When Pretesting Fails to Enhance Learning Concepts From Reading Texts

Hannah Hausman and Matthew G. Rhodes
Colorado State University

Prior research suggests that people can learn more from reading a text when they attempt to answer pretest questions first. Specifically, pretests on factual information explicitly stated in a text increases the likelihood that participants can answer identical questions after reading than if they had not answered pretest questions. Yet, a central goal of education is to develop deep conceptual understanding. The present experiments investigated whether conceptual pretests facilitate learning concepts from reading texts. In Experiment 1, participants were given factual or conceptual pretest questions; a control group was not given a pretest. Participants then read a passage and took a final test consisting of both factual and conceptual questions. Some of the final test questions were repeated from the pretest and some were new. Although factual pretesting improved learning for identical factual questions, conceptual pretesting did not enhance conceptual learning. Conceptual pretest errors were significantly more likely to be repeated on the final test than factual pretest errors. Providing correct answers (Experiment 2) or correct/incorrect feedback (Experiment 3) following pretest questions enhanced performance on repeated conceptual test items, although these benefits likely reflect memorization and not conceptual understanding. Thus, pretesting appears to provide little benefit for learning conceptual information.

Keywords: pretesting, conceptual learning, error correction, testing effect

Previous research suggests that, although rereading is a popular study strategy (Karpicke, Butler, & Roediger, 2009; Kornell & Bjork, 2007; Morehead, Rhodes, & Delozier, 2016), it is not effective (e.g., Callender & McDaniel, 2009). Rather, testing students after they read a passage can improve memory for tested information and improve transfer to novel questions (Agarwal, Karpicke, Kang, Roediger, & McDermott, 2008; Butler, 2010; McDaniel, Anderson, Derbish, & Morrisette, 2007; Hinze, Wiley, & Pellegrino, 2013; Kang, McDermott, & Roediger, 2007; Roediger & Karpicke, 2006; Weinstein, McDermott, & Roediger, 2010). Thus, testing can enhance learning after an initial reading of a text.

Although testing often occurs after reading, our particular interest in these experiments is the finding that testing before reading—referred to as *pretesting*—can enhance learning (for a review, see Carpenter & Toftness, 2017). For example, Richland, Kornell, and Kao (2009) had participants read a two-page text about colorblindness. Across five experiments, participants who attempted to answer open-ended pretest questions before reading the text (e.g., "What is total color blindness caused by brain damage called?") were significantly more likely to correctly answer pretest questions that were repeated on the final test than participants who were given additional time to read the passage. Such benefits of pretesting over additional studying emerged even though all participants' pretest responses were wrong (see also Kornell, 2014; Kornell, Hays, & Bjork, 2009; Kornell & Vaughn, 2016). Similar benefits of pretesting for learning repeated questions from reading texts have been demonstrated with other types of pretests and final tests (e.g., multiple choice, cued recall, free recall; Bull & Dizney, 1973; Frase, 1968; Hamaker, 1986; Hartley & Davies, 1976; Little & Bjork, 2016; Peeck, 1970; Pressley, Tanenbaum, McDaniel, & Wood, 1990; Rickards, 1976a, 1976b; Rickards, Anderson, & McCormick, 1976; Sagaria & Di Vesta, 1978).

In light of such evidence, Pashler and colleagues (2007, p. 21) recommended in an Institute for Education Science practice guide that instructors consider "using 'prequestions' to activate prior knowledge and focus students' attention on the material that will be presented in class." Despite the practical recommendation, there is little evidence on the generality of pretesting benefits. First, the benefits of pretests typically do not extend to new factual questions (but see Carpenter & Toftness, 2017; Little & Bjork, 2016). For example, Richland et al. (2009) found that participants who answered pretest questions performed similarly on novel final test questions (i.e., questions that had not been previously pretested) compared with participants given additional time to read the passage but who had not completed a pretest (see also Pressley et al., 1990). Other studies have found that, relative to a no-pretest control condition, pretests can impair performance on final test questions that did not appear on the pretest (Peeck, 1970; Rickards, 1976a, 1976b; Sagaria & Di Vesta, 1978). Indeed, the more learners selectively attend to pretested information and not to other information, the more that performance on novel questions may be impaired (Carpenter & Toftness, 2017). To the extent that instructors' goal is for students to correctly answer new exam questions, and not only the questions they have practiced in advance, pretests have limited value.

Second, and of key interest for the present studies, the effects of pretesting have not been investigated with conceptual questions requiring learners to make inferences. A few studies have found inconsistent effects of pretest questions interspersed throughout a text that require learners to recognize examples of a concept (Andre, Mueller, Womach, Smid, & Tuttle, 1980; Felker & Dapra, 1975; Rickards, 1976a; Watts & Anderson, 1971). Yet, the majority of previous research has used questions whose answers were directly stated in the text, which we refer to as factual questions. For example, Little and Bjork (2016) asked the factual multiple-choice question, "The majority of Yellowstone National Park resides in which state?" prior to reading a text on Yellowstone National Park. The text contained the answer to the question, amid the phrase, "Approximately 96% of Yellowstone National Park is located in the state of Wyoming." Thus, the answer centered on a single fact embedded in the text.

In contrast, broader understanding often requires integration of facts that converge on knowledge that is not explicitly stated in the text in order to make a sound inference, what we term a concept. For example, participants may read the following facts stated explicitly in a text: "Sexual reproduction leads to a variety of gene combinations in offspring, genetic variability produces different physical traits, and evolution is driven by these differences in physical traits." We were interested in participants' ability to infer that evolution would not happen if offspring had the exact genetic makeup of only one of its parents (passages and questions from Thiede, Wiley, & Griffin, 2011). Although students should acquire factual knowledge (e.g., genetic variability produces different physical traits), a central goal of education is developing conceptual understanding (e.g., evolution will not occur if offspring replicates the genetics of a parent). The present experiments replicated previous research with factual pretests and, critically, examined whether conceptual pretests enhance learning of repeated conceptual information. Experiment 1 also tested whether learning extends to related conceptual information that was not pretested.

Pretests and Conceptual Learning

Pashler and colleagues (2007) suggested two ways pretesting might enhance learning: attention and elaboration. These mechanisms are not mutually exclusive and both predict that conceptual pretests will enhance learning of repeated, and possibly new, conceptual questions.

Attention

First, conceptual pretests could orient learners to attend more to conceptual information in the passage than they would have without pretest questions. Prior research indicates that learners may attend to different information in a text based on previous exposure to test questions (Rothkopf & Bisbicos, 1967; Sagerman & Mayer, 1987) or instructions regarding the type of test they should anticipate (e.g., Hinze et al., 2013; Thiede et al., 2011). For example, Thiede and colleagues (2011) instructed some participants that they would read a passage and then be tested on their memory for the facts explicitly stated in the passage. Other participants were informed that they would read a passage and then be tested on their ability to make inferences across multiple pieces of information presented in the passage. All participants were tested on both facts

and inferences. Participants who expected a factual test performed better on factual questions than inference questions, whereas participants who expected an inference test performed better on inference questions than factual questions.

In addition to explicit instructions, the materials may also signal the type of information that is important (Rothkopf & Bisbicos, 1967; Sagerman & Mayer, 1987). For example, Sagerman and Mayer (1987) had participants read four different science passages. Following each of the first three passages, participants answered open-ended conceptual questions or identified facts that had been explicitly stated in the passage (a control condition did nothing). After the fourth passage, all participants answered open-ended conceptual questions. Participants who had previously answered conceptual questions performed better on new conceptual questions than control participants or those who had identified facts. In the present experiments, participants did not receive explicit instructions about the type of test they should anticipate, but were given factual or conceptual pretest questions prior to reading a passage. If conceptual pretests orient learners to read for conceptual understanding broadly, then such pretests should enhance performance on future conceptual questions, regardless of whether they are repeated pretest questions or novel questions.

In contrast, if conceptual pretests encourage learners to read the text specifically for the answers to those questions (Peeck, 1970; Rickards, 1976a, 1976b; Sagaria & Di Vesta, 1978), then learning should be limited to repeated questions. However, pretests do not necessarily have to limit attention—and thus, learning—to pretested information. Carpenter and Toftness (2017) demonstrated that factual pretests preceding video lessons enhanced performance on repeated and novel factual questions. The authors posited that, compared to reading a text, watching a video made it more difficult for participants attend to only the information that was related to pretest questions. In the present experiments, the answers to the conceptual questions were not explicitly stated in the text and required integration across multiple sentences. Therefore, searching the text for the answer to a conceptual question could have a similar effect to watching a video, requiring participants to process information that was both relevant and irrelevant to the pretest questions. In contrast, participants could use key words from the factual pretests to focus their attention on the answers, which were explicitly stated in the text. In short, conceptual pretests could enhance learning of both repeated and novel questions, even if factual pretests do not.

However, previous research suggests that focused attention alone may not explain why pretests enhance learning. Little and Bjork (2016, Experiment 2) demonstrated that pretests enhanced learning, but that participants did not spend more time studying information addressed in the pretests. Furthermore, attempts to experimentally increase attention to relevant information in a text has not yielded the same benefits as pretesting. For example, Richland and colleagues (2009) observed that participants learned more information through pretesting than when that same information was not pretested, but was emphasized in the text in a bold or italicized font. In fact, participants who answered pretest questions learned more than participants who read the questions prior to reading the passage, but were instructed to memorize the questions rather than answer them (Experiment 5). Similarly, Pressley and colleagues (1990) found that participants learned more from answering pretest questions than from rating how well the ques-

tions were written. However, it remains possible that pretesting enhanced learning because it increased attention to relevant information more than the other manipulations of attention.

Elaboration

In addition to increasing attention to relevant information, attempting to answer conceptual pretest questions should directly enhance learning by activating prior knowledge, as Pashler and colleagues (2007) suggested. Indeed, Richland et al. (2009) proposed that the information that comes to mind when answering a pretest question ". . .may create a fertile ground for later encoding of the answer when it is eventually provided" in the passage (p. 252). More formally, the elaborative retrieval hypothesis posits that retrieval activates relevant prior knowledge, facilitating integration of new information and enhancing learning (Carpenter, 2009, 2011; Carpenter & Yeung, 2017; Kornell, 2014; Roediger & Butler, 2011; but see Karpicke, Lehman, & Aue, 2014). However, most experiments examining the benefits of elaboration have used lists of word pairs as the materials (but see Hinze et al., 2013).

In addition to enhancing learning of simple materials, including word pairs, elaboration facilitates text comprehension. When reading, learners can encode a text at different levels, ranging from superficial memory for the words presented (the surface form), to an organized representation of the information explicitly stated in the text (the textbase), to a deep understanding of the meaning of the text that incorporates information beyond the text itself (the situation model). Creating the situation model is a constructive process requiring the reader to make inferences across ideas presented in the text and integrate background knowledge (Kintsch, 1994, 1998; Zwaan, 1999).

