AMERICAN PSYCHOLOGICAL ASSOCIATION

Journal of Educational Psychology

ISSN: 0022-0663

2021, Vol. 113, No. 7, 1281–1303
https://doi.org/10.1037/edu0000657

Expecting to Teach Affects Learning During Study of Expository Texts

Tricia A. Guerrero and Jennifer Wiley
Department of Psychology, University of Illinois at Chicago

Past research has suggested that there may be benefits in learning from expository science text when students study with the expectation that they will need to teach another student. The present experiments were designed to extend prior work by testing whether an effect would be seen on both immediate tests (similar to those used in most prior studies) as well as delayed tests (which are important for demonstrating long-term learning of material). The experiments also tested whether an effect would be seen when learning outcomes were measured using questions testing memory for the text and questions testing comprehension from the text, and the experiments explored whether effects would be seen for a text written below grade level (Experiment 1) versus at an appropriate grade level (Experiment 2). Across both experiments, results supported that expecting to teach improved learning outcomes even at a delay and improved both memory and comprehension from expository texts. These results suggest that expecting to teach may be a useful activity for supporting durable learning from expository texts.

Educational Impact and Implications Statement

Previous research has demonstrated that when students act as "teachers" and tutor their peers, it can lead to improvements in the learning of the "teachers." A more provocative claim is that just preparing to teach others (and not actually doing it) can also have a beneficial effect. The present experiments tested for benefits when students simply prepared to teach another student and found that preparing to teach (vs. preparing to test) led to better performance on test items measuring both memory for text and comprehension from text for the "teacher." Further, these results extended prior work by showing benefits for both memory and comprehension on a delayed test and with a text written at an appropriate grade level. These results provide support for the suggestion that solely providing students with the expectation that they will have to teach another student may be a useful educational activity.

Keywords: comprehension, expository text processing, learning activities, learning by teaching, memory

As a general rule, teaching is meant to impart knowledge from the teacher to the student, yet research has shown that teaching can be an effective way to improve learning for the "teacher" as well. Many studies have explored the benefits that can come from peer tutoring, cooperative learning, or reciprocal teaching activities (Annis, 1983; Cohen et al., 1982; King et al., 1998; Kobayashi, 2019; Palinscar & Brown, 1984; Roscoe & Chi, 2008; Simmons et al., 1995; Webb, 2013). Bargh and Schul (1980) outlined three main stages involved in the process of teaching: preparation to teach, presentation of material, and feedback received through answering students' questions. A provocative claim that some researchers have suggested is that solely engaging in the first of these stages (preparation to teach) can lead to benefits in learning (Bargh & Schul, 1980; Benware & Deci, 1984; Fiorella & Mayer, 2013; Nestojko et al., 2014).

The benefits of preparing to teach have been commonly tested through paradigms that provide participants with an expectation that they will have to teach the material after study. The benefits of expecting to teach have been explored for learning to solve mathematical and logical problems, as well as with procedural learning (Daou et al., 2016; Hoogerheide et al., 2014; Renkl, 1995). They have also been explored for learning from text which is the focus of the current research. In these experiments, the learning material usually consists of a textbook passage or an expository text, and participants expect to either take a test on the content or to teach it after they study (Bargh & Schul, 1980; Benware & Deci, 1984; Fiorella & Mayer, 2013; Nestojko et al.,

This article was published Online First February 15, 2021.

Tricia A. Guerrero https://orcid.org/0000-0002-0549-0980
Jennifer Wiley https://orcid.org/0000-0002-2590-7392

Portions of this work were completed as part of the master's thesis of Tricia A. Guerrero. The authors thank the committee members, Thomas Griffin, M. Anne Britt, and Karl Szpunar, for their suggestions and support. Additionally, the authors wish to thank Gabriella Lazinek, Grace Li, Bella Garcia, and Erum Baig for their assistance in data collection and coding. Preparation of this research was supported by the Institute of Education Sciences, U.S. Department of Education (Grant R305A160008) and by the National Science Foundation Graduate Research Fellowship Program. The opinions expressed are those of the authors and do not represent views of the Institute of Education Sciences, the U.S. Department of Education, or the National Science Foundation.

Correspondence concerning this article should be addressed to Tricia A. Guerrero, Department of Psychology, University of Illinois at Chicago, 1007 West Harrison Street, M/C 285, Chicago, IL 60607, United States. Email: tguerr9@uic.edu

2014). These prior experiments on learning from text have differed along several dimensions including the difficulty of the texts that are assigned, the types of test questions that are given, and the delay between study and test. Further, prior experiments have differed in their results, and a benefit from simply expecting to teach has not always been shown in learning outcomes. The goal for the current research was to test whether expecting to teach improves learning with both simpler (below grade level) and more difficult (at appropriate grade level) materials and to test whether benefits can be seen on questions that test memory for text and comprehension from text.

Prior Experiments on the Effects of Expecting to Teach on Learning From Text

In an early experiment on the effects of expecting to teach, Bargh and Schul (1980) tested for benefits on learning from text. Although many details about the nature of the texts and tests that were given are not reported in the original article, this experiment provided an initial suggestion that simply expecting to teach might improve learning. Experiment 1 manipulated whether participants were asked to learn a passage because "they would be teaching the contents of the passage to another subject without being able to refer to a copy of the passage" versus being asked to learn because they would answer multiple-choice questions about the text. Participants were given 15 min to study a 660-word text. Immediately after reading the text, they completed a test containing eight recall and eight recognition questions that claimed to assess both main ideas and details. In addition, before studying the main text, all participants completed an initial reading activity where they read a very short text (170 words, no topic or other information about the passage was provided) with the expectation that they would answer multiple-choice questions about it. Participants were tested on this initial passage using four recall and four recognition questions. Performance on this initial reading task served as a covariate in the main analysis. The main result from this experiment was that participants who were expecting to teach had higher performance on the main learning task when controlling for scores on the initial reading task (and also entering gender in the model). Although it is difficult to make any detailed conclusions about this finding due to the lack of information provided about the materials used, this result provided an initial suggestion that simply expecting to teach might improve learning compared with a condition in which participants were expecting to be tested on the material. Although the authors acknowledged the exploratory nature of this experiment, they suggested that the performance benefits seen for those expecting to teach may be attributable to differences during study, either by an increase in motivation or by prompting the "teacher" to build a more organized cognitive structure that facilitates memory and understanding.

In a follow-up experiment, Benware and Deci (1984) used a similar manipulation with a 25-page journal article on brain functioning (Flesch Kincaid Grade Level, 12.9). The students in the control group were told that they would be examined on the material, whereas the students in the experimental group were told that they would teach the contents of the article to another student. Students were instructed to read the article at home during a 2-week break, and learning was assessed when they returned to the laboratory using a 24-item test. The authors reported that the test consisted of half rote memory and half conceptual questions. The two groups did not differ in performance on rote memory questions. However, those who were expecting to teach scored better than those who were expecting to test on the questions that the authors categorized as requiring conceptual understanding. Again, as in Bargh and Schul (1980), no examples or details about the assessment items were provided, so it is not clear exactly what kinds of questions were used to test conceptual understanding or whether the answers were explicitly available in the text. It is also not clear when the students actually read the article in relation to the timing of the test. However, the results of this experiment provided another early suggestion that expecting to teach could improve learning over expecting to take a test and that benefits might be seen in comprehension of the text. Because the students who expected to teach also expressed more interest, enjoyment, and willingness to continue in the experiment, the authors attributed this to these students being more intrinsically motivated.

More recently, Nestojko et al. (2014) also showed that giving readers an expectation to teach prior to engaging with the material could improve learning across two experiments. In Experiment 1, participants were split into two conditions. Those expecting to teach were told that "sometime after you are finished reading, another participant will arrive. You will be asked to teach the information you read in the passage to that participant. You will not have access to the passage while you teach. Afterward, that participant will take a test based on the passage you teach them." Those expecting to test were told that "you will later be given a test on the material." They were then given 10 min to read a 1,541-word text about the historical inaccuracies found in a movie about the Crimean War (Flesch Kincaid Grade Level, 11.5; text and test questions from Rawson & Kintsch, 2005). After the reading phase, participants completed a 25-min distractor task that required them to remember word lists. Participants then completed two tests related to the text they read. The first test was a free recall activity in which participants were asked to recall as much information from the passage as they could. The second test was a short-answer test which asked them to retrieve important and unimportant details from the passage. Analyses revealed that participants who were expecting to teach had better performance on both assessments than those who were expecting to test.

A second experiment used a 1,300-word passage from a university-level psychobiology textbook about the brain (Flesch Kincaid Grade Level, 12.5). As in the first experiment, prior to reading participants were either told to expect to take a test or to expect to teach the material. They were then given 9 min to read the passage followed by a similar distractor task. However, instead of recall and short-answer tests, they were asked to complete a fill-in-the-blank test where participants were given a sentence with a missing word and asked to recall the verbatim word from the text they read. Half of the blanks referred to a main point, and half referred to a detail. Those expecting to teach showed better performance on questions relating to main points on the fill-in-the-blank test compared with participants who were expecting to test (with prior knowledge as a covariate). No differences emerged between groups for questions associated with details. These authors preferred the selection and organizing explanation for these effects. They posited that this benefit from expecting to teach was attributable to participants using strategies similar to those of teachers, specifically selecting and organizing important information from the text during study. Results of both experiments showed that overall participants were better able to recall important information when they were expecting to teach.

In contrast to the short delay used in the Nestojko et al. (2014) experiments, Fiorella and Mayer (2013) examined whether expecting to teach improves learning with both an immediate test (Experiment 1) and a test given after a 1-week delay (Experiment 2). Those who were expecting to teach were told "you will study a short lesson on how the Doppler Effect works and then be asked to teach the material that you learned. Specifically, you will be expected to provide a short lecture explaining how the Doppler Effect works as if you were teaching the material to someone else." Those who were expecting to test were told "you will study a short lesson on how the Doppler Effect works and then be asked to answer some questions about what you learned." After receiving the expectation, participants were given 10-min to study a 533-word expository science text on the Doppler Effect (Flesch Kincaid Grade Level, 7.9) which described the characteristics of sound waves.

Although the open-ended tests in the Fiorella and Mayer (2013) experiments included questions designed to assess both memory for important ideas stated in the text as well as questions designed to assess comprehension of the text (testing construction of a coherent mental model of the content or application to new contexts), only analyses with the items collapsed into a single score are reported. In Experiment 1 (immediate testing), analyses showed that those who were expecting to teach significantly outperformed the control group on the test. Similar to Nestojko et al. (2014), the authors suggested that expecting to teach prompted participants to attend to the most important information in the text and organize it into a coherent mental representation during study.

