking it from view of the sun.
    b. The probe had used a lot of battery power to reach the comet and had little left after its arrival.
    c. The probe landed harder than expected, causing it's the battery to malfunction.
8. Why are comets considered a "fossil record"?
    a. The materials that make up the nucleus are unchanged since the origin of the solar system.
    b. During their orbits, they pick up material from many regions of the solar system and can document its evolution.
    c. Material from comets helped create many of the planets, so they hold the key to understanding planet origins.
9. Which of the following were NOT found by the Rosetta probe?
    a. A previously undiscovered mineral
    b. Organic compounds similar to parts of DNA
    c. Water vapor
10. Information in the passage indicates that seeing a near-Earth comet requires all of the following EXCEPT:
    a. the viewer to have a powerful telescope
    b. solar winds and radiation pressure to blow against the coma
    c. the comet to have been pushed out of its typical orbit

11. The passage mentions astronomers observing all of the following about comets EXCEPT:
    a. comets that give off bright light from their nucleus
    b. orbits that take comets to the edges of the Sun's gravitational influence
    c. comets with sponge-like appearances
12. Scientists are most interested in comets directly because:
    a. They contain chemicals from the origin of the solar system.
    b. The temporary effects of close encounters with solar radiation can reveal how planets and atmospheres were formed.
    c. They come from the edge of the solar system, and so hold clues as to what lies beyond.

11.3 Appendix C. 3. The Great Barrier Reef Retrieval Practice Questions

1. All the following are examples of the intelligence of humpback whales except:
    a. They can guide ships away from storms.
    b. They can coordinate hunting techniques in groups.
    c. They have the potential to remember specific humans.
2. Humpback whales have their elaborate courtship rituals during which season?
    a. Winter
    b. Fall
    c. Spring

3. Whale watchers are especially interested in seeing humpback whales for what reason?
   a. Their breaching behavior
   b. Their large size
   c. Their playful nature
4. You could identify clownfish by all of the following except:
   a. How many fins they have
   b. Intense colors
   c. Length
5. What is not true about sea anemone tentacles?
   a. Sea anemone tentacles are short and look like flowers.
   b. Sea anemone tentacles are long and look like flowers.
   c. Sea anemone tentacles are long and look like tree branches.
6. What is characteristic of the symbiotic relationship between clownfish and sea anemones?
   a. Sea anemones provide defenses to counteract the clownfishes' noticeable colors.
   b. Clownfish provide a constant source of nutrition for the sea anemone.
   c. Sea anemones protect clownfish from infectious parasites and diseases.
7. The author notes that as of 2013 there have been 1600 shipwrecks in the Great Barrier Reef to make a point that:
   a. It’s challenging for sailors to sail through.
   b. The fish in this area are extremely aggressive.
   c. The ships used aren’t made to sail through this part of the ocean.

8. Coastal wetlands are beneficial to the Great Barrier in what way?
    a. They filter out pollutants.
    b. They bring more nutrients to the species within Great Barrier Reef.
    c. They help maintain proper water temperatures.
9. A reduction of what could help sustain the biodiversity of the Great Barrier Reef Ecosystem?
    a. Usage of agricultural pollutants
    b. Snorkeling tours
    c. Invasive fish species
10. Which of the following animals does not live in the Great Barrier Reef?
    a. Tortoises
    b. Coral
    c. Mollusks
11. How might a biologist determine a symbiotic relationship if they were looking at a different marine ecosystem that wasn't the Great Barrier Reef?
    a. They could observe and note which species interact in a way that provides mutual benefits for both parties.
    b. One would need to ensure that two species provide equal and opposite benefits to each other to ensure their survival.
    c. They could look for species that live in close proximity of one another and if the two interact daily, it is safe to assume it is a symbiotic relationship.

12. The article supports all of the following points about the Great Barrier Reef except:
    a. Reducing human interference will be the most effective way to reduce threats to the Great Barrier Reef.
    b. Clownfish would not survive in the ocean without sea anemones.
    c. The Great Barrier Reef is composed of multiple ecosystems.

11.4 Appendix C. 4. Acupuncture Retrieval Practice Questions

1. What might you feel if you get acupuncture?
    a. A feeling as if energy is moving within the body
    b. Nothing. Acupuncture is painless.
    c. Instant relief from pain
2. How do Western doctors view acupuncture?
    a. As having potential benefits, although how it works is still unclear
    b. As a type of alternative medicine that is not based on science
    c. As a good example of the placebo effect
3. During illness, what can happen to the body's Qi?
    a. Qi gets blocked
    b. Qi gets drained
    c. Qi breaks down