Learners may struggle to comprehend texts about less familiar topics because reading will not necessarily automatically activate the relevant prior knowledge needed to form a situation model (Kintsch & Rawson, 2005). Accordingly, activities that engender elaboration by encouraging learners to generate explanations while reading can enhance learning from texts. For example, eighth grade students improved their comprehension of the circulatory system more when they were instructed to explain the text to themselves while reading than when they only read the text. The prereading instructions encouraged students to explain what each sentence meant to them, the new information it provided, and how it related to what they had already read (Chi, De Leeuw, Chiu, & LaVancher, 1994). Similarly, college students who read their biology text once, but engaged in *elaborative interrogation*—that is, periodically answered "why" questions as they read—earned higher test scores than students who read the text twice. Critically, the test questions addressed information that was not probed by the "why" questions (Smith, Holliday, & Austin, 2010; see Dunlosky, Rawson, Marsh, Nathan, & Willingham, 2013, for the generality of the benefits of self-explanation and elaborative interrogation).

Answering conceptual pretest questions could similarly activate prior knowledge and thus promote learning from reading texts. For example, attempting to answer a pretest question regarding the consequence of higher levels of carbon dioxide (CO_2) in the atmosphere could activate prior knowledge, including that higher levels of CO_2 are associated with warmer temperatures and recent news of polar bears struggling to find ice in the Arctic. Such prior knowledge should contribute to a situation model of the effects of CO_2 and facilitate inferences from the text that higher levels of CO_2 lead to higher sea levels. As we have operationalized different types of questions, factual pretest questions can be answered from the textbase, with answers explicitly stated in the text. In contrast, conceptual questions require the reader to make inferences across sentences and can only be addressed by the situation model. Therefore, to the degree that elaboration encourages readers to construct a coherent situation model, pretests may be particularly effective for enhancing conceptual learning, relevant for repeated, and possibly novel, conceptual final test questions.

In sum, prior research using factual pretests and research on text comprehension, more broadly, predict that conceptual pretests should enhance learning of repeated questions by increasing attention for concepts and activating related prior knowledge. Novel conceptual questions may also benefit to the extent that answering conceptual pretest questions and reading the passages supports participants' construction of a situation model and requires them to deeply process related, but nonpretested information.

The Current Study

The present experiments sought to test the theoretical mechanisms and boundary conditions for how pretests enhance learning and have significant practical applications for education. Indeed, although learning concepts is foundational in education, it is not apparent whether pretesting facilitates learning of concepts. Experiment 1 was designed to replicate previous research by examining whether factual pretests enhance learning of repeated and novel factual questions from reading texts. Unlike previous research, Experiment 1 also examined whether conceptual pretests benefit repeated and novel conceptual questions. Participants answered factual or conceptual multiple-choice questions prior to reading passages. All of the pretest questions were repeated on a final test along with novel questions. A no-pretest condition was included for comparison.

To anticipate, although we replicated prior findings by showing that pretests benefit performance on repeated factual questions, we did not detect any benefit of conceptual pretests. Accordingly, in Experiments 2 and 3, we investigated why conceptual pretests failed to enhance learning. In Experiment 2, we examined whether error correction on conceptual pretests are driven by confidence and whether error correction for conceptual pretests can be ameliorated by corrective item-by-item feedback during the pretest. Finally, in Experiment 3, we further explored the type of feedback that would promote learning from conceptual pretests.

Experiment 1

The primary goal of Experiment 1 was to investigate whether conceptual pretesting enhances learning of repeated and/or novel conceptual information. Participants either answered factual or conceptual pretest questions before reading expository passages. A control condition did not take a pretest and only read the passages. All of the pretest questions were repeated on the final test along with novel questions. We examined whether participants in the pretest conditions performed better on repeated questions relative to novel questions and relative to participants who did not take pretests.

Method

Participants. One hundred eight participants at Colorado State University participated in this 1-hr, one-session experiment for partial course credit in introductory psychology, as approved by the university's institutional review board. Thirty-five participants (22 females, 12 males reported; mean age = 19.97 years) were randomly assigned to the factual pretest condition, 37 participants were randomly assigned to the conceptual pretest condition (22 females, 13 males reported; mean age = 19.54 years), and 36 participants did not take a pretest (21 females, 14 males reported; mean age = 19.51 years). Three additional participants (2 factual pretest participants and 1 conceptual pretest participant) were excluded because they answered fewer than 90% of the pretest or final test questions. The targeted sample size for all experiments was 34 participants per condition and was calculated using G*Power (Faul, Erdfelder, Buchner, & Lang, 2007) based on a two-tailed paired *t* test with an alpha level of .05, power of .80, and a Cohen's *d* of .50 (from Richland et al., 2009).

Materials. The materials were taken from Thiede and colleagues (2011; personal communication, K.W. Thiede, October, 2015) and included six expository text passages, five factual questions per passage, and five conceptual questions per passage. The passages were an average of 710 words long (*SD* = 193 words) and covered the following topics: evolution, food allergies, ice ages, IQ, monetary policy, and volcanoes. The test questions were multiple-choice with four possible answers. The answers to factual questions were explicitly stated in the text. For example, participants were asked, "How much of the earth is covered by glaciers during an ice age?" and the answer (one third) was explicitly stated in the text, "An ice age is a period of time—usually millions or tens of millions of years—when vast glaciers cover as much as a third of the Earth's land surface." In contrast, answering conceptual questions required the reader to make inferences based on information presented in the text. For example, participants were asked to select a consequence of higher levels of CO_2 in the atmosphere. The answer (higher sea levels) was not stated in the text. Instead, participants had to answer this question by making an inference based on the following facts that were presented in different paragraphs in the passage: higher levels of CO_2 in the

atmosphere lead to warmer temperatures and cooler temperatures lead to lower sea-levels due to water being frozen in glaciers.

Design and procedure. The experiment had three phases in which participants (1) read six passages, (2) completed a 10-min math distractor task, and (3) took a final test about the passages (see Figure 1). Participants were randomly assigned to answer either factual or conceptual pretest questions before each passage. The no pretest condition did not answer any pretest questions and only read the passages in the first phase of the experiment. On the final test, all participants answered both factual and conceptual questions. When participants answered pretest questions, those same questions were repeated on the final test along with novel questions. That is, participants who answered factual pretest questions answered repeated and novel factual final test questions, in addition to novel conceptual final test questions. Similarly, participants who answered conceptual pretest questions answered repeated and novel conceptual final test questions, in addition to novel factual final test questions. Finally, all of the factual and conceptual final test questions were novel for participants who did not answer pretest questions (i.e., the no pretest condition). Therefore, we used a 3 (pretest type: factual, conceptual, vs. none) × 2 (final test type: factual vs. conceptual) × 2 (question novelty: novel vs. repeated) mixed-factor design with pretest type was manipulated as a between-participants variable and final test type and question novelty manipulated as within-participants variables.

During the reading phase of the experiment, all participants read the six passages in the same order (evolution, food allergies, ice ages, IQ, monetary policy, and volcanoes) and were given a minimum amount of time to read each passage on the basis of passage length (120 s, 120 s, 210 s, 160 s, 120 s, 140 s, respectively). The minimum reading time was based on a reading pace of between 250 and 300 words per minute. Once the minimum reading time had been reached, participants could continue reading or advance to the next trial.

In the factual pretest condition, three factual pretest questions preceded each passage. For each participant, the three pretest questions were randomly selected from the five factual questions associated with the passage. The topic of the upcoming passage (e.g., evolution) was displayed above each question and respond-
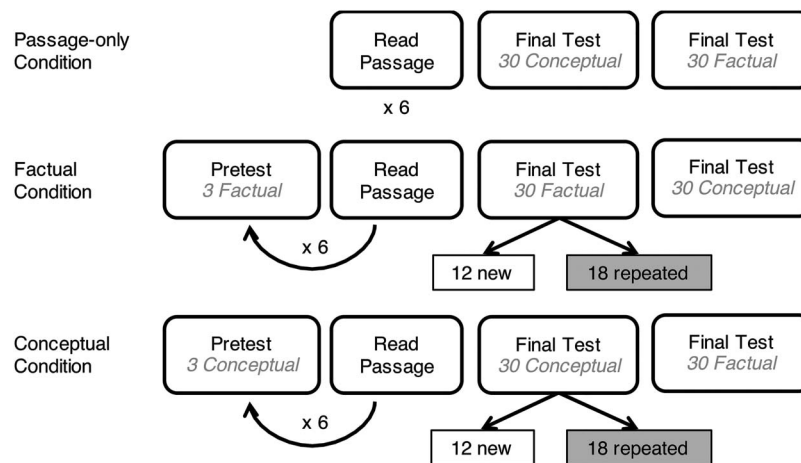

Figure 1. Experiment 1 procedure.

ing was self-paced. The procedure was the same in the conceptual pretest condition with the exception that, for each participant, the pretest questions were randomly selected from the conceptual questions associated with each passage. In total, participants in the pretest conditions answered three pretest questions, read the accompanying passage, and completed this cycle six times. Therefore, they answered either 18 factual or 18 conceptual pretest questions out of the possible 30 factual or 30 conceptual questions.

The final test came after a 10-min single-digit addition math distractor task. Participant answered all 30 factual and 30 conceptual test questions at their own pace (see Figure 1). Because of a programming error, all factual pretest participants answered factual final test questions first and all conceptual pretest participants answered conceptual final test questions first. Whether factual or conceptual pretest questions came first was only counterbalanced in the no-pretest condition.[1] Within each question type block (factual vs. conceptual), the questions were presented in the same order for all participants. The questions were arranged by topic and appeared in the same order as the passages (e.g., all factual evolution questions came before all factual food allergy questions).

Results

Pretest performance and reading times. Participants answered a similar proportion of factual ($M = 0.40$, $SD = 0.12$) and conceptual ($M = 0.42$, $SD = 0.15$) pretest questions correctly. In both conditions, participants performed significantly above chance (0.25), suggesting that they had some prior knowledge of the topics, $t(34) = 7.32$, $p < .001$, $d = 1.24$, $t(36) = 6.97$, $p < .001$, $d = 1.15$, respectively. We used an alpha level of .05 for this and all subsequent statistical analyses.

Following the pretests, participants in the factual ($M = 87.63$, $SD = 58.01$) and conceptual ($M = 92.53$, $SD = 57.34$) conditions spent an average of approximately 90 s reading the passage beyond the average required minimum of 145 s per passage. Participants who did not take pretests ($M = 100.56$, $SD = 50.17$) spent a similar amount of extra time reading. A one-way analysis of variance (ANOVA) revealed no significant differences in additional reading times among the three conditions, $F(2, 105) = 0.50$, $p = .61$, $\eta_p^2 = .009$. Furthermore, for each condition, reading time was uncorrelated with performance on any types of final test questions (novel factual, repeated factual, novel conceptual, repeated conceptual). The only exception was that participants in the no pretest condition who spent more time reading tended to be somewhat more accurate on factual final test questions, $r(36) = .35$, $p = .04$. Given that this was the only significant correlation detected, we treat it as spurious.