However, in Experiment 2 (1-week delay between study and testing) solely expecting to teach did not produce any benefits above the control group on the test. These results suggest that expecting to teach another student may improve performance on test questions in the short-term, but it is not clear whether this benefit will be robust over time. Of course, the goal of instruction is to establish long-lasting knowledge representations that result in durable learning of the material. Whereas immediate tests can show whether students are able to recall and use information from a lesson a short time after study, only delayed tests can show what information has been learned in the sense that it has been incorporated into long-term memory, is retained over time, and is available for use in new contexts in the future (e.g., Johnson & Mayer, 2009; Rawson & Kintsch, 2005; Roediger & Karpicke, 2006; Salomon & Perkins, 1989; Ziegler & Stern, 2014).

Differences in the timing of the test may be one reason why some prior experiments found benefits from expecting to teach whereas Fiorella and Mayer (Experiment 2) did not. Thus, both immediate and delayed tests are used in the present experiments to help clarify where the benefits may be seen. Although there were many other differences in methods and stimuli across prior experiments, one other main question that is tested in the present experiments is whether the effects of expecting to teach depend on whether learning from text is assessed using measures of memory for text or comprehension from text. A broader framework from the literature on comprehension and learning from text provides motivation for this distinction.

Measuring Memory for Text Versus Comprehension From Text

In the text comprehension literature, many researchers make a distinction between memory for text and comprehension from text (Kintsch, 1994; Mayer, 1989; Wiley & Voss, 1999). Theories of text comprehension describe how the text can be represented at various levels of understanding. According to the Construction-Integration Model (Kintsch, 1994), at the surface level, the text is represented verbatim. At the textbase level, the information from the text is processed into meaningful semantic units represented as propositions. Representation at both of these levels is limited to information explicitly found in the text. In contrast, at the situation model level, the meaning of the text is constructed by creating a model of how or why the process, phenomenon, or situation described by the text happens. This occurs by recognizing connections between ideas, through the generation of inferences between propositions, and by integration with the reader's prior knowledge. This level of representation of the meaning of the text is also referred to as a reader's mental model (Mayer, 1989), and for many science texts it will represent a causal model of the phenomenon or system being described (Graesser & Bertus, 1998; Wiley & Voss, 1999).

Some expository texts require less inference generation to build this model than others because of the nature of the text itself. This can occur when information that the reader is required to understand, such as relations between concepts, is explicitly stated in the text. In these texts, the textbase and the situation model levels are more similar and overlapping in how the meaning of the text is represented (McNamara et al., 1996; Tapiero & Otero, 2002; Wiley et al., 2005). In contrast, other expository texts require the reader to engage in more effortful processing to generate a model from relations that are left implicit in the text. In this case, the reader must actively construct meaning by generating inferences to fully comprehend the text. Hence, in these texts, the situation model is further detached from the textbase.

When test questions ask readers about ideas that have been explicitly stated in a text, they serve as measures of memory for text. However, when relations between ideas are left implicit, then solely having memory for the text would not capture the full message of a text. In the case of more difficult expository science texts, to understand the causal model underlying phenomena and to construct a coherent mental model, the reader will have to generate inferences that are implicit in the text. It is the construction of this mental model or situation model that best represents comprehension from difficult expository texts (Kintsch, 1994; Mayer, 1989; Wiley et al., 2005), and comprehension can be specifically measured by asking questions that target the implicit inferences that must be generated by the reader.

Given the distinction between measures of memory for text and comprehension from text, both the difficulty level of the texts and whether the test questions are testing the situation model will alter what type of learning is being assessed. And this could determine whether benefits are seen from expecting to teach. If the assigned text is short or very explicit, then it is possible that having a good memory for the text will be sufficient to ensure good performance on both questions that target the textbase and situation-model representations. If the assigned text is more difficult (with less overlap between the textbase and situation model), then performing well on questions that target the situation model will require more than just memory for the information explicitly stated in the text. In this case, questions that assess the quality of the reader's situation model will provide a more sensitive test of whether there are comprehension benefits.

Finally, as noted above, if the goal is to measure learning, then it is important to show benefits in performance at a delay. When a participant completes a test immediately after reading, their memory for surface features remains highly active. However, after a delay surface memory experiences decay, whereas the situation model representation remains robust to decay (Kintsch et al., 1990). Thus, improvements in performance on situation-model questions and on delayed tests would provide the strongest evidence that expecting to teach is improving comprehension from text. Viewed from this framework, because Bargh and Schul (1980) and Nestojko et al. (2014) used relatively immediate recall, recognition, and fill-in-the-blank tests which assessed participants' ability to retrieve information explicitly presented in the text, these results may be best interpreted as showing that expecting to teach is affecting memory for the text. They do not provide clear evidence that expecting to teach is improving comprehension. Further, the lack of a significant effect on delayed tests which included some comprehension items (Fiorella & Mayer, 2013 Experiment 2) also failed to provide any support that expecting to teach improved comprehension outcomes. Thus, the present experiments were designed to provide a strong test of whether expecting to teach yields benefits in comprehension by exploring if an effect would be seen on delayed tests, on questions that address the situation model, and also in the context of text written at an appropriate grade level.

The Current Experiments

The current experiments tested whether expecting to teach improves learning with both simpler (below grade level) and more difficult (at appropriate grade level) materials. An expository text used in prior research (written at the 7th–8th grade level, Fiorella & Mayer, 2013) served as the stimuli in Experiment 1, and a new text on Fermentation (written at a 10th–11th grade level) served as the stimuli in Experiment 2. The main between-subjects manipulation in both experiments was whether students expected to take a test or to teach following study. An orthogonal between-subjects manipulation was the delay between study and test. Half the sample was tested immediately and half after a 1-week delay.

The main hypotheses tested in these experiments were as follows:

1. Expecting to teach will benefit learning from text compared to expecting to test.

2. Expecting to teach will benefit memory for text compared to expecting to test.

3. Expecting to teach will benefit comprehension of text compared to expecting to test.

The first hypothesis was tested by analyzing whether test performance was better for those expecting to teach than those expecting to test. This prediction represents the main/general claim made by prior research. To ensure that effects of the expectancy manipulation reflect learning that is durable over time, it is important to see benefits in test performance for those expecting to teach compared with those expecting to test, especially at the delayed timepoint (when contents are no longer active in recent memory).

The second and third hypotheses were tested by an additional within-subjects manipulation which had participants complete measures of memory (text-based questions) as well as comprehension (inference questions based in the situation model). If expecting to teach improves memory for a text, this should result in better test performance on text-based items. This prediction would be consistent with results of prior research. In contrast, given the lack of clarity in the previous research, the most tentative of the three hypotheses was that expecting to teach would improve comprehension from text. If expecting to teach does improve comprehension, this should result in better test performance on inference test questions. Additionally, seeing better performance on inference questions that test the situation model at a delay would provide clear evidence that expecting to teach can result in long-term benefits in student understanding.

Assuming that expecting to teach might be found to improve learning, a more exploratory goal of the research was to attempt to understand how expecting to teach might change the learning process. It has been proposed that a possible mechanism by which expecting to teach leads to benefits in learning is because it causes participants to better select and organize information from the text during study (Bargh & Schul, 1980; Fiorella & Mayer, 2013; Nestojko et al., 2014). Alternatively, Benware and Deci (1984) primarily attributed the benefits to differences in student motivation, engagement, and effort. To explore these possible mechanisms, self-report ratings were collected and participants' notes were analyzed as a trace measure of how students engaged with the text during study. If benefits are seen in test performance as a result of expecting to teach, then measures derived from the ratings or notes may help to test exploratory research questions of whether these benefits can be shown to be attributable to differences in the learning process or motivation during study.

Experiment 1

The current experiment sought to replicate the prior work of Fiorella and Mayer (2013) by testing whether expecting to teach would lead to benefits in learning from a simple expository text (written below the grade level of participants). As in the prior work, participants were tested either immediately or after a 1-week delay. Second, the current experiment sought to extend prior work by testing whether benefits in test performance were seen in both measures of memory and comprehension. Additionally, participants' notes taken during study were analyzed to explore how expecting to teach may be affecting the learning process. Alternative explanations for the benefits of expecting to teach including arousal and motivation were also explored by collecting self-report ratings, and the relationship between changes in the learning process and the main learning outcome was examined.

Method

Participants

The sample consisted of 206 undergraduates (141 females, $M_{age} = 18.90$, $SD_{age} = 1.29$). Self-reported racial composition was 27% Hispanic, 19% White, 36% Asian, and 6% Black. A power calculation for the ANCOVA that was used to analyze the main dependent variable (test performance) was based in the average effect size (Cohen's *d: M* = .46) observed in Fiorella and Mayer (2013) and Nestojko et al. (2014). It suggested that the obtained sample size exceeded the minimum of 151 subjects needed to provide an 80% chance of detecting benefits attributable to expecting to teach.

Participants were randomly assigned to an expectancy condition (expect to test or expect to teach) and a time of test condition (immediate or delayed testing), resulting in four groups of participants. Equivalence of the groups in terms of reader characteristics was explored using measures of reading ability (ACT score) and prior knowledge. Descriptive and inferential statistics are shown in Table 1. As shown in Table 1, ACT scores and prior knowledge were not significantly different across groups. However, to help ensure that final test scores reflected learning during the study and not preexisting differences and to be consistent with analyses conducted in past research (Bargh & Schul, 1980; Nestojko et al., 2014), both reading ability and prior knowledge were used as covariates in analyses of learning outcomes in both experiments reported in this article.

Materials

All materials are included in Appendix A and are available on the Open Science Framework (https://osf.io/cr6b7/).

Expectation Manipulation. Instructions (taken from Benware & Deci, 1984) varied dependent on whether participants were assigned to groups expecting to teach or expecting to test. As shown in Table A1, all participants were given the same initial instruction followed by a slightly different continuation. A second place where instructions to the two groups varied was after studying, but prior to taking the test. Participants who were expecting to teach were told:

> Before teaching the other student, we would like you to answer some test questions to assess your understanding. Please answer each question to the best of your ability. (Fiorella & Mayer, 2013)

Participants who were expecting to teach did not actually have an opportunity to teach another student. At the end of the experiment, participants in this condition were told:

> This was a pilot study to test the effectiveness of the instructions provided to you. You do not actually have to teach another student, but future participants will.