4. Why might acupuncture therapy continue to reduce pain even weeks treatment?
    a. Endorphins can reduce inflammation and give the body time to heal.
    b. It causes the body to start consistently releasing more endorphins, which block pain signals from being sent to the brain.
    c. It stimulates endorphins in the muscles which promote relaxation and healing.
5. Why might multiple needles be needed during acupuncture?
    a. Because prolonged release of endorphins requires a buildup of signals from the body
    b. Because many areas of the body need to be pierced with a needle for treatment to work
    c. Because the area being treated is large and requires a greater release of endorphins
6. How might acupuncture and runner's highs be similar?
    a. They both stimulate the same areas of the brain.
    b. Western medicine is not able to explain their health benefits.
    c. They both provide a short-term rush of endorphins.
7. Why do acupuncturists believe acupuncture improves degenerative eye disease?
    a. It improves the functioning of the spleen.
    b. It releases endorphins, which reduce eye inflammation.
    c. It helps to unblock Qi within the eye.
8. Acupuncture may help with all of the following EXCEPT:
    a. Nausea from the flu
    b. Morning sickness
    c. Blurred vision

9. Why do acupuncturists use the acupoints on the outside of the foot to treat degenerative eye disease?
    a. They are connected via meridians to the spleen.
    b. They are connected via meridians to the eye.
    c. They are connected via meridians to areas of the brain involved in vision.
10. The passage suggests that acupuncture research:
    a. has found evidence that acupuncture reduces inflammation
    b. has demonstrated that acupuncture increases blood flow
    c. has <u>not</u> found evidence that acupoints on one area of the body are connected to other, seemingly unrelated, areas of the body
11. Which of the following best explains the author's perspective about how acupuncture works?
    a. There are multiple ways acupuncture may work and most of them are <u>not</u> well understood.
    b. Most of the health benefits from acupuncture are related to increased endorphin levels in the body.
    c. Each acupoint has the ability to stimulate a seemingly unrelated area of the brain which promotes healing of the affected area of the body.
12. The article supports all of the following points about acupuncture EXCEPT:
    a. Western medicine has been ignoring the benefits of acupuncture treatment for too long.
    b. Pressure placed at acupoints can cause activity in surprising areas of the brain.
    c. Acupuncture can help a lot of modern problems, including general pain and nausea.

12.0 Appendix D. Session One Restudy Sentence Facts

12.1 Appendix D. 1. Dinosaur Restudying Sentence Facts.

1. It was once believed that they were slow and not likely to survive, but **we now believe they could move quickly and easily**.
2. The author says, "Now, the *Brontosaurus* is imagined running on land " to make the point that **scientists' understanding of the *Brontosaurus* has changed within the last generation**.
3. The author would most likely agree that **dinosaurs lasted far longer than most animals before going extinct**.
4. **Larger animals typically have lower brain to body ratios** than smaller animals.
5. The brain grows at **two-thirds the rate of the body**.
6. A potential problem for judging dinosaur intelligence based on EQ is that **the ratio of brain size to brain mass works within animal types (e.g., mammals) but not across animal types because of variation in brain to body growth rates**.
7. The Davenport Ranch Sauropod Tracks evidence of dinosaur intelligence because **they demonstrate coordinated efforts among the herd**.
8. We can assume that dinosaurs looked after their young until **the young were large enough to survive on their own**.
9. It is implied that **dinosaurs in herds may have cared for each other’s young**.
10. A possible current belief about the *Brontosaurus* is that **its small brain to body ratio was likely related to its large body size**.

11. A question that can be answered by the passage is: **Why might dinosaurs have needed more intelligence than many modern reptiles**?
12. Interpretation of dinosaur fossils and behavior was influenced by **false beliefs about brain size demonstrates the author's beliefs about dinosaurs**.

12.2 Appendix D. 2. Comet Restudying Sentence Facts.

1. Comets are made of **dust and ice**.
2. The nucleus of a comet is **full of holes**.
3. It is **unlikely that a comet pushed out of its orbit will come close enough to Earth to be seen by astronomers**.
4. A comet near the Earth may be less visible to astronomers if it **approaches Earth before it has moved closest to the Sun**.
5. The tails of a comet are created by **radiation pressure blowing dust off the coma and solar winds ionizing and attracting the gasses**.
6. A comet tail is pointed **away from the Sun** as the comet travels around the Sun.
7. The Rosetta probe only collected data for 2 days after it landed on the comet because **the probe happened to land in a crater, blocking it from view of the sun**.
8. Comets are considered a "fossil record" because **the materials that make up the nucleus are unchanged since the origin of the solar system**.
9. The Rosetta probe **did not find a previously undiscovered mineral**.
10. A near-Earth comet **does not require the viewer to have a powerful telescope**.

11. The passage **does not mention astronomers observing that comets give off a bright light from their nucleus**.
12. Scientists are most interested in comets directly because **they contain chemicals from the origin of the solar system**.

12.3 Appendix D. 3. The Great Barrier Reef Restudying Sentence Facts.

1. Humpback whales demonstrate their intelligence by **guiding ships away from storms**.
2. Humpback whales have their elaborate courtship rituals during the **winter season**.
3. Whale watchers are especially interested in seeing humpback whales' **breaching behavior**.
4. You can't identify a Clownfish by **how many fins it has**.
5. Sea anemone tentacles are **long and look like tree branches**.
6. A symbiotic relationship between clownfish and sea anemones is shown by **sea anemones providing defenses to counteract the clown fish's noticeable colors**.
7. The author notes that as of 2013 there have been 1600 shipwrecks in the Great Barrier Reef to make the point that **it's challenging for sailors to sail through**.
8. Coastal wetlands are beneficial to the Great Barrier Reef because they **filter out pollutants**.
9. A **reduction of agricultural pollutants** could help sustain the biodiversity of the Great Barrier Reef Ecosystem.
10. **Tortoises** do not live in the Great Barrier Reef.