Final test performance. Consistent with previous research, factual pretest questions dramatically enhanced performance on those specific questions repeated on the final test (see Figure 2, left panel). This was true relative to the same participants' performance on novel factual final test questions, $t(34) = 5.12$, $p < .001$, $d = 0.88$, and relative to the factual final test performance of participants who did not take a pretest, $t(69) = 6.11$, $p < .001$, $d = 1.42$. However, the factual, conceptual, and no pretest conditions performed similarly on novel factual questions, $F(2, 105) = 0.21$, $p = .81$, $\eta_p^2 = .004$. Thus, factual pretests enhanced learning of repeated factual information, but did not enhance, or impair, learn-

ing of novel factual information. Conceptual pretests also did not impair learning of novel factual information.

In contrast with factual pretest questions, and contrary to our prediction, conceptual pretests did not enhance performance on repeated conceptual questions (see Figure 2, right panel). In fact, participants who answered conceptual pretest questions performed numerically, but not significantly, worse on repeated than novel conceptual final test questions, $t(36) = -1.67$, $p = .10$, $d = -0.28$, the opposite of what we predicted. Furthermore, their performance on repeated conceptual final test questions was similar to participants who did not take a pretest, $t(71) = -0.23$, $p = .82$, $d = -0.05$. The factual, conceptual, and no pretest conditions performed similarly on novel conceptual questions, $F(2, 105) = 0.48$, $p = .62$, $\eta_p^2 = .009$. Thus, conceptual pretests did not enhance, but also did not impair, learning of repeated or novel conceptual information. Factual pretests also did not impair learning of repeated or novel conceptual information.[2]

Bayesian analyses. A limitation of the frequentist approach to null-hypothesis significance testing is that the null hypothesis cannot be supported; it can only fail to be rejected. However, a key finding from Experiment 1 was a null result: conceptual pretests did not enhance learning of repeated relative to novel conceptual questions. A Bayesian approach to hypothesis testing allows the relative strength of evidence in favor of the null and alternative hypotheses to be compared (Jarosz & Wiley, 2014; Masson, 2011; Rouder, Speckman, Sun, Morey, & Iverson, 2009; Wagenmakers, 2007). Specifically, the Bayes factor serves as a useful alternative to traditional p values and is computed as the ratio of the likelihood of the data given the null hypothesis to the likelihood of the data given the alternative hypothesis (BF_{01}) or vice versa (BF_{10}). The strength of the evidence can be interpreted as weak ($1 < BF \leq 3$), positive ($3 < BF \leq 20$), strong ($20 < BF \leq 150$), and very strong (>150; Raftery, 1995; Wagenmakers, 2007). We followed the recommendations of Rouder and colleagues (2009) to use the JZS prior (which requires the fewest assumptions about the range of the effect size) and used their accompanying online Bayes factor calculator. We report Bayes factors for the key analyses of how factual and conceptual pretests affected final test performance.

The Bayesian analyses converged with the traditional analyses reported earlier. Specifically, the Bayes factors for the associated t tests very strongly supported that factual pretests enhanced learning of repeated factual questions. This was true relative to the same participants' performance on novel factual final test questions ($BF_{10} = 1741.29$) and relative to the factual final test performance

---

[1] Two key analyses were the within-participants comparison of performance on repeated versus novel factual questions in the factual pretest condition and repeated versus novel conceptual questions in the conceptual pretest condition. As within-participant comparisons, these analyses were not affected by the order of the question type blocks on the final test. We fully counterbalanced test block order in Experiment 2.

[2] The same pattern of results was observed when only comparing participants who answered the question type blocks in the same order on the final test. Relative to not taking a pretest, factual pretests enhanced performance on repeated factual final test questions, $t(52) = 3.39$, $p = .001$, $d = 0.97$, but not novel factual final test questions, $t(52) = -0.8$, $p = .43$, $d = -0.23$. There were no benefits of conceptual pretests. Relative to not taking a pretest, conceptual pretests did not enhance performance on repeated conceptual final test questions, $t(52) = 0.39$, $p = .70$, $d = 0.11$, or novel conceptual final test questions, $t(52) = 1.19$, $p = .24$, $d = 0.35$.
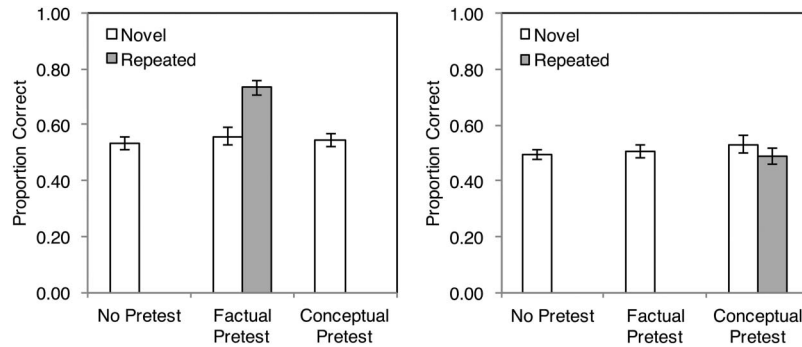
Figure 2. Proportion correct on factual final test questions (left) and conceptual final test questions (right) in Experiment 1. Error bars represent standard errors of the mean.

of participants who did not take a pretest ($BF_{10} = 2.04 \times 0^5$). In contrast, although participants performed numerically worse on repeated conceptual questions, the Bayes factors revealed evidence in favor of the null hypothesis that conceptual pretests do not affect memory for repeated conceptual questions. This was true relative to the same participants' performance on novel conceptual final test questions ($BF_{01} = 1.60$) and relative to the conceptual final test performance of participants who did not take a pretest ($BF_{01} = 4.40$).

Error correction.   Participants made a similar number of errors on the factual and conceptual pretests, yet only factual pretests enhanced learning. Figure 3 shows the outcome of factual (a) and conceptual (b) pretest errors on the final test. If participants selected an incorrect answer choice on the pretest, they could correct it on the final test, maintain their incorrect answer, or change to a different incorrect answer. We calculated the average of each

participant's proportion of errors that were corrected, maintained, and changed. The pattern of results was the same when the proportions were calculated out of the total number of all pretest errors, irrespective of which participant made them, and will not be discussed further.

The majority of factual pretest errors were corrected on the final test and were significantly more likely to be corrected ($M = .66$, $SD = .20$) than maintained ($M = .19$, $SD = .14$), $t(34) = 8.54$, $p < .001$, $d = 5.31$. In contrast, conceptual pretest errors were more likely to be maintained ($M = .44$, $SD = .12$) than corrected ($M = .33$, $SD = .19$), $t(36) = 2.38$, $p = .02$, $d = .40$. Thus, participants learned less from conceptual pretests than factual pretests because conceptual pretest errors were harder to correct: Conceptual pretest errors were significantly more likely to be maintained on the final test than factual pretest errors, $t(70) = 8.00$, $p < .001$, $d = 1.89$.

Discussion

Consistent with previous research, Experiment 1 demonstrated that factual pretests improved learning of repeated questions relative to novel factual questions and relative to not taking a pretest (Richland et al., 2009). Contrary to our hypothesis, conceptual pretests did not improve learning. The difference in the benefits of factual and conceptual pretests cannot be explained by engagement, as measured by the time participants spent reading the passages. There were no differences in reading times among the factual pretest, conceptual pretest, and no pretest conditions and reading time was unrelated to final test performance in each condition.

The difference in the benefits of factual and conceptual pretests also cannot be explained by background knowledge or question difficulty, as participants also answered a similar number of factual and conceptual pretest questions correctly. Similarly, participants who did not answer pretest questions answered a similar proportion of factual ($M = 0.53$, $SD = .13$) and conceptual ($M = 0.50$, $SD = .10$) final test questions correctly (see also Thiede et al., 2011).

Instead, the most striking differences between conditions were evident for error correction. Although both factual and conceptual pretests induced similar levels of errors, conceptual pretest errors were more likely than factual pretest errors to be repeated on the final test. Accordingly, the benefits of pretesting may not depend on whether the pretest questions can be answered correctly, but
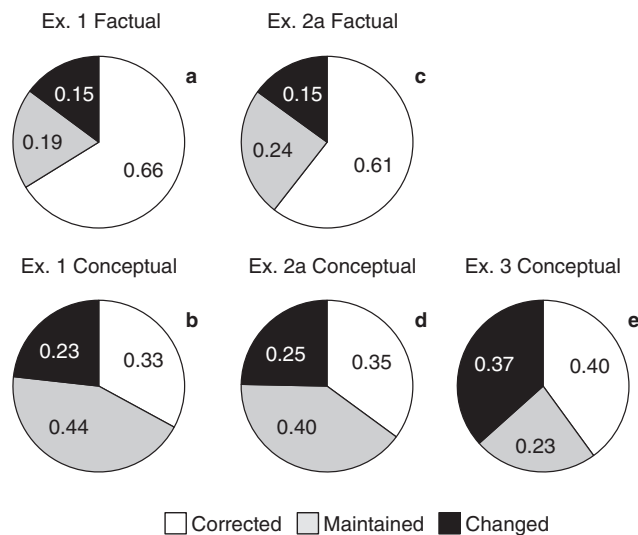

Figure 3. Proportion of incorrect pretest answers that were corrected, maintained, or changed to a new incorrect answer on the final test. The top row (a and c) shows the outcome of factual pretest errors in Experiments 1 and 2a, respectively. The bottom row (b, d, and e) shows the outcome of conceptual pretest errors in Experiments 1, 2, and 3, respectively. The proportions were calculated at the participant level and this figure shows the average of all of the participant proportions. Ex. = Experiment.

whether errors can be corrected. We further explored error correction in Experiment 2.

Experiment 2

Experiment 1 indicated that participants were far less likely to correct pretest errors made on conceptual questions than on factual questions. Experiment 2 tested two hypotheses regarding error correction. One hypothesis is that errors correction depends on participants' confidence in their errors. For example, prior research on the hypercorrection effect suggests that people are more likely to correct high confidence errors than low confidence errors (Butler, Fazio, & Marsh, 2011; Butterfield & Metcalfe, 2001, 2006; Fazio & Marsh, 2010; Sitzman, Rhodes, & Tauber, 2014). By extension, participants may have made more high-confidence factual errors than conceptual errors, increasing the chance of error correction for factual information. To understand the relationship between confidence and error correction, participants in Experiment 2 provided confidence ratings.

Experiment 2 also tested a second, related hypothesis that, regardless of confidence, participants did not make the necessary inferences while reading to correct pretest errors (cf. Otero & Kintsch, 1992). That is, answers to conceptual questions required participants to notice and integrate information presented across multiple sentences and paragraphs. Factual pretest participants did not have the same requirement because the answers to factual pretest questions were explicitly stated in the text. Therefore, half of the participants in Experiment 2a received correct-answer feedback after each factual or conceptual pretest question. If explicitly reading the answers is necessary for learning from pretests and correcting errors, then corrective feedback should dramatically improve learning of repeated relative to novel conceptual questions. If corrective feedback enhances learning, one might ask whether subsequently reading the passages has any additional benefit; we explore this question in Experiment 2b by including a group of participants who only answered pretest and final test question but were not exposed to the passages.