Arousal Ratings. Participants were asked two questions (How excited do you feel right now?; How stressed do you feel right now?) to assess their level of arousal based on the instructions given to them. Responses were assessed on a 10-point Likert scale ranging from *not*

Table 1
Means (Standard Deviations) of Reader Characteristics and Ratings for Experiment 1

	Immediate		Delayed			Statistic		
Variable	Test n = 51	Teach n = 53	Test n = 52	Teach n = 50		F	p	η_p^2
Reader characteristics								
ACT	23.16 (4.45)	24.00 (4.93)	24.69 (4.98)	24.12 (4.60)	1	0.04	.84	.00
					2	1.56	.21	.01
					Int	1.14	.29	.01
Prior knowledge	2.29 (1.17)	2.60 (1.57)	2.67 (1.53)	2.72 (1.43)	1	0.79	.37	.00
					2	1.53	.22	.01
					Int	0.43	.51	.00
Arousal measures								
Excited	4.47 (2.09)	5.06 (2.26)	5.23 (2.22)	5.76 (2.10)	1	3.40	.07	.02
					2	5.86	.02	.03
					Int	0.01	.93	.00
Stressed	4.47 (2.43)	5.51 (2.78)	4.62 (2.49)	5.02 (2.60)	1	4.03	.05	.02
					2	0.23	.63	.00
					Int	0.78	.38	.00
Study ratings								
Interesting	6.78 (2.02)	7.74 (1.82)	7.10 (1.84)	7.46 (1.86)	1	6.25	.01	.03
					2	0.00	.95	.00
					Int	1.25	.27	.01
Enjoyable	6.37 (2.01)	7.30 (1.99)	6.69 (2.01)	7.06 (2.15)	1	5.21	.02	.03
					2	0.02	.89	.00
					Int	0.98	.32	.01
Learn more	6.25 (1.91)	6.92 (2.46)	6.19 (2.10)	6.76 (2.33)	1	4.03	.05	.02
					2	0.14	.71	.00
					Int	0.03	.87	.00
Effort	5.57 (2.12)	6.55 (2.19)	6.23 (2.37)	6.92 (2.22)	1	7.21	.01	.03
					2	2.78	.10	.01
					Int	0.22	.64	.00
Difficulty	2.69 (1.61)	2.96 (1.75)	2.21 (1.60)	2.70 (1.59)	1	2.79	.10	.01
					2	2.60	.11	.01
					Int	0.22	.64	.00
Understanding	7.47 (1.86)	7.77 (1.87)	8.06 (1.72)	7.58 (2.24)	1	0.11	.75	.00
					2	0.54	.46	.00
					Int	2.11	.15	.01

Note. Statistics are calculated using a two-factor ANOVA. The main effects of expectation (1), time of test (2), and interaction (Int) are shown.

at all . . . (excited/stressed) to *very . . . (excited/ stressed)*. Descriptive statistics for each condition are shown in Table 1.

Expository Text. Participants read a 533-word expository text (Flesch Kincaid Grade Level, 7.9), taken from Fiorella and Mayer (2013, 2014), which described the Doppler Effect. Using familiar, real-life examples, it described how wavelength, wave frequency, and pitch are perceived by an observer and discussed the relationship between the characteristics of sound waves. This text contained 44 idea units along with five images and described a causal model of how movement affects the perception of sound waves.

Three experts were asked to identify the subset of idea units most important to developing an understanding of the Doppler Effect. Any idea unit identified as important by two or more experts was coded as more important (25 of 44). These idea units were assumed to contribute the most to the development of a mental model of the text. The remaining 19 idea units were designated as less important (but not irrelevant) for understanding the Doppler Effect.

Study Ratings. In an attempt to replicate past findings and to be consistent with the survey questions used in past research, participants were asked the following questions, adapted from Fiorella and Mayer (2013) and Benware and Deci (1984), to assess their level of effort, interest and motivation during study on a 10-point Likert scale from *strongly disagree* to *strongly agree*:

I found this text to be interesting.

I found this experiment enjoyable.

I would be interested in learning more about this topic.

I felt the subject matter was difficult.

I feel like I have a good understanding of this topic.

I invested a lot of effort into this topic.

Descriptive and inferential statistics for each condition are shown in Table 1.

Multiple-Choice Test Questions. As shown in Table A2, two categories of questions were used: questions that test memory for a text, *text-based questions*, and those that assess comprehension of a text, *inference questions* (Wiley & Guerrero, 2018). Regardless of question type, each item related to at least one of the idea units categorized as most important to developing an understanding of the Doppler Effect by the experts. Collectively, the questions tested all sections of the text.

Five text-based questions tested memory for surface-level features of the text and involved recognition of facts and details explicitly mentioned in the text. This mention could have been verbatim or paraphrased and required minimal processing as a result of the question stem and answer being located within the same sentence (Ozuru et al., 2007).

Five inference questions were used to assess the reader's situation model or mental model of the scientific process. These questions tested for the generation of implicit bridging inferences (which reflected the implied relationship between multiple pieces of information provided across sentences the text) as well as elaborative inferences (which required application of the text to novel or hypothetical situations). Although the answers to these questions were not explicitly stated in the text, the supporting information needed to construct each inference was. Proportion correct was calculated by computing the average score on each of the text-based and inference tests separately.

To validate the distinction between these question types, an independent set of participants ($n = 20$) were asked to read the text and sort the test questions into text-based and inference categories. Participants were presented with instructions that explained that text-based questions "can be found directly in the text by exactly matching or almost exactly matching something stated in a sentence" and inference questions "can be inferred by connecting or combining multiple sentences that could even be in separate paragraphs (and may require applying the ideas to a new situation)." They then received a sample text and example questions with correct answers prior to receiving the Doppler Effect materials. Participant ratings demonstrated consensus on the distinction between the question types as the raters showed a high level of agreement with the intended categories (overall agreement 89%, agreement for inference items 92% [range 75–100%], agreement for text-based items 85% [range 75–95%]).

To show that the questions could not be answered from common knowledge, a separate group of participants ($n = 20$) completed all multiple-choice questions without reading the text. Accuracy rates were 27% overall. This shows that in general there is low prior knowledge about the Doppler Effect and that students did not perform significantly different than chance (25%) on these questions without reading the text, $t(19) = 0.73$, $p = .47$, $d = .16$.

Finally, an additional group ($n = 20$) completed the test questions while the text was available for them to reference. Performance on test questions was high overall (71%) with similar performance on both text-based ($M = 76\%$, $SD = 14\%$) and inference questions ($M = 66\%$, $SD = 23\%$), $t(19) = 1.75$, $p = .10$, $d = .39$, suggesting substantial overlap in the textbase and situation model representation of the Doppler Effect text.

Free Recall. A free recall response was also used to assess learning. Participants were told to "type as much information from the passage as you can recall" (Nestojko et al., 2014).

Final Survey and Manipulation Check. To assess whether participants remembered that they were expected to teach another student, the first question of the final survey asked: "What were you told that you would be expected to do after studying the text?" Of the participants who were expecting to teach, eight additional participants who did not recall that they would have to teach another student were excluded from analyses, leaving only the 206 participants described in the Participants section.

Additionally, the prior knowledge measures were also taken from the prior work this experiment was attempting to replicate (Fiorella & Mayer, 2013). All participants were asked to rate their knowledge of the Doppler Effect on a 6-point Likert scale from *very low* to *very high*.[1] Participants were asked to report

[1] The full self-report prior knowledge questionnaire, taken from Fiorella and Mayer (2013), was collected, which included both the prior knowledge rating item in addition to several judgments of knowing for relevant concepts. Because of concerns from reviewers about validity and reliability, the judgments of knowing items were removed from the prior knowledge scores for the current analyses. The overall pattern of results did not change whether or not the judgments of knowing were included in the scores.

on their prior knowledge and reading ability (ACT and SAT scores) during the final survey to avoid any priming or stereotype threat that might be caused by answering these questions immediately before the experiment. SAT scores were converted to the ACT scale using The College Board conversion table (The College Board, 2018).

Participants were also asked the same prior knowledge questions during an initial screening survey administered during the first week of the introduction to psychology course. Prior knowledge responses obtained at initial screening (n = 175) were highly correlated with prior knowledge responses given during the final survey, $r = .87$, $p < .001$. Because complete data were only available for the final survey, these scores were used for analyses.

As a final question, participants who were expecting to teach were asked, "Based on the instructions provided to you at the beginning of the study, rate the degree to which you believed that you would actually have to teach another student." Responses were assessed on a 6-point Likert scale ranging from *did not believe that I would have to teach another student* to *definitely believed I would have to teach another student*. A one-sample t test (against the midpoint) indicated that participants did believe that they would have to teach another student, $M = 4.58$, $SD = 1.57$, $t(102) = 7.00$, $p < .001$, $d = 0.69$.

Procedure

At the beginning of the session, all participants received an expectation to teach or test instructions and then completed the arousal ratings prior to viewing the text. (These instructions varied only as a result of expectancy condition and did not vary as a result of time of test condition.) Participants then had 10 min to study the text which was presented on paper along with a blank sheet to be used for note-taking. During this time, they were told that they may study the text using any manner they wished. Following the study period, the text and notes were collected, and participants completed the study ratings.

Participants in the immediate conditions then moved to the test phase. Participants who were expecting to teach were told that their understanding of the text would be assessed prior to teaching. All participants then completed the multiple-choice tests, without access to the text, which was presented on the Qualtrics survey platform. The program required a response to each test question prior to moving on to the next section. No feedback was provided during the experiment. Following the test questions, participants completed a free recall response and then the final survey. Participants who were expecting to teach completed a final question regarding their belief in the original instruction they were given and were excused.

Participants in the delayed conditions had a 1-week delay between Session 1 and 2. They were excused from Session 1 after completing the study ratings. (This was the first point in the procedure that it became certain that they would not be teaching or testing at all during the first session. This was after they had already prepared for teaching or testing.) They began Session 2 by completing the multiple-choice test and continued with the remainder of the experiment in the same order as the immediate conditions.

Coding

Notes. The notes taken by participants were coded by two independent raters. First, participants' notes were coded for the quantity of idea units (out of 44) they contained from the text with high interrater reliability, ICC (3, 2) = .99. From this coding the proportions of more-important (out of 25) and less-important (out of 19) idea units included in the notes were computed. These measures were used to test whether students selected the more-important idea units in their notes. (Results are similar if raw numbers are used instead of proportions of total possible idea units in each importance category.)

Idea unit coding was also used to test whether students engaged in reorganization of information from the text as they wrote their notes. A "restructuring" score was computed using a Spearman's rho rank-order correlation between the original order of ideas in the text and the order that ideas were mentioned in each student's notes (with a correlation of 1 representing that all ideas included in a student's notes followed the exact order of the text and correlations less than 1 representing a mismatch between the order of idea units in the notes and the original text). This value was then subtracted from 1 so that higher scores would reflect greater reorganization or restructuring.