11. If a biologist was looking at a different marine ecosystem that wasn't the Great Barrier Reef, they may determine a symbiotic relationship by **observing and noting which species interact in a way that provides mutual benefits for both parties**.
12. The article does not support the point statement that **reducing human interference would be the most effective way to reduce threats to the Great Barrier Reef**.

12.4 Appendix D. 4. Acupuncture Restudying Sentence Facts.

1. You feel as if **energy is moving within the body** if you get acupuncture.
2. Western doctors view acupuncture as **having potential benefits, although how it works is still unclear**.
3. During illness, the body's **Qi can get blocked**.
4. Acupuncture therapy continues to reduce pain even weeks after treatment because **endorphins can reduce inflammation and give the body time to heal**.
5. Multiple needles might be needed during acupuncture because **prolonged release of endorphins requires a buildup of signals from the body**.
6. Acupuncture and runner's highs are similar because **they both stimulate the same areas of the brain**.
7. Acupuncturists believe acupuncture improves degenerative eye disease because it **improves the functioning of the spleen**.
8. Acupuncture may **not help with nausea from the flu**.
9. Acupuncturists use the acupoints on the outside of the foot to treat degenerative eye disease because **they are connected via meridians to the spleen**.

10. The passage suggests that acupuncture research has **found evidence that acupuncture reduces inflammation**.
11. The author's perspective of acupuncture is that **there are multiple ways acupuncture may work and most of them are not well understood**.
12. The article does not support that **Western medicine has been ignoring the benefits of acupuncture treatment for too long**.

13.0 Appendix E. Experiment 1 Supplemental Material

13.1 Additional Analyses

Additional analyses were run to explore the role of difficulty on session two test performance. With this in mind, for Experiment 1, there were three fixed effects: study strategy intervention, the between-subjects participant condition variable (i.e., retrieval high knowledge and retrieval low knowledge), and difficulty. The fixed effects were contrasted coded as they were with the two-way interaction analyses.

Preliminary inspection of mean accuracy across the question types indicated that participants performed best on the easy questions, but comparably well for the medium and hard questions; we thus collapsed the latter two conditions into "harder" questions. The easy questions were used as the main comparison group. Full-passage questions were excluded from analyses exploring the impact of difficulty since they were not coded as being "easier" or "harder." Thus, we had three easier questions and six harder questions.

We initially considered random intercepts for participants, passage topic, and the forty-eight final test questions. Following model comparisons via the likelihood ratio test, we eliminated random intercepts that did not contribute significantly to the model, and the final model contained only random intercepts for participant variance and question variance; there was little variance at the topic level, and differences among the individual questions were accounted for by the random intercept.

Table 3 presents results from the main model. There were no significant main effects of study strategy (p = .59) or the between-subjects participant condition variable (p = .90).

Participants performed significantly better on the easier questions compared to the harder questions ($p < .001$) (see Table 4 for mean accuracies). None of the two-way interactions were significant. There was a significant three-way interaction between study strategy, the between-subjects participant condition variable, and difficulty regarding the hard questions ($p < .01$). Meaning, participants who studied via restudying and had background topic material performed significantly better on the harder session two test questions compared to the participants who studied via retrieval practice and did not receive background topic material.

Table 3

Results from mixed-effects model for Experiment 1 with the variable of difficulty.

***Note*. Coefficients = the estimate of each effect of each variable on accuracy on session two final test, as reported in log odds; SE = standard error of the estimate; ** $p < .01$, *** $p < .001$**

	Estimate	SE	*z*-value	*p*-value
Study Strategy Intervention	0.06	.11	0.54	.590
Between-subjects Variable	-0.01	.12	-0.12	.905
Difficulty: Harder	-1.53	.21	-7.16	< .001 ***
Study Strategy Intervention X Between-subjects Variable	-0.13	.22	-0.60	.552
Study Strategy Intervention X Difficulty: Harder	-0.14	.13	-1.05	.293
Between-subjects Variable X Difficulty: Harder	0.07	.13	0.49	.621
Study Strategy Intervention X Between-subjects Variable X Difficulty: Harder	0.69	.27	2.59	< .01 **

Table 4

Mean accuracy on the session two final test of Experiment 1 as a function of session one condition assignment.

Condition	Study Strategy	Background Topic Knowledge (Yes/No)	Difficulty	Accuracy
Retrieve Low Knowledge	Retrieval Practice	Yes	Easier	.73
Retrieve Low Knowledge	Restudying	No	Easier	.74
Retrieve High Knowledge	Retrieval Practice	Yes	Easier	.74
Retrieve High Knowledge	Restudying	No	Easier	.74
Retrieve Low Knowledge	Retrieval Practice	Yes	Harder	.44
Retrieve Low Knowledge	Restudying	No	Harder	.39
Retrieve High Knowledge	Retrieval Practice	Yes	Harder	.36
Retrieve High Knowledge	Restudying	No	Harder	.44

13.2 Exploratory Measures

Recall that, after session one, participants were asked to complete two free response questions: one about their strategy during the task (see Table 5) and the other about their overall experience (see Table 6). For strategy use, we conducted an informal coding of the responses into eight themes: reading quickly, re-reading the passage, reading out loud, practicing recall, memorization, no strategy, and other. For experience, we conducted an informal coding of the responses into five themes: enjoyable, easy, not enjoyable, too long, and other.