Accordingly, in Experiment 2, we sought to examine the relationship between confidence and feedback on error correction for conceptual and factual pretest questions. Given that we observed no differences in novel test question performance among the pretest conditions and the no-pretest condition we did not include a no pretest control condition. Rather, participants in Experiment 2 were randomly assigned to receive factual or conceptual pretests. All participants provided a confidence rating regarding the accuracy of their answer following each pretest question. Further, half of the participants were given corrective feedback following their answer and confidence rating to each pretest question. The remaining half of the participants were given no feedback. Of interest was (a) whether confidence in pretest answers was related to the likelihood of error correction and (b) whether corrective feedback would enhance performance on repeated questions, particularly in the conceptual pretest condition.

Experiment 2a

Method

Participants. Two hundred forty-six participants at Colorado State University participated in this 1-hr, one-session experiment for partial course credit in introductory psychology. One hundred sixty-four participants were randomly assigned to participate in Experiment 2a. Forty-two (35 females, 7 males; mean age = 18.7 years) were randomly assigned to the factual pretest passage-only condition, 40 participants were randomly assigned to the conceptual pretest passage-only condition (32 females, 7 males reported; mean age = 19.0 years), 42 participants were randomly assigned to the factual pretest Feedback + Passage condition (35 females, 7 males; mean age = 18.9 years), and 40 participants were randomly assigned to the conceptual pretest Feedback + Passage condition (35 females, 5 males; mean age = 19.1 years). Four additional participants (two factual pretest and two conceptual pretest participants) were excluded because they answered fewer than 90% of the pretest or final test questions, did not complete the experiment in the 1-hr time-slot, or did not follow the experimenter's instructions. Another 82 participants were randomly assigned to two additional conditions, reported in Experiment 2b.

Materials. The materials were identical to Experiment 1. The only difference was that the order of the four answer choices for each question was randomized for the final test.

Design and procedure. The general procedure was similar to Experiment 1. Participants answered either 3 factual or conceptual pretest questions (randomly selected for each participant), read the associated passage, and completed that cycle for each of six different passages. After a math distractor task, participants answered both factual and conceptual final test questions, with the order of the questions counterbalanced across participants. We used a 2 (pretest type: factual vs. conceptual) × 2 (feedback type: passage-only vs. Feedback + Passage) × 2 (final test type: factual vs. conceptual) mixed-factor design with pretest type and feedback type manipulated as between-participants variables and final test type was manipulated as a within-participants variable.

The key differences between Experiments 1 and 2 were in the pretest phase. Following each pretest question, all participants rated their confidence in the accuracy of their answer on a scale of 0 (*not confident at all*) to 10 (*extremely confident*). Participants were given as much time as needed to make their ratings. With the exception of the confidence ratings, the factual and conceptual passage-only conditions were nearly identical to the factual and conceptual pretest conditions of Experiment 1, respectively. We refer to them as the passage-only conditions because participants did not receive corrective item-by-item pretest feedback. Therefore, the only feedback they received on their pretest answers was the information contained in the passage itself. After each pretest question in the Feedback + Passage conditions, participants rated their confidence in their answer and then were given feedback for 10 s. The feedback indicated whether their answer was correct or incorrect, showed the topic and question again, and provided the correct answer. Finally, the math distractor task was reduced from 10 min in Experiment 1 to 5 min in Experiment 2 to accommodate the time needed for confidence ratings and feedback.

Results

Pretest performance and reading times. As in Experiment 1, participants answered a similar proportion of factual and conceptual pretest questions correctly. This was true regardless of whether participants were given corrective feedback after each pretest question (factual: $M = 0.39$, $SD = 0.12$; conceptual: 0.37,

$SD = 0.14$) or not (factual: $M = 0.41$, $SD = 0.16$; conceptual: 0.39, $SD = 0.14$). Participants demonstrated some preexisting knowledge of the topics before the experiment, as their pretest performance exceeded chance (0.25; $ts > 5$, $ps < .001$).

Table 1 shows the time participants spent reading beyond the minimum required time of 145 seconds per passage, on average. A 2 (pretest type: factual vs. conceptual) $\times$ 2 (feedback type: passage-only vs. Feedback + Passage) factorial ANOVA of reading times revealed no main effect of pretest type, $F(1, 160) = 0.008$, $p = .93$, $\eta_p^2 < .01$, and a modest tendency for participants in the passage-only condition to spend more time reading, $F(1, 160) = 3.49$, $p = .06$, $\eta_p^2 = .02$. These main effects were qualified by a significant interaction, $F(1, 160) = 6.21$, $p = .01$, $\eta_p^2 = .04$. Participants spent numerically, but not significantly, more time reading following conceptual pretests than factual pretests in the passage-only condition, $t(80) = 1.84$, $p = .07$, $d = .41$. The pattern reversed when they received corrective item-by-item pretest feedback, $t(80) = 1.70$, $p = .09$, $d = .38$. Given these differences, we conducted all key analyses with and without controlling for reading time. The pattern of findings did not change, indicating that differences in reading time cannot account for the effects of pretesting and feedback, reported next.

Final test performance. Experiment 2a examined how corrective feedback after each pretest question affected learning. Our primary question was whether feedback enhanced performance on repeated questions relative to novel final test questions. To answer this question, we only analyzed factual final test performance for participants who took factual pretests (see Figure 4, left panel), as they were the only participants to have both novel and repeated factual final test questions. Similarly, we only analyzed the conceptual final test performance of participants who took conceptual pretests (see Figure 4, right panel), as they were the only participants to have both novel and repeated conceptual final test questions (see the Appendix for an omnibus analysis of the factual and conceptual conditions).

Factual information. To investigate the effect of feedback on learning factual information, we conducted a 2 (feedback type: passage-only vs. Feedback + Passage) $\times$ 2 (question novelty: novel vs. repeated) mixed-factor ANOVA of factual final test performance among factual pretest participants (see Figure 4, left panel). Participants performed significantly better on repeated questions than novel questions, $F(1, 82) = 180.12$, $p < .001$, $\eta_p^2 = .69$, and significantly better when they received feedback after each pretest question, $F(1, 82) = 11.64$, $p = .001$, $\eta_p^2 = .12$. Finally, there was a significant Question Novelty $\times$ Feedback Type interaction, $F(1, 82) = 28.00$, $p < .001$, $\eta_p^2 = .26$. Specifically, feedback after each pretest question enhanced performance for repeated factual questions on the final test, $t(82) = 7.90$, $p < .001$, $d = 1.74$, $BF_{10} = 5.73 \times 10^8$, but did not affect on performance on novel factual questions, $t(82) = -0.76$, $p = .45$,

$d = 0.17$, $BF_{01} = 3.41$. The same pattern of results emerged when controlling for additional reading time as a covariate.

The passage-only condition replicated Experiment 1 and demonstrated the same result: Pretesting significantly improved performance on repeated factual questions relative to novel factual questions, $t(41) = 5.40$, $p < .001$, $d = 0.84$, $BF_{10} = 5849$. The benefit of pretesting on repeated questions over novel questions was even larger when participants received corrective feedback after each pretest question, $t(41) = 14.23$, $p < .001$, $d = 2.38$, $BF_{10} = 2.96 \times 10^{14}$. Importantly, factual pretests did not impair performance on conceptual final test questions. Performance on novel conceptual questions did not differ between participants who took factual pretests and participants who took conceptual pretests in the passage-only condition, $t(80) = -1.40$, $p = .17$, $d = -.32$, $BF_{01} = 1.86$, or the Feedback + Passage condition, $t(80) = .10$, $p = .92$, $d = .03$, $BF_{01} = 4.33$. The same pattern of results emerged when controlling for additional reading time as a covariate.

Conceptual information. To investigate the effect of feedback on learning conceptual information, we conducted a 2 (feedback type: passage-only vs. Feedback + Passage) $\times$ 2 (question novelty: novel vs. repeated) mixed-factor ANOVA of conceptual final test performance among participants who took conceptual pretests (see Figure 4, right). Participants performed significantly better on repeated questions than novel questions, $F(1, 78) = 81.16$, $p < .001$, $\eta_p^2 = .51$, and significantly better when they received feedback after each pretest question in addition to reading the passage, $F(1, 78) = 26.11$, $p < .001$, $\eta_p^2 = .25$. These main effects were qualified by a significant Question Novelty $\times$ Feedback Type interaction, $F(1, 78) = 41.95$, $p < .001$, $\eta_p^2 = .35$. Specifically, feedback after each pretest question enhanced performance on repeated conceptual questions on the final test, $t(78) = 8.12$, $p < .001$, $d = 1.85$, but had little effect on performance on novel conceptual questions, $t(78) = .47$, $p = .47$, $d = 0.16$.

The passage-only condition replicated Experiment 1 and demonstrated the same result: Pretesting did not improve performance on repeated conceptual questions relative to novel conceptual questions, $t(39) = 1.74$, $p = .09$, $d = 0.28$, $BF_{01} = 1.49$. However, participants did perform significantly better on repeated than novel questions when they received feedback after each pretest question, $t(39) = 11.32$, $p < .001$, $d = 1.98$, $BF_{10} = 1.10 \times 10^{11}$. Performance on novel factual questions did not differ between participants who took conceptual pretests and participants who took factual pretests in the passage-only condition, $t(80) = .21$, $p = .84$, $d = .05$, $BF_{01} = 4.27$, or the Feedback + Passage condition, $t(80) = .63$, $p = .53$, $d = .14$, $BF_{01} = 3.66$.

Error correction. We calculated the average of each participant's proportion of pretest errors that were corrected, maintained, or changed on the final test. Figure 3c shows the pattern of error correction for the factual pretest passage-only condition and Figure 3d shows the pattern of error correction for the conceptual pretest

Table 1

Mean (Standard Deviation in Parentheses) of Additional Time Spent Reading in Experiment 2a

Feedback type	Factual pretest condition	Conceptual pretest condition
Passage only	75.14 (50.39)	93.77 (48.56)
Correct Answer + Passage	79.64 (41.83)	62.31 (43.34)

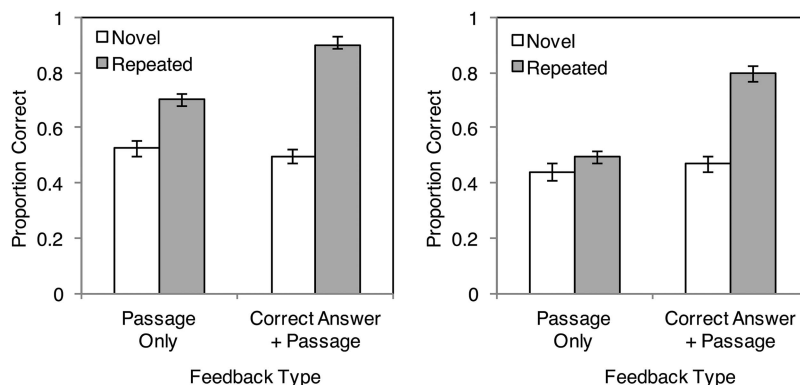*Note.* Additional reading times are reported in seconds.