Finally, in an attempt to evaluate the degree to which participants engaged in constructive processing during study, the notes were coded for the presence of any idea units that represented new information. This could have included idea units that brought in prior knowledge or the development of relationships between concepts that were not explicit in the text. Notes were coded for the proportion of new idea units out of the total number of idea units in the notes. However, only two participants included any new ideas in their notes. (They discussed sine waves using more specific vocabulary than was used in the text.) Due to the low incidence of new ideas introduced in the notes, no analyses were conducted on this measure.

Free Recall. Recall scores were calculated by coding for the number of idea units contained within the free recall responses of the total 44 idea units in the text. These responses were coded with high interrater reliability, ICC (3, 2) = .90.

Results

Analyses on the multiple-choice test scores are reported first. A significant main effect of expectation condition would provide support for the first hypothesis that expecting to teach benefits learning, especially if it occurs after a delay between study and test. If an effect of expectation condition is seen for text-based items, then this would provide support for the second hypothesis that expecting to teach benefits memory processes. If an effect of expectation condition is seen for inference items, then this would provide support for the third hypothesis that expecting to teach benefits comprehension processes. Finally, analyses of the effect of expecting to teach are performed on the recall responses, and exploratory analyses are conducted on measures obtained during the learning process. The analysis scripts and data are available on the Open Science Framework (https://osf.io/cr6b7/).

Table 2
Correlations Between Covariates and Dependent Variables

Variable	Prior knowledge	Test performance	Free recall	Important notes[a]	Motivation[b]
Experiment 1					
ACT	.27***	.33***	.26***	.15*	−.09
Prior knowledge	—	.26***	.12	.00	.14*
Test performance		—	.44***	.12+	.14*
Free recall			—	.21**	.10
Important notes[a]					.11
Experiment 2					
ACT	.21***	.49***	.38***	.15*	.06
Prior knowledge	—	.34***	.26***	.06	.37***
Test performance		—	.57***	.13+	.16*
Free recall			—	.25***	.12
Important notes[a]					.14*

[a] Proportion of more-important idea units included in the notes taken during study. [b] Motivation factor score.
$^{+}p < .10$. $^{*}p < .05$. $^{**}p < .01$. $^{***}p < .001$.

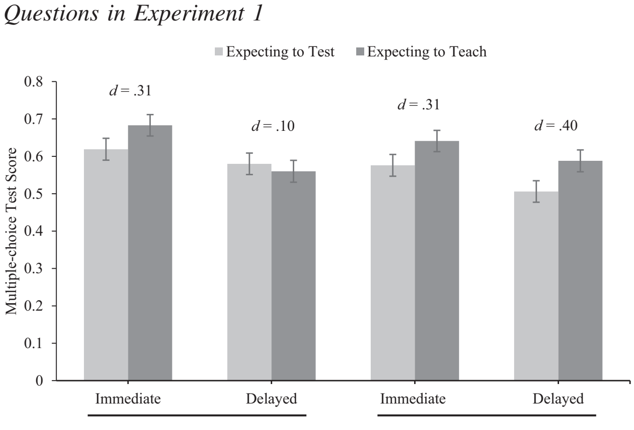
How Expecting to Teach Affects Learning Outcomes

Effects of Expecting to Teach on Multiple-Choice Test Performance. A 2 (Expectation: Teach, Test) × 2 (Time of Test: Immediate, Delayed) × 2 (Question Type: Text-based, Inference) mixed-design analysis of covariance (ANCOVA) controlling for ACT score and prior knowledge was conducted for performance on the test questions.[2] Initial tests indicated no violations of the assumptions of normality, Levene's test of homogeneity of variance, and linearity. Homogeneity of regression slopes was examined by testing interactions of the independent variables with the covariates; none were significant (*p*s > .28). Therefore, the assumption of homogeneity of regression slopes was also satisfied. Intercorrelations among covariates and the dependent variables are presented in Table 2.

As shown in Figure 1, a main effect of expectation indicated that those expecting to teach had higher test performance than those expecting to test, $F(1, 200) = 4.94$, $p = .03$, $\eta_p^2 = .02$. A main effect of time of test showed that participants had higher test scores when tested immediately than at a delay, $F(1, 200) = 10.97$, $p = .001$, $\eta_p^2 = .05$. Further, there was no significant interaction between these two factors, $F(1, 200) = 0.60$, $p = .44$, $\eta_p^2 = .00$, meaning the effect was found both immediately and at a delay. Thus, the first hypothesis that expecting to teach would improve learning (even when tested at a delay) was supported.

Additionally, test scores did not differ by question type, $F(1, 200) = 2.81$, $p = .10$, $\eta_p^2 = .01$. The expectation by question type interaction was not significant, $F(1, 200) = 1.81$, $p = .18$, $\eta_p^2 = .01$, nor was the time of test by question type interaction, $F(1, 200) = 0.26$, $p = .61$, $\eta_p^2 = .00$. The three-way interaction also failed to reach significance, $F(1, 200) = 1.66$, $p = .20$, $\eta_p^2 = .01$. These results suggest that the effect of expectation was seen regardless of question type (and delay). Thus, the hypotheses that expecting to teach would improve both memory and comprehension processes were supported.

Figure 1
Performance on Text-Based and Inference Multiple-Choice Test Questions in Experiment 1



Note. Multiple-choice test scores are shown for expectancy and time of test conditions. Error bars represent ± 1 standard error.

Effects of Expecting to Teach on Free Recall. Because of the positive skew and overdispersion that is commonly found in count variables, negative binomial regression was used to analyze these data (Coxe et al., 2009; Gardner et al., 1995).[3] ACT score (centered) and prior knowledge (centered) were entered as covariates along with expectation and time of test as dummy-coded between-subjects factors.

A likelihood ratio test indicated that the predictors accounted for a significant amount of the variance, $\chi^2(4) = 59.01$, $p < .001$. As shown in Table 3, those expecting to teach recalled significantly more idea units than those expecting to test. In addition, those who were tested immediately recalled significantly more idea units than those tested at a delay.

When the interaction between expectation and time of test was included it was not significant ($p = .94$), and it did not improve the

[2] Analyses are presented without the prior knowledge covariate in Appendix C.

[3] Following the suggestions of Coxe et al. (2009), a Poisson regression model was determined to be inappropriate because of the overdispersion scaling parameter (φ) far exceeding 1 (2.34). In this circumstance, Coxe et al. recommend using negative binomial regression. The negative binomial regression provided significantly better model fit over the standard Poisson regression, $\chi^2(1) = 103.16$, $p < .001$, and was thus utilized for the analyses on this measure.

fit of the model; hence, it was not included in the final model. As in the previous analysis, similar effects were found both immediately and at a delay. Because a benefit was seen even when recall was delayed, this result is consistent with the test results, and provides converging evidence that expecting to teach is improving learning.

Summary for Effects of Expectations on Learning Measures. The main effects of expectation, and the fact that an effect of expectation was seen even on delayed tests, provide evidence that expecting to teach is improving learning. Further, the fact that the effects were seen on both text-based and inference test items provides evidence that expecting to teach led to better memory and comprehension for this text. Given these findings, the next goal for analyses was to explore how expecting to teach might be changing the learning process and whether differences in selectivity, organization, or motivation may help to explain the benefits.

How Expecting to Teach Affects the Learning Process

Effects of Expecting to Teach on Note-Taking Content. A 2 (Expectation) × 2 (Time of Test) × 2 (Importance Level: More-important, Less-important) mixed-design ANOVA was conducted on the proportion of idea units included in the notes (out of the total possible). As shown in Figure 2, there was a main effect of expectation, $F(1, 202) = 8.33$, $p = .004$, $\eta_p^2 = .04$, in which those expecting to teach included a greater proportion of the idea units in their notes than those expecting to test. There was a main effect of time of test, $F(1, 202) = 5.17$, $p = .02$, $\eta_p^2 = .02$, in which all participants took more extensive notes in the delayed condition. However, there was no interaction between expectation and time of test, $F(1, 202) = 0.80$, $p = .37$, $\eta_p^2 = .00$.

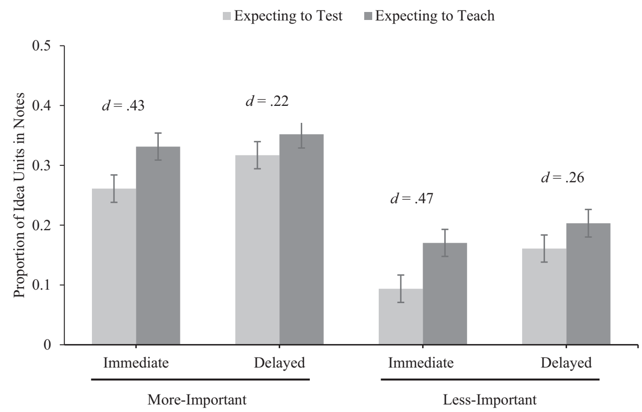
There was also a main effect for importance level, as all participants included a greater proportion of the more-important ideas in their notes than less-important ideas, $F(1, 202) = 174.72$, $p < .001$, $\eta_p^2 = .46$. Neither the expectation by importance level interaction, $F(1, 202) = 0.08$, $p = .78$, $\eta_p^2 = .00$, nor the time of test by importance level interaction reached significance, $F(1, 202) = 0.24$, $p = .62$, $\eta_p^2 = .00$. The three-way interaction was also not significant, $F(1, 202) = 0.00$, $p = .99$, $\eta_p^2 = .00$. The lack of an interaction between expectancy and importance level seems inconsistent with the proposed selection mechanism that the benefits of expecting to teach are caused by participants' focusing selectively on the most important ideas in the text during study. There was no evidence that the expectation manipulation affected the relative selection of more-important over less-important ideas in the notes. Nevertheless, as shown in Table 2, when students did take notes that included a larger proportion of the more-important information, it was positively related to better learning as measured by free recall and tended to be related to better learning as measured by test performance.

Table 3
Negative Binomial Regression Analysis for Free Recall in Experiment 1

Variable	β	*SE*	*z*	CI	*p*
Intercept	1.73	0.08	20.67	[1.56, 1.89]	<.001
ACT	0.04	0.01	4.07	[0.02, 0.06]	<.001
Prior knowledge	0.05	0.03	1.45	[−0.02, 0.12]	.15
Expectation	0.31	0.10	3.19	[0.12, 0.50]	.001
Time of test	−0.61	0.10	−6.16	[−0.80, −0.41]	<.001

Note. CI = confidence interval.