Table 5

Free response about strategy use at the end of session one for Experiments 1 and 2.

Coded Response	Frequency of Response (%)		Example
	Experiment 1	Experiment 2	
Reading in detail	23.4	27.4	"I would try and look at key words to help me get the overall picture of what the specific passages were trying to say."
Reading quickly	23.4	20.9	"I skimmed through the paragraphs and still understood [the] majority of the information given in the article."
Re-reading passage	20.2	13.0	"I just read the passages as many times as the time allowed for."
Reading out loud	3.2	2.2	"I would read the passages out loud because this is the strategy that works for me to retain information."
Practicing recall	8.1	7.8	"My strategy was to read a paragraph and break up the information to retain it and then if I had time go back and recap the important details and if I forget key parts to go back and read the passage."
Memorization	4.0	3.5	"Picked a comfortable position to be able to sit and focus on the readings. Trying to memorize key names and terms."

No strategy	12.9	18.7	“I didn’t have much of a strategy other than just try and understand what I was reading, and hope that the understanding would help me answer the questions.”
Other	4.8	6.6	“I read the passages slowly in order to absorb the information better. I also sat down in a quiet stress-free environment.”

Table 6

Free response about overall study experience at the end of session one for Experiments 1 and 2.

Coded Response	Frequency of Response (%)		Example
	Experiment 1	Experiment 2	
Enjoyable	45.2	58.3	"I [was] pretty [interested] to learn all this knowledge by reading research, but [it] also help[ed] me realize how the knowledge would be easily forgotten as time passed."
Easy	12.9	7.4	"It was a pretty straightforward and easy task."
Not enjoyable	11.3	4.8	"Overall, the task was not as enjoyable as I had hoped it would be, the passages went slower than my reading speed which caused me to lose focus and wait for the next page to load."
Too long	13.7	13.9	"There was not enough time to read each passage."
Other	16.9	15.6	"It kind of reminded me of when I took the SATs or other standardized testing, except I couldn't go back to check what I missed or didn't understand."

At the end of the experiment, participants were asked to answer a few additional questions regarding the study. Participants answered a series of Likert-scaled questions about their familiarity with the four topics they read about and about their experience with the study (see Table 7). These responses were scaled as 1 being low value (e.g., not familiar, or enjoyable) and 5 being a high value (e.g., very familiar or very enjoyable).

Table 7

Free response about strategy use at the end of session one for Experiments 1 and 2.

Question	Average Response	
	Experiment 1	Experiment 2
How familiar were you with the material on Dinosaurs before the experiment?	2.19	2.27
How familiar were you with the material on the Great Barrier Reef before the experiment?	2.36	2.14
How familiar were you with the material on Comets before the experiment?	1.70	1.90
How familiar were you with the material on Acupuncture before the experiment?	1.98	2.02
How much did you enjoy the task?	2.88	2.77
How engaging did you find the task?	2.85	2.87
How focused were you during the task?	3.40	3.23
For the science topics you read about twice, do you think reading the additional information helped you answer the questions?	3.51	3.41
Did you have enough time provided to you to read through the passages?	3.08	3.19

Participants also completed the MSLQ and Big 5 Questionnaires and self-reported their undergraduate GPA (see Table 8). The MSLQ measures were Likert-scaled with 1 being a low value on the scale and 7 being a high value on the scale. The Big 5 measures were Likert-scaled with 1 being a low value on the scale and 5 being a high value on the scale. GPA was reported on a 4.0 scale.

Table 8

Motivated Strategies for Learning Questionnaire and Big 5 Questionnaire average responses, and self-reported undergraduate GPA for Experiments 1 and 2.

Question	Average Response	
	Experiment 1	Experiment 2
MSLQ: Motivation Measures	5.24	5.07
MSLQ: Learning Measures	4.49	4.36
Big 5: Openness	3.38	3.25
Big 5: Conscientiousness	3.30	3.23
Big 5: Extraversion	2.50	2.47
Big 5: Agreeableness	3.62	3.46
Big 5: Neuroticism	1.91	2.00
GPA	3.46	3.45

At the end of session two, participants filled out another free response question regarding their overall experience with the task (see Table 9). These responses were coded in the same way as the overall experience question that followed session one and participants answered similarly at the conclusion of the study.

Table 9

Free response about overall study experience for the end of session two for Experiments 1 and 2.