Figure 4. Proportion correct on factual final test questions (left) among participants who took factual pretests and proportion correct on conceptual final test questions (right) among participants who took conceptual pretests in Experiment 2a. Error bars represent standard errors of the mean.

passage-only condition. As in Experiment 1, when participants did not receive feedback on their factual pretest errors, they corrected significantly more errors ($M = .60$, $SD = .21$) than they maintained ($M = .25$, $SD = .17$), $t(41) = 6.16$, $p < .001$, $d = .95$. In contrast, when participants did not receive feedback on their conceptual errors, they maintained ($M = .40$, $SD = .17$) and corrected ($M = .35$, $SD = .16$) a similar number of errors, $t(39) = 1.01$, $p = .32$, $d = .16$. Replicating Experiment 1, participants learned less from conceptual pretesting than factual pretesting because conceptual pretest errors were harder to correct: Participants maintained a significantly larger proportion of their incorrect conceptual pretest answers than their incorrect factual pretest answers on the final test, $t(80) = 4.02$, $p < .001$, $d = 0.90$.

Consistent with the finding that corrective feedback dramatically enhanced learning of repeated questions, participants corrected the majority of their factual ($M = 0.89$, $SD = .10$) and conceptual pretest errors ($M = 0.78$, $SD = .20$) when they were given feedback after each pretest question.

Confidence ratings. In Experiment 2, participants rated their confidence in the accuracy of each pretest answer on a 0 (*not confident at all*) through 10 (*extremely confident*) scale.[3] We report average confidence ratings in Table 2 for descriptive purposes. As Table 2 shows, participants were more confident in their correct pretest answers than their incorrect pretest answers. Confidence ratings were also similar for factual and conceptual questions and, on average, the magnitude of confidence ratings, on average, did not distinguish between factual and conceptual pretest errors.

However, the key questions were whether, on an item-by-item basis, confidence distinguished between pretest answers that were correct and incorrect on the pretest and between pretest answers that were corrected or maintained on the final test. We calculated the Goodman-Kruskal gamma correlation (Nelson, 1984) between each participant's pretest confidence and accuracy. We note that variations in degrees of freedom reflect instances in which participants had invariant confidence ratings or accuracy and a correlation could not be calculated.

Participants gave higher confidence ratings to correct than incorrect pretest answers and the correlations were significantly greater than 0 in both the factual ($M = .25$, $SD = .33$), $t(41) = 5.01$, $p < .001$, $d = 0.77$, and conceptual pretest conditions ($M =$

.22, $SD = .34$), $t(39) = 4.13$, $p < .001$, $d = 0.65$. However, the strength of the relationship between confidence and accuracy was similar in the factual and conceptual conditions, $t(80) = 0.48$, $p = .63$, $d = 0.11$. Thus, confidence ratings discriminated between correct and incorrect answers on the pretest and discrimination was similar for conceptual and factual pretests.

We also examined whether confidence ratings distinguished between errors that were corrected or maintained by calculating the gamma correlation between each participant's confidence in their pretest errors and whether those errors were corrected on the final test (cf. Butterfield & Metcalfe, 2001). A positive correlation would indicate that higher confidence errors were more likely to be corrected whereas a negative correlation would indicate that lower confidence errors were more likely to be corrected. Confidence was negatively correlated with both factual error correction ($M = -.03$, $SD = .58$) and conceptual error correction ($M = -.16$, $SD = .50$). However, the negative correlation did not differ from zero in the factual condition, $t(41) = -0.28$, $p = .79$, $d = 0.04$, and was only marginally different from zero in the conceptual condition $t(38) = -2.00$, $p = .05$, $d = 0.32$. Critically, the gamma correlation between confidence and error correction did not differ between the factual and conceptual conditions, $t(79) = 1.12$, $p = .27$, $d = .25$. Thus, confidence was not a strong predictor of error correction and the relationship between confidence and error correction did not differ between the factual and conceptual conditions. Therefore, confidence likely does not account for differences in error correction between these conditions.

The general pattern of results did not change when participants received feedback after each pretest question (see Table 2 for confidence ratings on the pretest). The gamma correlations between confidence and accuracy on the pretest was significantly different from zero for participants who took factual pretests ($M = .24$, $SD = .37$), $t(41) = 4.14$, $p < .001$, $d = .64$, and conceptual pretests ($M = .21$, $SD = .38$), $t(38) = 3.56$, $p = .001$, $d = .57$. The correlations did not differ significantly across the two pretest

---

[3] Four participants in the passage-only condition and 2 participants in the Feedback + Passage condition gave one confidence rating outside of the allowable 0 to 10 range. Those six ratings were treated as missing data and removed from confidence analyses.

Table 2
Mean (Standard Deviation in Parentheses) Pretest Confidence Ratings in Experiment 2

Feedback type	Pretest accuracy	Factual questions	Conceptual questions
Passage only	Incorrect	4.67 (1.39)	4.87 (1.43)
Passage only	Correct	5.51 (1.66)	5.59 (1.44)
Correct Answer + Passage	Incorrect	4.31 (1.26)	4.50 (1.68)
Correct Answer + Passage	Correct	5.13 (1.63)	5.02 (1.88)

Note. Confidence ratings were given on a scale of 0 (*not confident at all*) to 10 (*extremely confident*).

conditions, $t(79) = .29$, $p = .46$, $d = .06$. Furthermore, the gamma correlation between confidence and error correction was not significantly different from zero for participants who took factual pretests ($M = .002$, $SD = 0.68$), $t(26) = .02$, $p = .99$, $d = .003$, or conceptual pretests ($M = .04$, $SD = 0.67$), $t(30) = .02$, $p = .76$, $d = .05$. In addition, gamma correlations between confidence and error correction for the factual and conceptual pretest conditions did not differ, $t(58) = -0.19$, $p = .85$, $d = .05$. Thus, confidence distinguished between correct and incorrect answers on the pretest, but did not predict factual or conceptual error correction when participants received feedback after each pretest answer.

Discussion

Experiment 2a replicated Experiment 1: When participants were not given feedback, factual pretests enhanced learning of repeated questions relative to novel questions, but conceptual pretests did not. Relative to novel conceptual questions, performance was numerically worse on repeated conceptual questions in Experiment 1 ($d = .28$), but numerically better on repeated conceptual questions in Experiment 2a ($d = .28$). Furthermore, Bayesian analyses revealed weak evidence in favor of the null hypothesis that conceptual pretests lead to no differences in performance on repeated and novel conceptual questions in Experiment 1 ($BF_{01} = 1.60$) and in Experiment 2a ($BF_{01} = 1.49$).

Participants were also significantly more likely to repeat their conceptual pretest errors on the final test than their factual pretest errors. Data from confidence ratings do not appear to account for these differences. In particular, participants gave similar average confidence ratings following factual and conceptual pretest errors (see Table 2), and more importantly, there was no correlation between confidence in a factual or conceptual pretest error and the probability of correcting that error on the final test.

Another more likely explanation for why conceptual pretests did not enhance learning is that, in contrast to factual questions, the answers to conceptual pretest questions are not explicitly stated in the text but must be inferred. Experiment 2a tested this hypothesis by giving corrective feedback to one half of the participants, explicitly stating the correct answer after each pretest question. Such corrective feedback dramatically improved learning of repeated, but not novel, conceptual questions (see Figure 4, right). Corrective feedback also improved learning of repeated factual questions, although to a lesser extent because factual pretests already benefitted learning even without feedback (see Figure 4, right). Therefore, pretests can enhance learning when the answers are in the text, but immediate corrective feedback enhances learning more (Rothkopf, 1966). Experiment 2b further examined learning from the passage versus feedback.

Experiment 2b

Given the large benefit of corrective feedback on learning, we also examined whether participants benefited from reading the passages at all after receiving feedback.

Method

Among the 246 participants who completed Experiment 2, 86 were randomly assigned to participate in Experiment 2b. Participants took a pretest with factual ($N = 41$; 31 females, 10 males; mean age = 19.0 years) or conceptual ($N = 41$; 31 females, 10 males; mean age = 18.3 years) questions, rated their confidence, and received corrective feedback. However, unlike Experiment 2a, they did not read the passages and, instead, immediately completed a 5-min math distractor task and then took the final test.

Results and Discussion

Similar to the Correct Answer + Passage conditions of Experiment 2a, participants answered most of the repeated factual ($M = .91$, $SD = .07$) and repeated conceptual ($M = .86$, $SD = .16$) questions correctly on the final test. However, because participants did not read the passages, their performance on novel factual ($M = .34$, $SD = .02$) and novel conceptual ($M = .37$, $SD = .02$) final test questions was similar to their performance on factual ($M = .38$, $SD = .02$) and conceptual ($M = .38$, $SD = .02$) pretest questions. As in the Feedback + Passage conditions of Experiment 2, participants who took factual pretests correctly answered significantly more repeated than novel factual questions, $t(40) = 29.67$, $p < .001$, $d = 4.73$, $BF_{10} = 1.14 \times 10^{25}$; participants who took conceptual pretests correctly answered significantly more repeated than novel conceptual questions, $t(41) = 18.08$, $p < .001$, $d = 2.82$, $BF_{10} = 1.10 \times 10^{18}$ (see Figure 5).

Although performance on repeated questions was comparable in the correct-answer + passage and feedback-only conditions, we note that the delay between the pretest and the final test was approximately 15 min shorter than in the Feedback + Passage condition. Thus, differences in the retention interval may, in part, account for the similar levels of performance. Nevertheless, the feedback-only condition answered most questions correctly on the final test, suggesting that much of the learning from pretests with feedback could come from feedback and not from reading the passage more effectively as a result of the feedback.

Discussion of Experiments 2a and 2b

Experiments 2a and 2b suggests that corrective feedback significantly enhanced learning from pretests, although the exact mech-
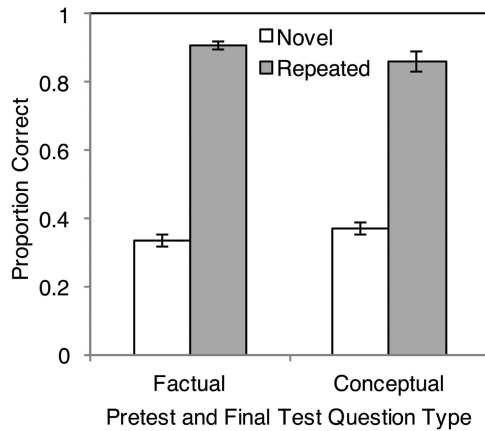
Figure 5. Proportion correct on factual final test questions among participants who took factual pretests (left two bars) and proportion correct on conceptual final test questions among participants who took conceptual pretests (right two bars) in Experiment 2b. Participants received correct-answer feedback after each question, but did not read the passages. Error bars represent standard errors of the mean.

anism is not clear. Specifically, the corrective feedback both signaled whether an answer was correct or incorrect and also provided the correct answer. Experiment 2b suggests that being provided the correct answer was critical to learning, given that participants learned similar amounts from feedback when they did and did not read the passage. Thus, although conceptual pretests with feedback enhanced performance on repeated questions, the benefits likely reflect memorization of the feedback, rather than true conceptual learning. Because memorizing the answer to a conceptual question does not require the reader to make inferences beyond information explicitly stated in the text, providing corrective feedback had the effect of transforming conceptual questions into factual ones. Therefore, if an instructor's goal is to encourage students to make inferences while they read, the results of Experiments 1 and 2 suggest that conceptual pretests would likely be ineffective, even if corrective feedback was provided.