Figure 2
Proportion of More- and Less-Important Idea Units Present in the Notes in Experiment 1

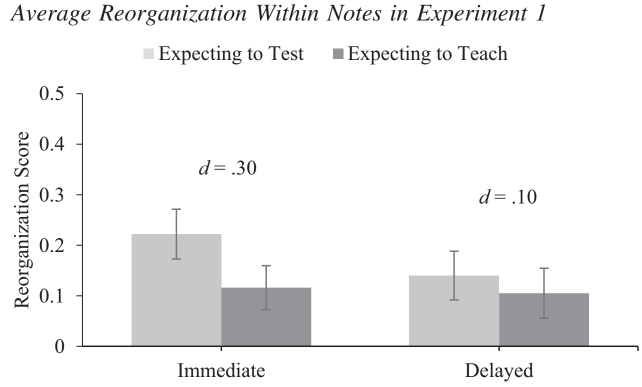


Note. Proportion of more- and less-important idea units present in the notes are shown for expectancy and time of test conditions. Error bars represent ± 1 standard error.

Effects of Expecting to Teach on the Organization of Notes. The amount of reorganization in the order of ideas from the text to the notes is shown in Figure 3. A 2 (Expectation) × 2 (Time of Test) ANOVA on the amount of reorganization seen in the notes revealed no differences across conditions. Neither the main effect for expectation, $F(1, 202) = 2.07$, $p = .15$, $\eta_p^2 = .01$, for time of test, $F(1, 202) = 0.91$, $p = .34$, $\eta_p^2 = .01$, nor the interaction were significant, $F(1, 202) = 0.52$, $p = .47$, $\eta_p^2 = .00$. In general, the notes showed very little restructuring. If anything, the nonsignificant trend showed more reorganization in the order of idea units in the condition expecting to test. This is inconsistent with the proposed organization mechanism that the benefits of expecting to teach stem from reorganizing ideas during study.

Arousal and Study Ratings. As shown in Table 1, the expectancy conditions differed on several ratings. Consistent with

Figure 3
Average Reorganization Within Notes in Experiment 1

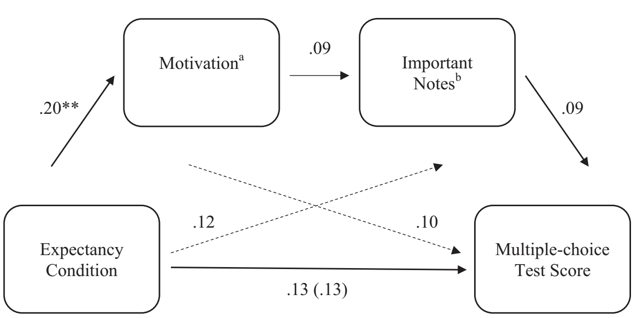


Note. Reorganization scores are shown for expectancy and time of test conditions. Error bars represent ± 1 standard error.

concerns raised by Renkl (1995), participants who were expecting to teach reported feeling more stressed in this experiment. However, those expecting to teach also reported the text to be more interesting, the experiment to be more enjoyable, a greater desire to learn more about the topic, and investing more effort into the task. Because the study ratings were highly related (Interesting, Enjoyable, Learn More, and Effort), a single factor was computed using Principal Components Analysis which accounted for 65% of the variance from the study ratings. As shown in Table 2, this motivation factor was positively related to learning outcomes. These results are consistent with the proposed role of motivation in expecting to teach effects.

In an attempt to test this proposal more formally, a sequential mediation model was tested (using Lavaan; Rosseel, 2012). The first mediator was motivation. The second mediator was the important notes measure (proportion of more-important idea units included in the notes). The independent variable used in the analysis was expectation, and the dependent variable was total performance on the multiple-choice test.[4] Although motivation and important notes were significantly related to test performance using simple correlations, when entered into the mediation model, they were no longer significantly predictive of learning. As shown in Figure 4, although the standardized regression coefficient between condition and motivation was significant, the standardized regression coefficients between motivation and test score, motivation and notes, and notes and test score were not. The significance of the indirect effect was computed, using the logic of Preacher and Hayes (2008), for each of 10,000 bootstrapped samples, and the 95% confidence interval was computed by determining the indirect effects at the 2.5th and 97.5th percentiles. The bootstrapped standardized effect was .001, and the 95% confidence interval crossed zero [−.002, .005]. Thus, the indirect effect was not significant.[5] This suggests that although expecting to teach did improve motivation, it appears motivation could not be shown to be responsible for the effects of expecting to teach on learning either via changes in selection of important ideas in the notes, or otherwise.

Figure 4
Mediation Model for Experiment 1



Note. Standard regression coefficient for the relationship between expectancy condition and multiple-choice test score as mediated by motivation and the proportion of most-important idea units in the notes. The standardized regression coefficient between expectancy condition and multiple-choice test score, controlling for motivation and notes, is in parentheses.
[a] Motivation factor score. [b] Proportion of more-important idea units included in the notes taken during study.
** $p < .01$.

Summary of Results. These results provide evidence that expecting to teach is affecting learning of an expository science text on the Doppler Effect. Importantly, expecting to teach led to better learning on a delayed test, suggesting that it resulted in durable learning over time. Further, the fact that the effects were seen on both text-based and inference test items provides evidence that expecting to teach led to better memory and comprehension for this text. Because the materials used in this experiment were written below grade-level for these college readers, the main goal for Experiment 2 was to test how expecting to teach might affect learning with a more difficult expository text, written at an appropriate grade level for college readers.

Experiment 2

Experiment 1 provided initial support that expecting to teach seems to benefit both memory and comprehension of a simple expository text. Experiment 2 tested whether those benefits would remain with materials written at an appropriate grade level for readers. Whereas simpler expository texts, such as the Doppler Effect text used in Experiment 1, may explicitly state important causal relationships directly in the text, more difficult expository texts may require the reader to generate more inferences to construct a coherent mental model from the text. It is possible that when students are asked to learn from materials that are more appropriate for their grade level (rather than written below their grade level) and that leave more relations implicit, this will require additional processing that could attenuate the benefits of expecting to teach (McNamara et al., 1996; Voss & Silfies, 1996). In contrast, if expecting to teach draws attention to these implicit inferences that need to be constructed, then expecting to teach may improve comprehension even with more difficult expository texts and could even amplify the benefits.

Method

Participants

The sample consisted of 214 undergraduates (129 females, $M_{age} = 18.90$, $SD_{age} = 1.12$). Self-reported racial composition was 38% Hispanic, 27% White, 22% Asian, and 8% Black.

As in Experiment 1, participants were randomly assigned to an expectancy condition (expect to test or expect to teach) and a time of test condition (immediate or delayed testing) resulting in 4

[4] Performance on the text-based and inference questions were correlated ($r = .20$, $p = .003$). Because of this a total score on the multiple-choice test was calculated for the dependent variable used in the mediation analysis. However, patterns remained the same even when test performance on text-based and inference questions was tested separately.

[5] Mediation models with the individual motivation measures from the study ratings, motivation factor, and important notes as sole mediators were also tested. Neither mediator alone resulted in a significant indirect effect.

groups of participants. As shown in Table 4, ACT scores and prior knowledge were not significantly different across groups.

Materials

All materials are included in Appendix B and are available on the Open Science Framework (https://osf.io/cr6b7/).

Expectation Manipulation and Arousal Ratings. The same expectation instructions and arousal ratings from Experiment 1 were used. Descriptive and inferential statistics can be found in Table 4.

Expository Text and Study Ratings. Participants read a 1,331-word grade-level appropriate expository text (Flesch Kincaid Grade Level, 10.8) that described how yeast is used to produce alcohol in the fermentation process as well as the production process of fermented beverages which contained 147 idea units (see Appendix B). The text describes a causal model of microorganism reproduction by describing the process of how new molecules are created through anaerobic fermentation and how those molecules are then converted into alcohol. Additionally, the text describes how various characteristics of yeast strains can influence alcohol production and how differences in the rates that sugars are metabolized during the fermentation process can alter the resulting product.

Three experts were asked to identify the subset of idea units most important to developing an understanding of the fermentation process. Any idea unit identified as important by two or more experts was coded as more important (67 of 147). These idea units were assumed to contribute the most value to the development of a mental model of the text. The remaining 80 idea units were designated as less important (but not irrelevant) for understanding the fermentation process.

The same study ratings from Experiment 1 were collected. Participants found the fermentation text (Difficulty: $M = 3.47$, $SD = 2.03$; Understanding: $M = 6.92$, $SD = 2.07$) used in Experiment 2 to be significantly more difficult, $t(418) = 4.60$, $p < .001$, $d = .45$, and harder to understand, $t(418) = -4.11$, $p < .001$, $d = .40$, than the Doppler Effect text (Difficulty: $M = 2.64$, $SD = 1.65$; Understanding: $M = 7.72$, $SD = 1.93$) used in Experiment 1. In addition to having a higher word count and being written at a higher grade level, other metrics from automated text analyses (Coh-Metrix; Graesser et al., 2004; McNamara et al., 2014) showed that the fermentation text ($z = 0.85$) contained a lower degree of referential cohesion than the Doppler Effect ($z = 1.53$), indicating that the reader would be required to make more connections themselves. The fermentation text ($z = -2.55$) also contained less explicit logical connectives than the Doppler Effect

Table 4

Means (Standard Deviations) of Reader Characteristics and Ratings for Experiment 2

Variable	Immediate: Test $n = 55$	Immediate: Teach $n = 52$	Delayed: Test $n = 56$	Delayed: Teach $n = 51$		Statistic: F	p	η_p^2
Reader characteristics								
ACT	24.22 (4.59)	24.56 (4.42)	23.77 (3.51)	23.61 (4.35)	1	0.02	.88	.00
					2	1.46	.23	.01
					Int	0.19	.67	.00
Prior knowledge	2.96 (1.67)	2.38 (1.24)	2.48 (1.35)	2.65 (1.31)	1	1.16	.28	.01
					2	0.33	.57	.00
					Int	3.75	.05	.02
Arousal measures								
Excited	4.33 (2.45)	4.33 (2.53)	4.64 (2.47)	5.25 (2.41)	1	0.82	.37	.00
					2	3.40	.07	.02
					Int	0.83	.36	.00
Stressed	4.73 (2.60)	5.42 (3.07)	4.96 (2.94)	5.24 (3.01)	1	1.48	.23	.01
					2	0.00	.95	.00
					Int	0.29	.59	.00
Study ratings								
Interesting	6.85 (2.41)	6.54 (2.31)	7.09 (2.09)	7.04 (2.19)	1	0.35	.55	.00
					2	1.42	.23	.01
					Int	0.19	.67	.00
Enjoyable	6.33 (2.33)	5.56 (2.36)	6.50 (2.27)	6.51 (2.24)	1	1.45	.23	.01
					2	3.19	.08	.01
					Int	1.53	.22	.01
Learn more	5.84 (2.86)	5.25 (2.61)	6.48 (2.27)	6.55 (2.48)	1	0.55	.46	.00
					2	7.68	.01	.04
					Int	0.87	.35	.00
Effort	5.98 (2.69)	6.67 (2.13)	6.45 (1.97)	6.35 (2.04)	1	0.96	.33	.01
					2	0.06	.81	.00
					Int	1.65	.20	.01
Difficulty	3.35 (2.31)	3.94 (2.18)	3.52 (1.94)	3.08 (1.52)	1	0.08	.78	.00
					2	1.57	.21	.01
					Int	3.52	.06	.02
Understanding	7.02 (2.20)	6.56 (2.08)	7.29 (1.78)	6.78 (2.20)	1	2.89	.09	.01
					2	0.76	.38	.00
					Int	0.01	.94	.00

Note. Statistics are calculated using a two-factor ANOVA. The main effects of expectation (1), time of test (2), and interaction (Int) are shown.

text ($z = -1.29$), indicating that relations between ideas within the fermentation text may be more implicit, again requiring the reader to make more connections themselves. Thus, these metrics as well as student perceptions of the texts and the pilot studies on test items (below) all provided converging evidence that the fermentation text was a more difficult text than the Doppler Effect text used in Experiment 1.