Question	Frequency of Response (%)	
	Experiment 1	Experiment 2
Enjoyable	61.3	62.2
Easy	8.1	6.1
Not interesting	2.4	3.0
Too long	14.5	15.6
Other	13.7	13.0

14.0 Appendix F: Experiment 2 Supplemental Material

14.1 Additional Analyses

For Experiment 2, there were four fixed effects: study strategy intervention, background topic knowledge intervention, a between-subjects variable of feedback (named "experiment"), and difficulty. Study strategy intervention, background topic knowledge intervention, and the between-subject experiment variable were contrast coded as they were with the two-way interaction analyses. Difficulty was analyzed as it was in three-way interaction analysis for Experiment 1.

Regarding the four-way interaction model, there were no significant main effects of study strategy ($p = .20$) and background topic knowledge ($p = .36$). There was a significant main effect of difficulty where participants performed significantly better on the easier questions compared to the harder questions ($p < .001$). There was a significant main effect of the experiment variable where participants performed significantly better on the session two test if they received feedback in session one compared to not receiving feedback in session one ($p < .05$). There was a significant two-way interaction between study strategy intervention and the experiment variable where the effect of retrieval-practice in session one was amplified if participants also received feedback ($p = .01$). There were no significant three-way interactions, and the four-way interaction was also not significant. While some questions, specifically the "easier" questions, are more likely to be answered correctly, the other effects regarding study strategy and background accessibility are consistent across question difficulty.

14.2 Exploratory Measures

Before the experiment began, for Experiment 2 only, participants were asked four Likert-scaled questions about their familiarity with the four topics they would read about on a Likert-scale of one to five, where one indicated a low value and five indicated a high value (see Table 10).

Table 10

Likert questions about topic familiarity for Experiment 2 at the beginning session of one.

Question	Average Response
How familiar are you with the topic of Dinosaurs?	2.40
How familiar are you with the topic of the Great Barrier Reef?	2.10
How familiar are you with the topic of Comets?	1.95
How familiar are you with the topic of Acupuncture?	1.97

As with Experiment 1, participants filled out free response questions regarding their strategy (see Table 5) used during the task as well as their overall experience with the study (see Table 6) at the end of session one. Following session two, participants also completed Likert-scaled questions regarding topic familiarity and their experience with the task (see Table 7). Further, participants also completed the MSLQ, the Big 5, and self-reported their GPA (see Table 8), and a free response question regarding their overall experience at the end of the study (see Table 9).

Bibliography

Adesope, O. O., Trevisan, D. A., & Sundararajan, N. (2017). Rethinking the use of tests: A meta-analysis of practice testing. *Review of Educational Research*, *87*, 659–701. https://doi.org/10.3102/0034654316689306

Agarwal, P. K., Bain, P. M., & Chamberlain, R. W. (2012). The Value of Applied Research: Retrieval Practice Improves Classroom Learning and Recommendations from a Teacher, a Principal, and a Scientist. *Educational Psychology Review*, *24*(3), 437–448. https://doi.org/10.1007/s10648-012-9210-2

Agarwal, P. K., Finley, J. R., Rose, N. S., & Roediger, H. L. (2017). Benefits from retrieval practice are greater for students with lower working memory capacity. *Memory (Hove, England)*, *25*(6), 764–771. https://doi.org/10.1080/09658211.2016.1220579

Ariel, R., Dunlosky, J., & Bailey, H. (2009). Agenda-Based Regulation of Study-Time Allocation: When Agendas Override Item-Based Monitoring. *Journal of Experimental Psychology. General*, *138*, 432–447. https://doi.org/10.1037/a0015928

Barenberg, J., & Dutke, S. (2019). Testing and metacognition: Retrieval practise effects on metacognitive monitoring in learning from text. *Memory*, *27*(3), 269–279. https://doi.org/10.1080/09658211.2018.1506481

Barnett, S. M., & Ceci, S. J. (2002). When and where do we apply what we learn?: A taxonomy for far transfer. *Psychological Bulletin*, *128*(4), 612–637. https://doi.org/10.1037/0033-2909.128.4.612

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*, 1–48. https://doi.org/10.18637/jss.v067.i01

Bertilsson, F., Stenlund, T., Wiklund-Hörnqvist, C., & Jonsson, B. (2021). Retrieval practice: Beneficial for all students or moderated by individual differences? *Psychology Learning & Teaching*, *20*(1), 21–39. https://doi.org/10.1177/1475725720973494

Bjork, R. A., & Bjork, E. L. (1992). A new theory of disuse and an old theory of stimulus fluctuation. In A. Healy, S. Kosslyn, & R. Shiffrin (Vol. Eds.), *Learning processes to cognitive processes: Essays in honors of William Estes: Vol. 2*, (pp. 35–67). Hillsdale, NJ: Erlbaum.