Although participants likely memorized the answers to pretest questions when they were provided as feedback, it does not preclude the possibility that they could also benefit from only being told whether their pretest answer was correct or incorrect (Marsh, Lozito, Umanath, Bjork, & Bjork, 2012). The advantage of correct/incorrect feedback over corrective feedback is that it could change the information participants attend to when reading, without providing them an answer to memorize. Therefore, Experiment 3 examined whether correct/incorrect feedback could enhance conceptual learning from pretests, without the drawback of transforming a conceptual pretest question into a factual one by explicitly providing the answer.

Experiment 3

The purpose of Experiment 3 was to investigate whether correct/incorrect feedback on pretests enhances conceptual learning. We only included a conceptual pretest condition because the primary focus of this article is conceptual learning and because participants benefited from factual pretests regardless of whether feedback was

provided. Experiment 3 was nearly identical to the Conceptual Feedback + Passage condition of Experiment 2a. However, rather than corrective feedback, participants were only told whether their pretest answers were correct or incorrect. If feedback can only enhance learning by explicitly providing the correct answer, then conceptual pretests with correct/incorrect feedback should have minimal benefit. Pretests with correct/incorrect feedback should still benefit learning if feedback helps participants discriminate between correct and incorrect answers so that they can read the passage for the correct answers.

Method

Participants. Forty-three participants (14 females, 29 males; mean age = 19.09 years) at Colorado State University participated in this 1-hr, one-session experiment for partial course credit in introductory psychology.

Materials, design, and procedure. Experiment 3 was almost identical to the Feedback + Passage condition of Experiment 2. The two differences were that all participants took conceptual pretests and the feedback after each question only stated whether the previous answer was correct or incorrect. The feedback did not provide the correct answer.

Results

Pretesting performance and reading times. Consistent with previous experiments, participants answered more conceptual pretest questions correctly ($M = .37$, $SD = .19$) than would be expected by chance alone (.25), suggesting that participants had some preexisting knowledge of the passage topics, $t(42) = 7.04$, $p < .001$, $d = 1.07$. Following the pretests, participants spent an average of an additional 65.2 seconds ($SD = 47.10$) reading the passages beyond the minimum average time required of 145 seconds per passage. However, extra reading time was not significantly correlated with performance on repeated conceptual or novel factual questions, aside from a modest, positive correlation between reading time and performance on novel conceptual questions, $r(43) = .30$, $p = .05$.

Final test performance. Because all participants answered conceptual pretest questions, all factual final test questions were novel and participants performed similarly to the prior experiments ($M = .52$, $SD = .13$). Pretesting significantly enhanced performance on repeated conceptual questions ($M = .56$, $SD = .16$) relative to novel conceptual questions ($M = .43$, $SD = .19$) on the final test, $t(42) = 4.20$, $p < .001$, $d = .64$, $BF_{10} = 174.13$ (see Figure 6). Thus, although participants did not benefit from conceptual pretests without feedback (Experiments 1 and 2a), they did benefit from conceptual pretests when told which pretest answers were correct.

Error correction. One explanation for the improvement in performance on repeated conceptual questions is that correct/incorrect feedback identified pretest errors, permitting participants to read the text with the goal of finding the correct answers. Another explanation is that correct/incorrect feedback simply helped participants avoid selecting the same incorrect answer from the pretest on the final test. When participants made a pretest error they learned that the response option chosen was incorrect. Therefore, the correct/incorrect feedback reduced the number of possible
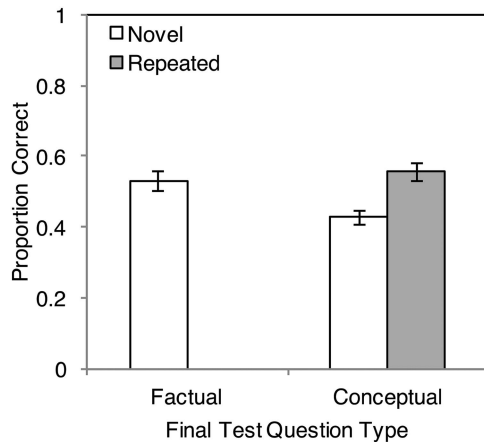
Figure 6. Proportion correct on factual and conceptual final test questions in Experiment 3. All participants took conceptual pretests and received correct/incorrect feedback on their pretest answers. Error bars represent standard errors of the mean.

correct answers from four to three. If participants could remember each of their incorrect pretest answers, then they could correct one third of their pretest errors on the final test through guessing alone. A one sample t test revealed that participants corrected significantly more pretest errors ($M = .40$, $SD = .19$) than would be expected through guessing alone (.33), $t(42) = 2.29$, $p = .03$, $d = .35$.

However, additional error correction data suggests that participants used the strategy of recalling their incorrect pretest answer and avoided selecting that same answer on the final test. Unlike Experiments 1 and 2, when participants did not correct an error, they were more likely to change to a new incorrect answer (0.37) than maintain their original incorrect answer (0.23; see Figure 3e). Thus, correct/incorrect feedback likely enhanced performance on repeated questions both by helping participants learn correct information from reading the passage and avoid repeating pretest errors on the final test.

Confidence. Participants gave similar confidence ratings, on average, to their correct pretest answers ($M = 5.09$, $SD = 1.55$) than their incorrect pretest answers ($M = 4.94$, $SD = 1.45$). Yet, the gamma correlation between confidence and pretest accuracy revealed that confidence discriminated between correct and incorrect pretest answers ($M = .23$, $SD = .38$) and was significantly different form zero, $t(42) = 4.00$, $p < .001$, $d = 1.23$. In contrast, confidence in pretest errors did not predict whether an error was corrected or not on the final test: The gamma correlation between confidence and error correction ($M = .04$, $SD = .47$) did not differ from zero, $t(41) = 0.57$, $p = .57$, $d = 0.09$.

Discussion

Experiment 3 revealed that participants benefited from pretest feedback for conceptual pretest questions, even when it only indicated whether an answer was correct. Such feedback may enhance learning by encouraging participants to selectively read the passage for the correct answers to pretest errors and to avoid previous errors on a subsequent test. Future research could differentiate between these explanations.

However, we note that the robust benefits of corrective feedback in Experiment 2 may not be solely explained by the fact that it indicates the accuracy of each pretest answer. For example, a comparison of effect sizes for the benefit of pretests for repeated conceptual questions over novel conceptual questions indicated that the impact of feedback was approximately three times larger when it included the correct answer (Experiment 2a, $d = 1.98$) than when feedback only indicated whether an answer was correct (Experiment 3, $d = .64$). Thus, corrective feedback may have enhanced performance on repeated conceptual questions in Experiment 2 by explicitly providing the correct answers.

General Discussion

Previous research has demonstrated that pretests can enhance learning of factual information. The present studies replicated past research: Participants were more likely to answer factual final test questions correctly if they attempted to answer those questions on a pretest. The benefits of factual pretests emerged relative to not taking a pretest (Experiment 1) and relative to novel factual questions that did not appear on the pretest (Experiments 1 and 2a). Unlike previous research, we asked whether pretests could also enhance conceptual learning. Contrary to our predictions, conceptual pretests did not enhance learning of repeated conceptual questions. This was true relative to not taking a pretest (Experiment 1) or relative to novel conceptual questions (Experiments 1 and 2a).

Participants learned less from conceptual pretests because they were significantly more likely to repeat conceptual than factual pretest errors on the final test (Experiments 1 and 2a). Experiment 2a suggested that differences in error correction cannot be attributed to differences in confidence. Confidence in a pretest error was unrelated to the probability that the error would be corrected on the final test, a finding that was true in both the factual and conceptual conditions.

A more likely explanation is that reading the passages served as less effective feedback for conceptual than factual pretests because the answers to the conceptual questions were not stated explicitly in the text and required the reader to make an inference. Consistent with the explicit feedback account, participants benefited from conceptual pretests for repeated questions when they received corrective item-by-item feedback on the pretest (Experiment 2a). However, by giving explicit feedback, the repeated conceptual questions likely became factual questions because answering them correctly no longer required making an inference. Indeed, participants answered a similar number of repeated conceptual questions correctly when they received correct-answer feedback, regardless of whether they then read the passage (Experiment 2b). Furthermore, providing feedback did not aid conceptual learning more broadly, as the benefits of feedback did not extend to novel conceptual questions (Experiment 2a).

Participants also benefited from feedback on conceptual pretests when the feedback only indicated whether an answer was correct and did not explicitly provide the answer, but the benefits of correct/incorrect feedback in Experiment 3 were smaller than the benefits of corrective-answer feedback observed in Experiment 2. The error correction data (Figure 3e) suggested that participants benefited from correct/incorrect feedback, in part, by learning which answer to avoid on the final test rather than learning the

correct answer from making inferences while reading the passage. Taken together, the present studies provide no evidence that pretesting helps readers learn underlying concepts in a way that allows them to make sound inferences beyond information explicitly stated in the text, even when feedback is provided. Future research could explore ways of implementing pretests to enhance learning of both novel factual and conceptual questions.

Theoretical Implications

Pashler et al. (2007) suggested that pretesting could enhance learning by directing readers' attention to important material in the text and by activating relevant prior knowledge. Pretesting did not increase attention to the text in general, as reading times were equivalent in the pretest and no-pretest conditions (Experiment 1). One possibility, not examined in these experiments, is that participants used their reading time differently depending on their pretest condition, perhaps by selectively attending to information referenced in the pretest questions (but see Little & Bjork, 2016). However, previous research suggests that the benefits of pretesting cannot be fully explained by selective attention to relevant information. Richland and colleagues (2009) demonstrated that bolding important information in the text yielded the same amount of learning as pretesting (see also Pressley et al., 1990). One caveat is that prior research assumed that the increase in attention from bolding the answers in the text, for example, was equivalent to the increase in attention from answering pretest questions, which may not necessarily be true. Thus, the role of increased attention cannot be ruled out entirely. Yet, if factual pretests enhance learning, even in part, by increasing attention to relevant information in the text, conceptual pretests should have had similar effect on conceptual information.