Multiple-Choice Test Questions. As shown in Table B1 and similar to Experiment 1, five text-based and five inference questions were created. As in Experiment 1, all of the test questions were related to the idea units categorized as most important to developing an understanding of the fermentation process by the experts. Collectively, the questions tested all sections of the text.

To validate the distinction between these question types, as in Experiment 1, an independent set of participants ($n = 20$) were asked to read the text and sort the test questions into text-based and inference categories. Participant ratings demonstrated consensus on the distinction between the question types as the raters showed a high level of agreement with the intended categories (overall agreement 81%, agreement for inference items 75% [range 60–90%], agreement for text-based items 86% [range 70–100%]). The majority of text-based questions that were miscategorized as requiring an inference involved a paraphrase between the text and test question. All inference questions that were miscategorized as text-based involved a bridging inference.

To show that the questions could not be answered from common knowledge, a separate group of participants ($n = 20$) completed all multiple-choice questions without reading the text. Accuracy rates were 23% overall. This shows that in general there is low prior knowledge about the process of fermentation, and that students did not perform significantly different than chance (25%) on these questions without reading the text, $t(19) = -0.89$, $p = .38$, $d = .20$.

Finally, an additional group ($n = 20$) completed the test questions while the text was available for them to reference. Performance on test questions was high overall (78%), but it was especially high for text-based questions (Text-based: $M = 88\%$, $SD = 15\%$; Inference: $M = 69\%$, $SD = 31\%$), resulting in a significant difference in performance between question types, $t(19) = 2.45$, $p = .02$, $d = .55$, and suggesting less overlap between the textbase and situation model for this text than the Doppler Effect text.

Free Recall. The same free recall instructions from Experiment 1 were used.

Final Survey and Manipulation Check. Participants were asked to complete the same final survey as was used in Experiment 1. Of the students who were expecting to teach, when asked to recall what they would be expected to do after studying the text, 11 additional participants did not recall that they would have to teach another student and were excluded from analyses leaving the 214 participants described above. Additionally, a one-sample t test (against the midpoint) indicated that participants did believe that they would have to teach another student, $M = 5.13$, $SD = 1.48$, $t(102) = 11.15$, $p < .001$, $d = 1.10$.

Parallel to Experiment 1, participants were asked to report their prior knowledge both during an initial screening survey at the start of the semester and in a final survey completed after the experiment. Prior knowledge responses given during initial screening ($n = 167$) were highly correlated with prior knowledge responses given during the final survey, $r = .89$, $p < .001$. Because complete data were only available for the final survey, these scores were used in analyses.

Procedure

The procedure was identical to Experiment 1. The only exception was that participants were given 20 min to study the fermentation text instead of the 10 min provided for the Doppler Effect text used in Experiment 1.

Coding

Notes. Two raters coded the notes for the idea units included by participants with high interrater reliability, ICC (3, 2) = .94. As in Experiment 1, this coding was used to compute the proportion of more-important idea units included in the notes, less-important idea units included in the notes, and the extent to which the order of the idea units in the notes matched the order in the original text. Similar to Experiment 1, only one participant included a new idea in their notes. (They stated that they had learned about fermentation in their biology class.) Because of the low incidence of new ideas introduced in student notes, no analyses were conducted on this measure.

Free Recall. Recall scores were calculated by coding for the number of idea units contained within the free recall responses out of the total 147 idea units in the text. These responses were coded in an identical manner to Experiment 1 and resulted in high interrater reliability, ICC (3, 2) = .98.

Results

All analyses followed the logic and procedures adopted in Experiment 1.

How Expecting to Teach Affects Learning Outcomes

Effects of Expecting to Teach on Multiple-Choice Test Performance. A 2 (Expectation) × 2 (Time of Test) × 2 (Question Type) mixed-design ANCOVA controlling for ACT scores and prior knowledge was conducted for performance on the test questions.[6] Initial tests indicated no violations of the assumptions of normality, homogeneity of variance, and linearity. Homogeneity of regression slopes was examined by testing interactions of the independent variables with the covariates; none were significant ($ps > .17$). Therefore, the assumption of homogeneity of regression slopes was also satisfied. Intercorrelations among covariates and the dependent variables are presented in Table 2.

As shown in Figure 5, a main effect of expectation indicated that those expecting to teach had higher test performance than those expecting to test, $F(1, 208) = 3.95$, $p = .048$, $\eta_p^2 = .02$. A main effect of time of test shows that participants had higher test scores when tested immediately than at a delay, $F(1, 208) = 31.25$, $p < .001$, $\eta_p^2 = .13$. Further, there was no significant interaction between these two factors, $F(1, 208) = 0.48$, $p = .49$, $\eta_p^2 = .00$, meaning the effect was found both immediately and at a delay.

[6] Analyses are presented without the prior knowledge covariate in Appendix C.

Thus, the first hypothesis that expecting to teach would improve learning was again supported.

A main effect of question type showed that participants had higher mean scores on text-based questions than on inference questions, $F(1, 208) = 43.28$, $p < .001$, $\eta_p^2 = .17$, yet the expectation by question type interaction was not significant, $F(1, 208) = 1.31$, $p = .25$, $\eta_p^2 = .01$. In addition, neither the time of test by question type interaction, $F(1, 208) = 0.84$, $p = .36$, $\eta_p^2 = .00$, nor the three-way interaction was significant, $F(1, 208) = 0.02$, $p = .89$, $\eta_p^2 = .00$. These results suggest that the effect of expectation was seen regardless of question type (and delay). Thus, the hypotheses that expecting to teach would improve both memory and comprehension processes were supported.

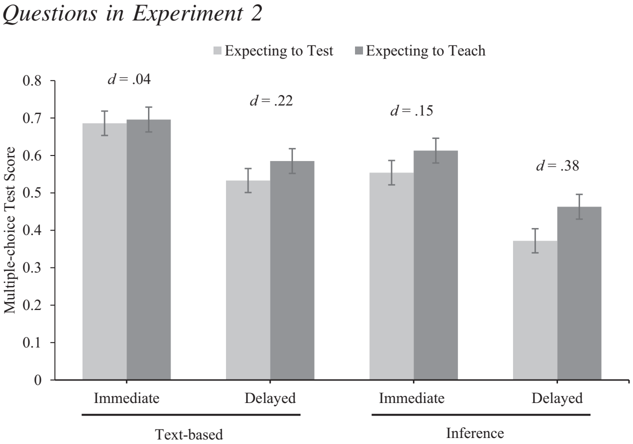
Effects of Expecting to Teach on Free Recall. As Experiment 1, negative binomial regression was used to analyze these data.[7] ACT score (centered) and prior knowledge (centered) were entered as covariates along with expectation and time of test as dummy-coded between-subjects factors. A likelihood ratio test indicated that together the predictors accounted for a significant amount of the variance, $\chi^2(4) = 83.11$, $p < .001$. As shown in Table 5, those expecting to teach recalled significantly more idea units than those expecting to test. In addition, those who were tested immediately recalled significantly more idea units than those tested at a delay.

When the interaction between expectation and time of test was included in the model it was not significant ($p = .81$) and did not improve the fit of the model; hence, it was not included in the final model. Consistent with the prior analyses, similar effects were found both immediately and at a delay. Because a benefit was seen even when recall was delayed, this result is consistent with the test results, and provides converging evidence that expecting to teach is benefiting learning.

Summary for Effects of Expectations on Learning Measures. Again, the main effect of expectation, and the fact that an effect of expectation was seen even on delayed tests, provide evidence that expecting to teach is benefiting learning even with more difficult text. Further, the fact that the effects were seen on both text-based and inference test items provides evidence that expecting to teach led to better memory and comprehension for this text. Given these findings, the next goal for analyses was to explore how expecting to teach might be changing the learning process, and whether differences in selectivity, organization, or motivation may help to explain the benefits.

Figure 5

Performance on Text-Based and Inference Multiple-Choice Test Questions in Experiment 2



Note. Multiple-choice test scores are shown for expectancy and time of test conditions. Error bars represent ± 1 standard error.

Table 5

Negative Binomial Regression Analysis for Free Recall in Experiment 2

Variable	β	*SE*	*z*	CI	*p*
Intercept	2.29	0.07	30.92	[2.14, 2.43]	<.001
Prior knowledge	0.11	0.03	3.37	[0.04, 0.17]	<.001
ACT	0.05	0.01	4.96	[0.03, 0.07]	<.001
Expectation	0.19	0.09	2.12	[0.02, 0.36]	.03
Time of test	−0.58	0.09	−6.56	[−0.75, −0.41]	<.001

Note. CI = confidence interval.

How Expecting to Teach Affects the Learning Process

Effects of Expecting to Teach on Note-Taking Content. A 2 (Expectation) × 2 (Time of Test) × 2 (Importance Level: More-important, Less-important) mixed-design ANOVA was conducted on the proportion of idea units included in the notes (out of the total possible). As shown in Figure 6, there was a significant three-way interaction, $F(1, 210) = 6.48$, $p = .01$, $\eta_p^2 = .03$. Two other significant effects help to explain the interaction. First, there was a main effect for importance level, $F(1, 210) = 97.24$, $p < .001$, $\eta_p^2 = .32$, which indicated that all participants included a greater proportion of the more-important idea units than the less-important idea units. There was also a main effect of expectation, $F(1, 210) = 8.27$, $p = .004$, $\eta_p^2 = .04$, which indicated that those who were expecting to teach included a greater proportion of the idea units in their notes than those expecting to test. The main effect of time of test was not significant, $F(1, 210) = 0.01$, $p = .91$, $\eta_p^2 = .00$. The two-way interaction between expectancy and importance level was not significant, $F(1, 210) = 1.51$, $p = .22$, $\eta_p^2 = .01$, nor was the two-way interaction between expectation and time of test, $F(1, 210) = 3.47$, $p = .06$, $\eta_p^2 = .02$. The time of test by importance level interaction was also not significant, $F(1, 210) = 0.78$, $p = .38$, $\eta_p^2 = .00$.