Bruett, H., Fang, X., Kamaraj, D. C., Haley, E., & Coutanche, M. N. (2018). Expertise moderates incidentally learned associations between words and images. *Frontiers in Psychology*, *9*. https://doi.org/10.3389/fpsyg.2018.02085

Buchin, Z. L., & Mulligan, N. W. (2023). Retrieval-Based Learning and Prior Knowledge. *Journal of Educational Psychology*, *115*(1), 22–35. https://doi.org/10.1037/edu0000773

Butler, A. C., Karpicke, J. D., & Roediger, H. L. (2007). The effect of type and timing of feedback on learning from multiple-choice tests. *Journal of Experimental Psychology. Applied*, *13*(4), 273–281. https://doi.org/10.1037/1076-898X.13.4.273

Butler, A. C., Karpicke, J. D., & Roediger, H. L. (2008). Correcting a metacognitive error: Feedback increases retention of low-confidence correct responses. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *34*(4), 918–928. https://doi.org/10.1037/0278-7393.34.4.918

Butler, A. C., & Roediger, H. L. (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

Cacioppo, J. T., Petty, R. E., & Kao, C. F. (1984). The efficient assessment of need for cognition. *Journal of Personality Assessment*, *48*(3), 306–307. https://doi.org/10.1207/s15327752jpa4803_13

Carpenter, S. K. (2009). Cue strength as a moderator of the testing effect: The benefits of elaborative retrieval. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *35*(6), 1563–1569. https://doi.org/10.1037/a0017021

Carpenter, S. K., Lund, T. J. S., Coffman, C. R., Armstrong, P. I., Lamm, M. H., & Reason, R. D. (2016). A classroom study on the relationship between student achievement and retrieval-enhanced learning. *Educational Psychology Review*, *28*(2), 353–375. https://doi.org/10.1007/s10648-015-9311-9

Carpenter, S. K., Pashler, H., Wixted, J. T., & Vul, E. (2008). The effects of tests on learning and forgetting. *Memory & Cognition*, *36*(2), 438–448. https://doi.org/10.3758/MC.36.2.438

Castel, A. D., McCabe, D. P., Roediger, H. L., & Heitman, J. L. (2007). The Dark Side of Expertise: Domain-Specific Memory Errors. *Psychological Science*, *18*(1), 3–5. https://doi.org/10.1111/j.1467-9280.2007.01838.x

Chan, J. (2009). Long-term effects of testing on the recall of nontested materials. *Memory (Hove, England)*, *18*, 49–57. https://doi.org/10.1080/09658210903405737

Chase, W. G., & Simon, H. A. (1973). Perception in chess. *Cognitive Psychology, 4*(1), 55–81. https://doi.org/10.1016/0010-0285(73)90004-2

Chen, O., Castro-Alonso, J. C., Paas, F., & Sweller, J. (2018). Undesirable Difficulty Effects in the Learning of High-Element Interactivity Materials. *Frontiers in Psychology*, *9*. https://www.frontiersin.org/articles/10.3389/fpsyg.2018.01483

Chi, M. T. H., Feltovich, P. J., & Glaser, R. (1981). Categorization and Representation of Physics Problems by Experts and Novices. *Cognitive Science, 5*(2), 121–152. https://doi.org/10.1207/s15516709cog0502_2

Clownfish Facts. (n.d.). Great Barrier Reef Foundation. Retrieved April 5, 2022, from https://www.barrierreef.org/the-reef/animals/clownfish

Cogliano, M., Kardash, C. M., & Bernacki, M. L. (2019). The effects of retrieval practice and prior topic knowledge on test performance and confidence judgments. *Contemporary Educational Psychology*, *56*, 117–129. https://doi.org/10.1016/j.cedpsych.2018.12.001

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, *11*(6), 671–684. https://doi.org/10.1016/S0022-5371(72)80001-X

Detterman, D. (1993). The case for the prosecution: Transfer as an epiphenomenon. *Transfer on Trial: Intelligence, Cognition, and Instruction*.

Dweck, C. S. (1986). Motivational processes affecting learning. *American Psychologist*, *41*(10), 1040–1048. https://doi.org/10.1037/0003-066X.41.10.1040

Efklides, A. (2014). How does metacognition contribute to the regulation of learning? An integrative approach. *Psihologijske Teme*, *23*(1), 1–30.

Fazio, L. K., Huelser, B. J., Johnson, A., & Marsh, E. J. (2010). Receiving right/wrong feedback: Consequences for learning. *Memory (Hove, England)*, *18*(3), 335–350. https://doi.org/10.1080/09658211003652491

Gick, M. L., & Holyoak, K. J. (1987). The cognitive basis of knowledge transfer. In S. M. Cormier & J. D. Hagman (Eds.), *Transfer of learning: Contemporary research and applications* (pp. 9–46). Academic Press.

Gignac, G. E., & Szodorai, E. T. (2016b). Effect size guidelines for individual differences researchers. *Personality and Individual Differences, 102,* 74–78. https://doi.org/10.1016/j.paid.2016.06.069

Glaser, J., & Richter, T. (2023). The Testing Effect in the Lecture Hall: Does it Depend on Learner Prerequisites? *Psychology Learning & Teaching*, *22*(2), 159–178. https://doi.org/10.1177/14757257221136660

Glenberg, A. M., Wilkinson, A. C., & Epstein, W. (1982). The illusion of knowing: Failure in the self-assessment of comprehension. *Memory & Cognition*, *10*(6), 597–602. https://doi.org/10.3758/BF03202442

Hanham, J., Leahy, W., & Sweller, J. (2017). Cognitive Load Theory, Element Interactivity, and the Testing and Reverse Testing Effects. *Applied Cognitive Psychology*, *31*(3), 265–280. https://doi.org/10.1002/acp.3324