Participants could have failed to learn from conceptual pretests because they failed to notice the important information, which spanned multiple sentences or paragraphs. The present experiments demonstrate that, when the answer is not an explicit fact that can be identified in the passage, presenting the relevant text is not sufficient to engender learning. Future research could examine whether pretesting can enhance learning of conceptual information if the important information is emphasized in the text or other similar attention-directing supports are provided.

Although definitive conclusions regarding attention during reading are not possible, participants' above chance performance on the factual and conceptual pretests indicates that they activated relevant prior knowledge. Thus, even for the pretest questions that were answered incorrectly, participants had some knowledge of the general topic (e.g., evolution). According to the elaborative retrieval hypothesis, making a retrieval attempt activates prior knowledge and facilitates integrating and encoding new information (Carpenter, 2009, 2011; Carpenter & Yeung, 2017; Kornell, 2014; Roediger & Butler, 2011), even if the retrieval attempt is unsuccessful (Kornell, 2014; Kornell et al., 2009; Kornell & Vaughn, 2016; Metcalfe, 2017; Richland et al., 2009). For example, Richland et al. (2009) found robust benefits of factual pretests, even though all of participants' pretest responses were incorrect.

Although elaborative retrieval has typically been used to explain learning of simple materials, particularly word pairs, an analogous mechanism has been proposed to explain how and when retrieval practice will enhance learning from texts (Hinze et al., 2013). For

true comprehension, readers cannot only learn the text exactly as it was presented, but must incorporate their prior knowledge and make appropriate inferences to form a more coherent representation of the text (i.e., the situation model; Kintsch, 1994, 1998). Hinze and colleagues (2013) suggested that retrieval enhances learning by encouraging readers to integrate and organize their prior knowledge and the information in the text. Other activities that encourage elaboration during reading, such as elaborative interrogation (e.g., Smith et al., 2010) and self-explanation (e.g., Chi et al., 1994), are also thought to enhance learning by helping readers construct a more coherent situation model (but see Dunlosky et al., 2013). By activating prior knowledge, conceptual pretests should have similarly enhanced learning.

Contrary to our prediction based on elaboration, the benefits of pretesting were restricted to repeated factual questions, and did not extend to novel factual, repeated conceptual, or novel conceptual questions. That is, learning was limited to information in the text that was directly pretested. Given the specificity of the benefits of pretesting, these data suggest that engaging in elaboration is not sufficient to support broader learning.

Instead, the benefits of elaboration may depend on the degree to which the information activated by retrieval—including pretest errors—can be integrated with the information provided in the text (e.g., Chan, 2009). Whether such integration happens could depend on whether the text reminds the reader of their pretest answers or other information that came to mind during the pretest (e.g., van den Broek, Bohn-Gettler, Kendeou, Carlson, & White, 2011; Wahlheim & Jacoby, 2013). Indeed, scientific misconceptions are more likely to be corrected when the misconception and the correct information are activated at the same time during reading (van den Broek & Kendeou, 2008). The factual, but not conceptual, question stems and answers appeared explicitly in the passages. Thus, even if factual and conceptual pretest questions both activate related information, such elaboration may only have benefited learning of factual questions because the text provided a stronger reminder for information activated by factual than conceptual questions.

Educational Implications

The clear implication for educators is that pretests have limited benefits. Although factual pretests enhanced learning of identical questions, the benefits did not extend to novel factual questions (but see Carpenter & Toftness, 2017; Little & Bjork, 2016). However, instructors may not want to repeat pretest questions on quizzes or exams, limiting the utility of this strategy. Critically, pretests did not enhance learning of novel or repeated conceptual questions. The current experiments do not rule out benefits of conceptual pretests under some circumstances, but simply answering pretest questions before reading a passage does not engender conceptual learning. We measured conceptual learning by testing participants' abilities to make inferences beyond information provided explicitly in the text. Future research could examine whether pretests enhance conceptual learning when measured in different ways. For example, might pretests help students learn to identify or generate new examples of a given concept?

Although we found the benefits of pretesting were limited, answering test questions after reading a text, watching a lecture, or otherwise studying information can dramatically enhance learning

(Dunlosky et al., 2013; Rowland, 2014; Roediger & Butler, 2011), including in real classrooms (Jensen, McDaniel, Woodard, & Kummer, 2014; Larsen, Butler, & Roediger, 2009; McDaniel et al., 2007; McDaniel, Agarwal, Huelser, McDermott, & Roediger, 2011; McDaniel, Thomas, Agarwal, McDermott, & Roediger, 2013; McDermott, Agarwal, D'Antonio, Roediger, & McDaniel, 2014; Roediger, Agarwal, McDaniel, & McDermott, 2011). Like pretests, the benefits of tests after an initial learning phase are often limited to tested information (but see Hinze et al., 2013; Little & Bjork, 2015; Little, Bjork, Bjork, & Angello, 2012). Unlike pretests in the present studies, the benefits of tests after an initial learning phase are not limited to factual information. Practice tests can help learners make inferences based on a text they have read (Butler, 2010; Jensen et al., 2014; Johnson & Mayer, 2009; Karpicke & Blunt, 2011) and identify examples of a previously studied concept (Larsen et al., 2009; McDaniel et al., 2013). Thus, we recommend that students' limited time is better spent answering practice test questions after reading a text, rather than prior to reading, to maximize learning of both tested facts and concepts.

References

Andre, T., Mueller, C., Womack, S., Smid, K., & Tuttle, M. (1980). Adjunct application questions facilitate later application, or do they? *Journal of Educational Psychology, 72,* 533–543. http://dx.doi.org/10.1037/0022-0663.72.4.533

Agarwal, P. K., Karpicke, J. D., Kang, S. H. K., Roediger, H. L., III, & Mcdermott, K. B. (2008). Examining the testing effect with open-and closed-book tests. *Applied Cognitive Psychology, 22,* 861–876. http://dx.doi.org/10.1002/acp.1391

Bull, S. G., & Dizney, H. F. (1973). Epistemic curiosity-arousing prequestions: Their effect on long-term retention. *Journal of Educational Psychology, 65,* 45–49. http://dx.doi.org/10.1037/h0034817

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

Butler, A. C., Fazio, L. K., & Marsh, E. J. (2011). The hypercorrection effect persists over a week, but high-confidence errors return. *Psychonomic Bulletin & Review, 18,* 1238–1244. http://dx.doi.org/10.3758/s13423-011-0173-y

Butterfield, B., & Metcalfe, J. (2001). Errors committed with high confidence are hypercorrected. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27,* 1491–1494. http://dx.doi.org/10.1037/0278-7393.27.6.1491

Butterfield, B., & Metcalfe, J. (2006). The correction of errors committed with high confidence. *Metacognition and Learning, 1,* 69–84. http://dx.doi.org/10.1007/s11409-006-6894-z

Callender, A. A., & McDaniel, M. A. (2009). The limited benefits of rereading educational texts. *Contemporary Educational Psychology, 34,* 30–41. http://dx.doi.org/10.1016/j.cedpsych.2008.07.001

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,* 1563–1569. http://dx.doi.org/10.1037/a0017021

Carpenter, S. K. (2011). Semantic information activated during retrieval contributes to later retention: Support for the mediator effectiveness hypothesis of the testing effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37,* 1547–1552. http://dx.doi.org/10.1037/a0024140

Carpenter, S. K., & Toftness, A. R. (2017). The effect of prequestions on learning from video presentations. *Journal of Applied Research in Memory & Cognition, 6,* 104–109. http://dx.doi.org/10.1016/j.jarmac.2016.07.014

Carpenter, S. K., & Yeung, K. L. (2017). The role of mediator strength in learning from retrieval. *Journal of Memory and Language, 92,* 128–141. http://dx.doi.org/10.1016/j.jml.2016.06.008

Chan, J. C. K. (2009). When does retrieval induce forgetting and when does it induce facilitation? Implications for retrieval inhibition, testing effect, and text processing. *Journal of Memory and Language, 61,* 153–170. http://dx.doi.org/10.1016/j.jml.2009.04.004

Chi, M. T. H., De Leeuw, N., Chiu, M.-H., & Lavancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science, 18,* 439–477.

Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving students' learning with effective learning techniques: Promising directions from cognitive and educational psychology. *Psychological Science in the Public Interest, 14,* 4–58. http://dx.doi.org/10.1177/1529100612453266

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

Fazio, L. K., & Marsh, E. J. (2010). Correcting false memories. *Psychological Science, 21,* 801–803. http://dx.doi.org/10.1177/0956797610371341

Felker, D. B., & Dapra, R. A. (1975). Effects of question type and question placement on problem-solving ability from prose material. *Journal of Educational Psychology, 67,* 380–384. http://dx.doi.org/10.1037/h0076616

Frase, L. T. (1968). Effect of question location, pacing, and mode upon retention of prose material. *Journal of Educational Psychology, 59,* 244–249. http://dx.doi.org/10.1037/h0025947

Hamaker, C. (1986). The effects of adjunct questions on prose learning. *Review of Educational Research, 56,* 212–242. http://dx.doi.org/10.3102/00346543056002212

Hartley, J., & Davies, I. K. (1976). Preinstructional strategies: The role of pretests, behavioral objectives, overviews and advance organizers. *Review of Educational Research, 46,* 239–265. http://dx.doi.org/10.3102/00346543046002239

Hinze, S. R., Wiley, J., & Pellegrino, J. W. (2013). The importance of constructive comprehension processes in learning from tests. *Journal of Memory and Language, 69,* 151–164. http://dx.doi.org/10.1016/j.jml.2013.03.002

Jarosz, A., & Wiley, J. (2014). What are the odds? A practical guide to computing and reporting Bayes factors. *The Journal of Problem Solving, 7*(1). Advance online publication. http://dx.doi.org/10.7771/1932-6246.1167

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

Johnson, C. I., & Mayer, R. E. (2009). A Testing Effect with Multimedia Learning. *Journal of Educational Psychology, 101,* 621–629. http://dx.doi.org/10.1037/a0015183

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

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

Karpicke, J. D., Butler, A. C., & Roediger, H. L., III. (2009). Metacognitive strategies in student learning: Do students practise retrieval when

they study on their own? *Memory, 17,* 471–479. http://dx.doi.org/10.1080/09658210802647009

Karpicke, J. D., Lehman, M., & Aue, W. R. (2014). Retrieval-based learning: An episodic context account. *Psychology of Learning and Motivation, 61,* 237–284. http://dx.doi.org/10.1016/B978-0-12-800283-4.00007-1

Kintsch, W. (1994). Text comprehension, memory, and learning. *American Psychologist, 49,* 294–303. http://dx.doi.org/10.1037/0003-066X.49.4.294

Kintsch, W. (1998). *Comprehension: A Paradigm for Cognition.* New York, NY: Cambridge University Press.