To follow-up the significant three-way interaction, the immediate and delayed conditions were examined separately. Divergences in statistical effects were seen in two places. First, in the immediate condition, only the importance level effect was significant, $F(1, 105) = 74.09$, $p < .001$, $\eta_p^2 = .41$. There was no effect of expectation, $F(1, 105) = 0.51$, $p = .48$, $\eta_p^2 = .01$, nor an interaction, $F(1, 105) = 1.11$, $p = .29$, $\eta_p^2 = .01$. In contrast, in the delayed condition, the importance level effect remained significant, $F(1, 105) = 33.00$, $p < .001$, $\eta_p^2 = .24$, but there was also an effect of expectation, $F(1, 105) = 11.22$, $p = .001$, $\eta_p^2 = .10$, and an interaction, $F(1, 105) = 5.84$, $p = .02$, $\eta_p^2 = .05$. These results indicate that expecting to teach did not lead to more extensive

[7] Following Experiment 1, the negative binomial regression was also used to analyze this measure. Again, the negative binomial regression provided a significantly better model fit than the standard Poisson regression, $\chi^2(1) = 252.43$, $p < .001$.

note-taking in the immediate condition in Experiment 2 but did lead to more extensive note-taking in the delayed condition. Further, participants who were expecting to teach in the delayed condition included a significantly greater proportion of more-important idea units in their notes than all other participants. The presence of the interaction between expectation and importance level among students in the delayed condition provides the strongest evidence in these experiments in support of the proposed selection mechanism, suggesting that part of the benefits of expecting to teach in this condition may be caused by some students focusing to a greater extent on more important ideas while studying the more difficult expository text. At the same time, the main effect for importance level in both conditions showed that all students tended to include more more-important ideas than less-important ideas in their notes. Similar to the results seen in Experiment 1, Table 2 shows that students who took notes that included a larger portion of the more-important ideas did better on the recall task, and tended to perform better on the multiple-choice tests.

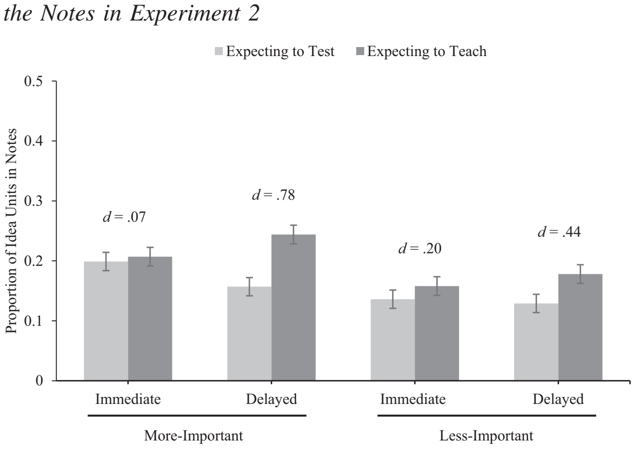
Effects of Expecting to Teach on the Organization of Notes. The amount of reorganization in the order of ideas from the text to the notes is shown in Figure 7. A 2 (Expectation) × 2 (Time of Test) ANOVA on the amount of reorganization seen in the notes revealed no differences across conditions. Neither the main effect of expectation, $F(1, 210) = 1.19$, $p = .28$, $\eta_p^2 = .01$, nor the main effect of time of test, $F(1, 210) = 0.10$, $p = .75$, $\eta_p^2 = .00$, were significant. The interaction was also not significant, $F(1, 210) = 1.30$, $p = .26$, $\eta_p^2 = .01$. In general, the notes showed very little restructuring. This is inconsistent with the proposed organization mechanism that the benefits of expecting to teach stem from the reorganization of ideas during study.

Arousal and Study Ratings. As shown in Table 4, in Experiment 2 no differences were seen in any ratings as a result of expectation condition. Students did not report feeling more stressed in Experiment 2, but they also did not report being more motivated. As in Experiment 1, a motivation factor was computed using Principal Components Analysis, which accounted for 66% of the variance from the study ratings (Interesting, Enjoyable, Learn More, and Effort). As shown in Table 2, the motivation factor was again positively related to test performance. However, the lack of a difference in these ratings across expectation conditions is inconsistent with a proposed role for motivation for the effects of expecting to teach. Instead, the only differences that emerged were for participants in the delayed condition to report a greater desire to learn more about the topic than participants in the immediate condition, and a tendency to report the experiment as being more enjoyable.

Even though no differences in motivation were found across expectation conditions in Experiment 2, the same serial mediation model used in Experiment 1 is reported for completeness (using Lavaan; Rosseel, 2012). Although motivation and important notes were significantly related to test performance using simple correlations, when entered into the mediation model only motivation was significantly predictive of learning.[8] As shown in Figure 8, the standardized regression coefficient between condition and motivation was not significant; the standardized regression coefficient between motivation and important notes was significant; and the standardized regression coefficient between important notes and test score was not. The significance of the indirect effect was computed, using the logic of Preacher and Hayes (2008), for each of 10,000 bootstrapped samples, and the 95% confidence interval was computed by determining the indirect effects at the 2.5th and 97.5th percentiles. The bootstrapped standardized effect was .00, and the 95% confidence interval crossed zero [−.003, .002]. Thus, the indirect effect was not significant.[9] Hence, motivation could not be shown to be responsible for the benefits of expecting to teach on learning via changes in selection of important ideas in the notes, or otherwise, in either experiment.

Figure 6

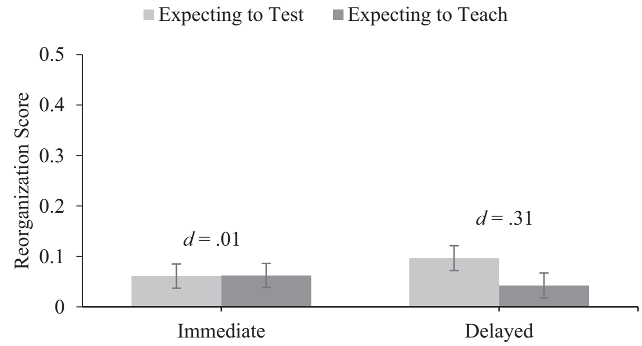
Proportion of More- and Less-Important Idea Units Present in the Notes in Experiment 2



Note. Proportion of more- and less-important idea units present in the notes are shown for expectancy and time of test conditions. Error bars represent ± 1 standard error.

Figure 7

Average Reorganization Within Notes in Experiment 2

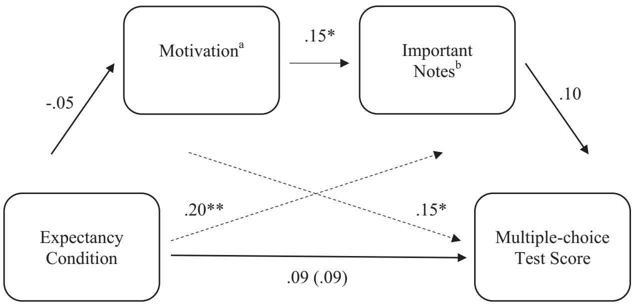


Note. Reorganization scores are shown for expectancy and time of test conditions. Error bars represent ± 1 standard error.

[8] Performance on the text-based and inference questions were correlated ($r = .53$, $p < .001$). Because of this a total score on the multiple-choice test was calculated for the dependent variable used in the mediation analysis. However, patterns remained the same even when test performance on text-based and inference questions was tested separately.

[9] Mediation models with the individual motivation measures from the study ratings, motivation factor, and important notes as sole mediators were also tested. Neither mediator alone resulted in a significant indirect effect.

Figure 8
Mediation Model for Experiment 2



Note. Standard regression coefficient for the relationship between expectancy condition and multiple-choice test score as mediated by motivation and the proportion of most-important idea units in the notes. The standardized regression coefficient between expectancy condition and multiple-choice test score, controlling for motivation and notes, is in parentheses. [a] Motivation factor score. [b] Proportion of more-important idea units included in the notes taken during study. * $p < .05$. ** $p < .01$.

Summary of Results. Similar to Experiment 1, these results provide evidence that expecting to teach is affecting learning of an expository science text on the fermentation process. Again, expecting to teach led to better learning on a delayed test, suggesting that it results in durable learning over time. Further, the fact that the effects were seen on both text-based and inference test items provides evidence that expecting to teach led to better memory and comprehension for this text. Importantly, the results of Experiment 2 show that benefits from expecting to teach could be seen even with a more difficult expository text.

General Discussion

The present set of experiments tested whether expecting to teach helps to improve learning from both simpler and more difficult expository texts. Participants were given either the expectation to teach or test followed by the opportunity to study the text. They then completed tests either immediately or after a 1-week delay. Experiment 1 used a simpler (below grade level) text and Experiment 2 used a more difficult (appropriate grade level) text. Importantly, consistent findings were shown across experiments with respect to learning outcomes.

The first hypothesis was that expecting to teach would benefit learning compared to expecting to test. This hypothesis was supported by those expecting to teach having higher mean test performance both on immediate and delayed tests. Because benefits were seen on a delayed test, this suggests that expecting to teach leads to durable learning that persists over time. Although Fiorella and Mayer (2013, Experiment 2) did not find a significant statistical difference between those who were expecting to test and those who were expecting to teach when tested at a delay, it is possible that their nonsignificant results were because they did not have enough power to detect the effect because of the smaller sample size.

The second hypothesis was that expecting to teach would benefit memory for the text compared to those expecting to test. This hypothesis was supported by those expecting to teach having higher mean performance on multiple-choice text-based test questions. These findings are consistent with those of Nestojko et al. (2014). They also add to the existing literature by showing that expecting to teach can lead to memory benefits both immediately and over time. Further, support was found for the third hypothesis, that expecting to teach would benefit comprehension compared to expecting to test. Those expecting to teach had higher mean performance on the multiple-choice inference test questions. This result is consistent with the findings of Benware and Deci (1984), and effect sizes found on inference questions at a delay (Experiment 1: $d = .40$; Experiment 2: $d = .38$) are quite similar to those found in prior research and used for power calculations ($d = .46$). For a variety of reasons, prior studies on the benefits of expecting to teach were not in a position to clearly test for improvements specifically in comprehension outcomes. However, by employing separate sets of text-based and inference questions, the present experiments extend prior findings and provide clear support that expecting to teach can improve comprehension from expository science texts. Additionally, the better performance seen on inference questions that test the situation model at a delay provides clear evidence that expecting to teach resulted in long-term benefits in student understanding.