Harackiewicz, J. M., Barron, K. E., & Elliot, A. J. (1998). Rethinking achievement goals: When are they adaptive for college students and why? *Educational Psychologist*, *33*(1), 1–21. https://doi.org/10.1207/s15326985ep3301_1

Haskell, R. E. (2001). Chapter 2 - Transfer of Learning: What It Is and Why It's Important. In R. E. Haskell (Ed.), *Transfer of Learning* (pp. 23–39). Academic Press. https://doi.org/10.1016/B978-012330595-4/50003-2

Huet, N., & Mariné, C. (2005). Clustering and expertise in a recall task: The effect of item organization criteria. *Learning and Instruction*, *15*(4), 297–311. https://doi.org/10.1016/j.learninstruc.2005.07.005

Humpback Whale. (n.d.). Great Barrier Reef Foundation. Retrieved April 5, 2022, from https://www.barrierreef.org/the-reef/animals/humpback-whale

John, O. P., Donahue, E. M., & Kentle, R. L. (1991). *Big Five Inventory (BFI)* [Database record]. APA PsycTests. https://doi.org/10.1037/t07550-000

Jonsson, B., Wiklund-Hörnqvist, C., Stenlund, T., Andersson, M., & Nyberg, L. (2021). A learning method for all: The testing effect is independent of cognitive ability. *Journal of Educational Psychology*, *113*(5), 972–985. https://doi.org/10.1037/edu0000627

Kalakoski, V., & Saariluoma, P. (2001). Taxi drivers' exceptional memory of street names. *Memory & Cognition*, *29*(4), 634–638. https://doi.org/10.3758/BF03200464

Kang, S. H. K., McDermott, K. B., & Roediger, H. L. (2007). Test format and corrective feedback modify the effect of testing on long-term retention. *European Journal of Cognitive Psychology*, *19*(4–5), 528–558. https://doi.org/10.1080/09541440601056620

Karpicke, J. D., & Roediger III, H. L. (2008). The critical importance of retrieval for learning. *Science*, *319*, 966–968. https://doi.org/10.1126/science.1152408

Kornell, N., Bjork, R. A., & Garcia, M. A. (2011). Why tests appear to prevent forgetting: A distribution-based bifurcation model. *Journal of Memory and Language*, *65*(2), 85–97. https://doi.org/10.1016/j.jml.2011.04.002

Lehman, M., Smith, M. A., & Karpicke, J. D. (2014). Toward an episodic context account of retrieval-based learning: Dissociating retrieval practice and elaboration. *Journal of*

Experimental Psychology. Learning, Memory, and Cognition, *40*(6), 1787–1794. https://doi.org/10.1037/xlm0000012

Löffler, E., von der Linden, N., & Schneider, W. (2016). Influence of Domain Knowledge on Monitoring Performance Across the Life Span. *Journal of Cognition and Development*, *17*(5), 765–785. https://doi.org/10.1080/15248372.2016.1208204

McClelland, D. C. (1973). Testing for competence rather than for "intelligence." *American Psychologist*, *28*(1), 1–14. https://doi.org/10.1037/h0034092

Morris, C. D., Bransford, J. D., & Franks, J. J. (1977). Levels of processing versus transfer appropriate processing. *Journal of Verbal Learning and Verbal Behavior*, *16*(5), 519–533. https://doi.org/10.1016/S0022-5371(77)80016-9

Norberg, K. A. (2022) Avoiding miscomprehension: A metacognitive perspective for how readers identify and overcome comprehension failure [Unpublished doctoral dissertation]. University of Pittsburgh.

Pan, S. C., & Rickard, T. C. (2018). Transfer of test-enhanced learning: Meta-analytic review and synthesis. *Psychological Bulletin*, *144*(7), 710–756. https://doi.org/10.1037/bul0000151

Pashler, H., Cepeda, N. J., Wixted, J. T., & Rohrer, D. (2005). When Does Feedback Facilitate Learning of Words? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*, 3–8. https://doi.org/10.1037/0278-7393.31.1.3

Pintrich, P. R., & De Groot, E. V. (1990). *Motivated Strategies for Learning Questionnaire (MSLQ)* [Database record]. APA PsycTests. https://doi.org/10.1037/t09161-000

Pyc, M. A., & Rawson, K. A. (2010). Why testing improves memory: Mediator effectiveness hypothesis. *Science (New York, N.Y.)*, *330*(6002), 335. https://doi.org/10.1126/science.1191465

Roediger, H. L., & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science*, *17*(3), 249–255. https://doi.org/10.1111/j.1467-9280.2006.01693.x

Roediger III, H. L., Putnam, A. L., & Smith, M. A. (2011). Ten benefits of testing and their applications to educational practice. In *The psychology of learning and motivation: Cognition in education, Vol. 55* (pp. 1–36). Elsevier Academic Press. https://doi.org/10.1016/B978-0-12-387691-1.00001-6

Rohrer, D., & Pashler, H. (2010). Recent Research on Human Learning Challenges Conventional Instructional Strategies. *Educational Researcher*, *39*(5), 406–412. https://doi.org/10.3102/0013189X10374770