Kintsch, W., & Rawson, K. A. (2005). Comprehension. In M. J. Snowling & C. Hulme (Eds.), *The science of reading: A handbook* (pp. 209–226). Malden, MA: Blackwell. http://dx.doi.org/10.1002/9780470757642.ch12

Kornell, N. (2014). Attempting to answer a meaningful question enhances subsequent learning even when feedback is delayed. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 40,* 106–114. http://dx.doi.org/10.1037/a0033699

Kornell, N., & Bjork, R. A. (2007). The promise and perils of self-regulated study. *Psychonomic Bulletin & Review, 14,* 219–224. http://dx.doi.org/10.3758/BF03194055

Kornell, N., Hays, M. J., & Bjork, R. A. (2009). Unsuccessful retrieval attempts enhance subsequent learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35,* 989–998. http://dx.doi.org/10.1037/a0015729

Kornell, N., & Vaughn, K. E. (2016). How retrieval attempts affect learning: A review and synthesis. In B. H. Ross (Ed.), *Psychology of learning and motivation* (Vol. 65, pp. 183–215). Amsterdam, the Netherlands: Elsevier.

Larsen, D. P., Butler, A. C., & Roediger III, H. L. (2009). Repeated testing improves long-term retention relative to repeated study: A randomised controlled trial. *Medical Education, 43,* 1174–1181. http://dx.doi.org/10.1111/j.1365-2923.2009.03518.x

Little, J. L., & Bjork, E. L. (2015). Optimizing multiple-choice tests as tools for learning. *Memory & Cognition, 43,* 14–26. http://dx.doi.org/10.3758/s13421-014-0452-8

Little, J. L., & Bjork, E. L. (2016). Multiple-choice pretesting potentiates learning of related information. *Memory & Cognition, 44,* 1085–1101. http://dx.doi.org/10.3758/s13421-016-0621-z

Little, J. L., Bjork, E. L., Bjork, R. A., & Angello, G. (2012). Multiple-choice tests exonerated, at least of some charges: Fostering test-induced learning and avoiding test-induced forgetting. *Psychological Science, 23,* 1337–1344. http://dx.doi.org/10.1177/0956797612443370

Marsh, E. J., Lozito, J. P., Umanath, S., Bjork, E. L., & Bjork, R. A. (2012). Using verification feedback to correct errors made on a multiple-choice test. *Memory, 20,* 645–653. http://dx.doi.org/10.1080/09658211.2012.684882

Masson, M. E. J. (2011). A tutorial on a practical Bayesian alternative to null-hypothesis significance testing. *Behavior Research Methods, 43,* 679–690. http://dx.doi.org/10.3758/s13428-010-0049-5

McDaniel, M. A., Agarwal, P. K., Huelser, B. J., McDermott, K. B., & Roediger, H. L., III. (2011). Test-enhanced learning in a middle school science classroom: The effects of quiz frequency and placement. *Journal of Educational Psychology, 103,* 399–414. http://dx.doi.org/10.1037/a0021782

McDaniel, M. A., Anderson, J. L., Derbish, M. H., & Morrisette, N. (2007). Testing the testing effect in the classroom. *The European Journal of Cognitive Psychology, 19*(4–5), 494–513. http://dx.doi.org/10.1080/09541440701326154

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

McDermott, K. B., Agarwal, P. K., D'Antonio, L., Roediger, H. L., III, & McDaniel, M. A. (2014). Both multiple-choice and short-answer quizzes enhance later exam performance in middle and high school classes. *Journal of Experimental Psychology: Applied, 20,* 3–21. http://dx.doi.org/10.1037/xap0000004

Metcalfe, J. (2017). Learning from Errors. *Annual Review of Psychology, 68,* 465–489. http://dx.doi.org/10.1146/annurev-psych-010416-044022

Morehead, K., Rhodes, M. G., & DeLozier, S. (2016). Instructor and student knowledge of study strategies. *Memory, 24,* 257–271. http://dx.doi.org/10.1080/09658211.2014.1001992

Nelson, T. O. (1984). A comparison of current measures of the accuracy of feeling-of-knowing predictions. *Psychological Bulletin, 95,* 109–133. http://dx.doi.org/10.1037/0033-2909.95.1.109

Otero, J., & Kintsch, W. (1992). Failures to detect contradictions in a text: What readers believe versus what they read. *Psychological Science, 3,* 229–235. http://dx.doi.org/10.1111/j.1467-9280.1992.tb00034.x

Pashler, H., Bain, P., Bottge, B., Graesser, A., Koedinger, K., McDaniel, M., & Metcalfe, J. (2007). *Organizing instruction and study to improve student learning (NCER 2007–2004).* Washington, DC: U.S. Department of Education, National Center for Educational Research, Institute of Education Sciences. Retrieved from http://ncer.ed.gov

Peeck, J. (1970). Effect of prequestions on delayed retention of prose material. *Journal of Educational Psychology, 61,* 241–246. http://dx.doi.org/10.1037/h0029104

Pressley, M., Tanenbaum, R., McDaniel, M. A., & Wood, E. (1990). What happens when university students try to answer prequestions that accompany textbook material? *Contemporary Educational Psychology, 15,* 27–35. http://dx.doi.org/10.1016/0361-476X(90)90003-J

Raftery, A. E. (1995). Bayesian model selection in social research. *Sociological Methodology, 25,* 111–163. http://dx.doi.org/10.2307/271063

Richland, L. E., Kornell, N., & Kao, L. S. (2009). The pretesting effect: Do unsuccessful retrieval attempts enhance learning? *Journal of Experimental Psychology: Applied, 15,* 243–257. http://dx.doi.org/10.1037/a0016496

Rickards, J. P. (1976a). Interaction of position and conceptual level of adjunct questions on immediate and delayed retention of text. *Journal of Educational Psychology, 68,* 210–217. http://dx.doi.org/10.1037/0022-0663.68.2.210

Rickards, J. P. (1976b). Type of verbatim question interspersed in text: A new look at the position effect. *Journal of Reading Behavior, 8,* 37–45. http://dx.doi.org/10.1080/10862967609547164

Rickards, J. P., Anderson, M. C., & Mccormick, C. B. (1976). Processing effects of common-word and number questions inserted in reading materials. *The Journal of Educational Research, 69,* 274–277. http://dx.doi.org/10.1080/00220671.1976.10884897

Roediger, H. L., III, Agarwal, P. K., McDaniel, M. A., & McDermott, K. B. (2011). Test-enhanced learning in the classroom: Long-term improvements from quizzing. *Journal of Experimental Psychology: Applied, 17,* 382–395. http://dx.doi.org/10.1037/a0026252

Roediger, H. L., III, & Butler, A. C. (2011). The critical role of retrieval practice in long-term retention. *Trends in Cognitive Sciences, 15,* 20–27. http://dx.doi.org/10.1016/j.tics.2010.09.003

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

Rothkopf, E. Z. (1966). Learning from written instructive materials: An exploration of the control of inspection behavior by test-like events. *American Educational Research Journal, 3,* 241–249. http://dx.doi.org/10.3102/00028312003004241

Rothkopf, E. Z., & Bisbicos, E. E. (1967). Selective facilitative effects of interspersed questions on learning from written materials. *Journal of Educational Psychology, 58,* 56–61. http://dx.doi.org/10.1037/h0024117

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review, 16,* 225–237. http://dx.doi.org/10.3758/PBR.16.2.225

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

Sagaria, S. D., & Di Vesta, F. (1978). Learner expectations induced by adjunct questions and the retrieval of intentional and incidental information. *Journal of Educational Psychology, 70,* 280–288. http://dx.doi.org/10.1037/0022-0663.70.3.280

Sagerman, N., & Mayer, R. E. (1987). Forward transfer of different reading strategies evoked by adjunct questions in science text. *Journal of Educational Psychology, 79,* 189–191. http://dx.doi.org/10.1037/0022-0663.79.2.189

Sitzman, D. M., Rhodes, M. G., & Tauber, S. K. (2014). Prior knowledge is more predictive of error correction than subjective confidence. *Memory & Cognition, 42,* 84–96. http://dx.doi.org/10.3758/s13421-013-0344-3

Smith, B. L., Holliday, W. G., & Austin, H. W. (2010). Students' comprehension of science textbooks using a question-based reading strategy. *Journal of Research in Science Teaching, 47,* 363–379. http://dx.doi.org/10.1002/tea.20378

Thiede, K. W., Wiley, J., & Griffin, T. D. (2011). Test expectancy affects metacomprehension accuracy. *British Journal of Educational Psychology, 81,* 264–273. http://dx.doi.org/10.1348/135910710X510494

van den Broek, P., Bohn-Gettler, C., Kendeou, P., Carlson, S., & White, M. J. (2011). When a reader meets a text: The role of standards of coherence in reading comprehension. In M. T. McCrudden, J. P. Magliano, & G. Schraw (Eds.), *Text relevance and learning from text,* (pp. 123–139). Charlotte, NC: IAP Information Age Publishing.

van den Broek, P., & Kendeou, P. (2008). Cognitive processes in comprehension of science texts: The role of co-activation in confronting misconceptions. *Applied Cognitive Psychology, 22,* 335–351. http://dx.doi.org/10.1002/acp.1418

Wagenmakers, E. J. (2007). A practical solution to the pervasive problems of p values. *Psychonomic Bulletin & Review, 14,* 779–804. http://dx.doi.org/10.3758/BF03194105

Wahlheim, C. N., & Jacoby, L. L. (2013). Remembering change: The critical role of recursive remindings in proactive effects of memory. *Memory & Cognition, 41,* 1–15. http://dx.doi.org/10.3758/s13421-012-0246-9

Watts, G. H., & Anderson, R. C. (1971). Effects of three types of inserted questions on learning from prose. *Journal of Educational Psychology, 62,* 387–394. http://dx.doi.org/10.1037/h0031633

Weinstein, Y., McDermott, K. B., & Roediger, H. L., III. (2010). A comparison of study strategies for passages: Rereading, answering questions, and generating questions. *Journal of Experimental Psychology: Applied, 16,* 308–316. http://dx.doi.org/10.1037/a0020992

Zwaan, R. A. (1999). Situation models: The mental leap into imagined worlds. *Current Directions in Psychological Science, 8,* 15–18. http://dx.doi.org/10.1111/1467-8721.00004

Appendix

Omnibus ANOVA of Final Test Performance in Experiment 2a

We conducted a 2 (feedback type: passage-only vs. feedback + passage) × 2 (pretest type: factual vs. conceptual) × 2 (question novelty: novel vs. repeated) mixed-factor ANOVA (Table A1). Participants performed significantly better when they received feedback after each pretest question in addition to reading the passage, significantly better following factual than conceptual pretests, and significantly better on repeated than novel question. All of the two-way interaction effects were significant, but the three-way interaction was not.

Table A1
Analysis of Variance of Final Test Performance on Feedback Type, Pretest Type, and Question Novelty

Source	df	F	p	η_p^2
Feedback type	1	38.06	<.001	.19
Pretest type	1	27.45	<.001	.15
Question novelty	1	251.98	<.001	.61
Feedback Type × Pretest type	1	4.05	.05	.03
Feedback Type × Question novelty	1	68.95	<.001	.30
Pretest Type × Question novelty	1	10.75	.001	.06
Feedback Type × Pretest Type × Question novelty	1	.57	.45	.004