Rowland, C. A. (2014). The effect of testing versus restudy on retention: A meta-analytic review of the testing effect. *Psychological Bulletin*, *140*(6), 1432–1463. https://doi.org/10.1037/a0037559

Schindler, S., Pfattheicher, S., & Reinhard, M.-A. (2019). Potential negative consequences of mindfulness in the moral domain. *European Journal of Social Psychology*, *49*(5), 1055–1069. https://doi.org/10.1002/ejsp.2570

Shanks, L. L., & Serra, M. J. (2014). Domain familiarity as a cue for judgments of learning. *Psychonomic Bulletin & Review*, *21*(2), 445–453. https://doi.org/10.3758/s13423-013-0513-1

Singley, M. K., & Anderson, J. R. (1989). *The transfer of cognitive skill.* Cambridge, MA: Harvard University Press.

Soderstrom, N. C., & Bjork, R. A. (2015). Learning versus performance: An integrative review. *Perspectives on Psychological Science: A Journal of the Association for Psychological Science*, *10*(2), 176–199. https://doi.org/10.1177/1745691615569000

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

The Great Barrier Reef, Eighth Grade Reading Passage. (n.d.). Retrieved April 5, 2022, from https://www.readworks.org

Thiede, K. W. (1999). The importance of monitoring and self-regulation during multitrial learning. *Psychonomic Bulletin & Review, 6,* 662–667. https://doi.org/10.3758/BF03212976

Toppino, T. C., & Cohen, M. S. (2009). The testing effect and the retention interval: Questions and answers. *Experimental Psychology*, *56*(4), 252–257. https://doi.org/10.1027/1618-3169.56.4.252

Toth, J. P., Daniels, K. A., & Solinger, L. A. (2011). What You Know Can Hurt You: Effects of Age and Prior Knowledge on the Accuracy of Judgments of Learning. *Psychology and Aging*, *26*(4), 919–931. https://doi.org/10.1037/a0023379

Troyer, M. (2017). It's all in your head: Effects of expertise on real-time access to knowledge during written sentence processing. *Proceedings of the 39th Annual Meeting of Cognitive Science Society*. https://cogsci.mindmodeling.org/2017/papers/0234/paper0234.pdf

Tullis, J. G., Finley, J. R., & Benjamin, A. S. (2013). Metacognition of the testing effect: Guiding learners to predict the benefits of retrieval. *Memory & Cognition*, *41*(3), 429–442. https://doi.org/10.3758/s13421-012-0274-5

Types of Coral Reefs. (n.d.). Coral Reef Alliance. Retrieved April 5, 2022, from https://coral.org/en/coral-reefs-101/types-of-coral-reef-formations/

Unsworth, N., Spillers, G., & Brewer, G. A. (2012). Working memory capacity and retrieval limitations from long-term memory: An examination of differences in accessibility. *Quarterly Journal of Experimental Psychology (2006)*, *65*. https://doi.org/10.1080/17470218.2012.690438

van Loon, M. H., de Bruin, A. B. H., van Gog, T., & van Merriënboer, J. J. G. (2013). Activation of inaccurate prior knowledge affects primary-school students' metacognitive judgments and calibration. *Learning and Instruction*, *24*, 15–25. https://doi.org/10.1016/j.learninstruc.2012.08.005

Vicente, K. J., & Wang, J. H. (1998). An ecological theory of expertise effects in memory recall. *Psychological Review, 105*(1), 33–57. https://doi.org/10.1037/0033-295X.105.1.33

Weissgerber, S., & Reinhard, M.-A. (2018). Pilot Study on the Relationship of Test Anxiety to Utilizing Self-Testing in Self-Regulated Learning. *International Journal of Psychological Studies*, *10*(4), Article 4. https://doi.org/10.5539/ijps.v10n4p95

Wheeler, M. A., Ewers, M., & Buonanno, J. F. (2003). Different rates of forgetting following study versus test trials. *Memory (Hove, England)*, *11*(6), 571–580. https://doi.org/10.1080/09658210244000414

Wiklund-Hörnqvist, C., Jonsson, B., & Nyberg, L. (2014). Strengthening concept learning by repeated testing. *Scandinavian Journal of Psychology*, *55*(1), 10–16. https://doi.org/10.1111/sjop.12093

Witherby, A. E., & Carpenter, S. K. (2022). The rich-get-richer effect: Prior knowledge predicts new learning of domain-relevant information. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *48*(4), 483–498. https://doi.org/10.1037/xlm0000996

Xiaofeng, M., Xiao-e, Y., Yanru, L., & AiBao, Z. (2016). Prior knowledge level dissociates effects of retrieval practice and elaboration. *Learning and Individual Differences*, *51*, 210–214. https://doi.org/10.1016/j.lindif.2016.09.012

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

Yeo, D. J., & Fazio, L. K. (2019). The optimal learning strategy depends on learning goals and processes: Retrieval practice versus worked examples. *Journal of Educational Psychology*, *111*(1), 73–90. https://doi.org/10.1037/edu0000268