Other exploratory analyses examined study ratings and behaviors that participants engaged in during study. These analyses were intended to provide insight into whether either of the two possible mechanisms that have been previously suggested might help to explain the benefits of expecting to teach on learning: either that expecting to teach increases motivation and engagement in the task (Benware & Deci, 1984) or that it improves the ability to select and organize information during study (Bargh & Schul, 1980; Fiorella & Mayer, 2013; Nestojko et al., 2014).

In terms of a potential explanation based in motivation, there was a general positive relation found between motivation and learning in both experiments. However, differences in motivation ratings between conditions were only found in Experiment 1, and mediation analyses indicated that these differences were not fully responsible for benefits seen in learning outcomes. Renkl (1995) also was unable to demonstrate mediation by motivation in his studies. Because the present experiments were designed as an attempt to directly replicate and extend the findings of past research, the current experiments used the same motivation questions that have been used in the precedent work. However, these questions consisted of only a few self-report ratings. It is possible that future research using more extensive and established motivation and engagement scales could help to clarify the role of motivation in expecting to teach effects (Fredricks et al., 2011; Martin, 2007; Vallerand et al., 1992).

In terms of a potential explanation for expecting to teach effects based in changes in study behaviors, this research analyzed the notes that students took while studying. Although past research on expecting to teach has allowed participants to take notes during study, there have not been any prior analyses reported using this trace data. In the present studies, very little evidence was found for organization processes during notetaking. On the other hand, all students tended to include more important information in their notes, and the extent to which they did so tended to predict learning in both experiments. Yet, the expectation condition did not cause students to be more selective than others. When looking for differences between the conditions, the only evidence that emerged was for more selective notetaking in the delayed condition in Experiment 2. Although it seems likely that real teachers do engage in selection and organization of ideas as they prepare to teach, there was no evidence in this trace data to suggest that the expecting to teach benefit was attributable to participants engaging in these behaviors. However, it may be important to note that the participants in these studies (and students in classrooms) are most likely novices at teaching. They do not have the training and experience that real teachers do. Given that students who engaged in selecting out the most important information from a text during study tended to do better on the tests, it may be that clarifying the goals of teaching to include these behaviors would increase the benefits of expecting to teach. Students may need to be told explicitly that their goal for teaching is to understand how and why the scientific phenomena is occurring, and that they need to select out the most important ideas and organize them into a coherent mental model. This may help participants to engage in more effective study behaviors in the expecting to teach condition, and in turn this may produce even stronger effects from expecting to teach.

Limitations

The main goal of these experiments was to attempt to replicate and extend the findings of expecting to teach manipulations. As part of this goal, texts of different difficulty and topics were used which provided an advantage in the ability to generalize the findings. However, this can also be seen as a limitation. An alternative approach would be to use a more elegant design that holds the topic and tests constant and only varies the text difficulty. Although this would have helped to determine how manipulating difficulty in a particular way affects expecting to teach (and could have possibly helped to eliminate other differences between the experiments that may have affected the motivation analyses), it would not have helped to ensure generalizability.

In the current experiments, study time was held constant across expectancy conditions. Although this was a strategic design decision intended to rule out the benefits of expecting to teach as being merely due to differences in time spent studying, it is possible that this may have also limited the benefits of the manipulation. When time is not constrained, participants who are given more demanding reading goals often spend more time on the text (Yeari et al., 2015). Had participants been given unconstrained study preparation time in these experiments, those expecting to teach may have spent more time on the task as well. And, in this additional time, they might possibly have been able to engage in more active construction of their lesson or a new phase involving revision and reorganization of their notes. Past work suggests that these kinds of activities are useful for learning (Hinze et al., 2013; Luo et al., 2016), and engaging in them could have increased the benefits of expecting to teach.

Another potential limitation of this research was that the instructions provided to participants were intentionally left vague, so as to not influence what the participants thought “expecting to teach” or “expecting to test” meant. It is possible that the interpretation of the task goals that were implied by “teaching” and “testing” varied across participants. Of course, pursuing different goals can influence the reading process, and lead to differing mental representations of the text (Coleman et al., 1997; Geiger & Millis, 2004; Ishiwa et al., 2013; Kintsch, 1994; Schmalhofer & Glavanov, 1986; van den Broek et al., 2001; Zwaan, 1994). Without clear signaling of how to approach the task of preparing to teach, it is unlikely that students in the expecting to teach condition were all adopting the same goals for reading. Some other experiments have provided a bit more clarification for the participants about what expecting to teach meant. In the case of Fiorella and Mayer (2013), participants were told as part of the teaching expectation that they would be “explaining how the Doppler Effect works” to another student, which provides participants with a more specific goal. This may be a particularly effective goal, because expecting to generate an explanation after reading has been shown to improve learning in and of itself (Coleman et al., 1997; Fukaya, 2013). Expecting to teach may also be seen as giving students more of a purpose for reading (Britt et al., 2017). However, as outlined by these authors, just providing a reader with a purpose for their reading does not imply that they will know how to achieve it. Again, this suggests that by better defining the goals for “teaching,” participants might be better able to engage in appropriate and effective study behaviors as they prepare to teach.

In addition, the instructions provided to all students in these experiments stated that the test that would be given would be like a “typical” test. For the purposes of replication and generalization, these instructions were taken from Benware and Deci (1984). However, research has shown that providing an expectation about the test format (e.g., essay vs. multiple-choice; McDaniel et al., 1994; Thiede, 1996) or question type (e.g., inference vs. text-based; Griffin et al., 2019; Jensen et al., 2014; Thiede et al., 2010) can affect learning. When the nature of a test is left undefined, students may generally assume that their goal for learning is to try to remember the information and not necessarily to comprehend it (Thiede et al., 2010). Thus, informing students about the nature of the final test also helps them to alter their reading goals. In the future, it seems promising to test whether manipulations that include both clear test-expectancies and more explicit task goals might increase the benefits from expecting to teach.

Finally, prior knowledge has been shown to influence text comprehension because it aids the reader in making connections to what they already know (Graesser & Bertus, 1998; Kintsch, 1994). To aid in replication, the current study adopted the self-report measures of prior knowledge used in the previous work (Fiorella & Mayer, 2013, 2014) and the practice of selecting topics for which most undergraduates might be expected to have low prior knowledge. Of course, including more objective tests of knowledge about the topics would be useful to better elucidate how prior knowledge may affect the benefits of expecting to teach in future research.

Conclusions

An important contribution of the present experiments is that expecting to teach was found to improve learning even at a delay. Improved performance on delayed assessments was critical for showing that expecting to teach can promote durable learning over time. It also offers some promise that these benefits might be seen in more authentic classroom contexts, as students are generally not tested immediately after study in their classes. Given the nonsignificant findings of the mediation analyses, there is still much more work to be done to understand the mechanisms underlying the benefits of expecting to teach.

The preparation stage is only the first of the three stages of the teaching process that were originally outlined by Bargh and Schul (1980). There is already research to suggest that engaging in the other stages of the teaching process, which are inherently more generative in nature, can lead to additional benefits over just preparing to teach (Fiorella & Mayer, 2013, 2014; Koh et al., 2018; Kobayashi, 2019). The results of the present experiments support the provocative claim that a portion of the benefit in learning by teaching is solely attributable to the first preparation stage and from expecting to teach.

References

Annis, L. F. (1983). The processes and effects of peer tutoring. *Human Learning: Journal of Practical Research & Applications*, *2*(1), 39–47.

Bargh, J. A., & Schul, Y. (1980). On the cognitive benefits of teaching. *Journal of Educational Psychology*, *72*(5), 593–604. https://doi.org/10.1037/0022-0663.72.5.593

Benware, C. A., & Deci, E. L. (1984). Quality of learning with an active versus passive motivational set. *American Educational Research Journal*, *21*(4), 755–765. https://doi.org/10.3102/00028312021004755

Britt, M. A., Rouet, J. F., & Durik, A. M. (2017). *Literacy beyond text comprehension: A theory of purposeful reading*. Taylor & Francis. https://doi.org/10.4324/9781315682860

Cohen, P. A., Kulik, J. A., & Kulik, C. L. (1982). Educational outcomes of tutoring: A meta-analysis of findings. *American Educational Research Journal*, *19*(2), 237–248. https://doi.org/10.3102/00028312019002237

Coleman, E. B., Brown, A. L., & Rivkin, I. D. (1997). The effect of instructional explanations on learning from scientific texts. *Journal of the Learning Sciences*, *6*(4), 347–365. https://doi.org/10.1207/s15327809jls0604_1

Coxe, S., West, S. G., & Aiken, L. S. (2009). The analysis of count data: A gentle introduction to Poisson regression and its alternatives. *Journal of Personality Assessment*, *91*(2), 121–136. https://doi.org/10.1080/00223890802634175

Daou, M., Lohse, K. R., & Miller, M. W. (2016). Expecting to teach enhances motor learning and information processing during practice. *Human Movement Science*, *49*, 336–345. https://doi.org/10.1016/j.humov.2016.08.009

Fiorella, L., & Mayer, R. E. (2013). The relative benefits of learning by teaching and teaching expectancy. *Contemporary Educational Psychology*, *38*(4), 281–288. https://doi.org/10.1016/j.cedpsych.2013.06.001

Fiorella, L., & Mayer, R. E. (2014). Role of expectations and explanations in learning by teaching. *Contemporary Educational Psychology*, *39*(2), 75–85. https://doi.org/10.1016/j.cedpsych.2014.01.001

Fredricks, J., McColskey, W., Meli, J., Mordica, J., Montrosse, B., & Mooney, K. (2011). *Measuring student engagement in upper elementary through high school: A description of 21 instruments. (Issues & Answers Report, REL 2011–No. 098)*. U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Southeast. http://ies.ed.gov/ncee/edlabs

Fukaya, T. (2013). Explanation generation, not explanation expectancy, improves metacomprehension accuracy. *Metacognition and Learning*, *8*, 1–18. https://doi.org/10.1007/s11409-012-9093-0

Gardner, W., Mulvey, E. P., & Shaw, E. C. (1995). Regression analyses of counts and rates: Poisson, overdispersed Poisson, and negative binomial models. *Psychological Bulletin*, *118*(3), 392–404. https://doi.org/10.1037/0033-2909.118.3.392

Geiger, J. F., & Millis, K. K. (2004). Assessing the impact of reading goals and text structures on comprehension. *Reading Psychology*, *25*(2), 93–110. https://doi.org/10.1080/02702710490435637

Graesser, A. C., & Bertus, E. L. (1998). The construction of causal inferences while reading expository texts on science and technology. *Scientific Studies of Reading*, *2*(3), 247–269. https://doi.org/10.1207/s1532799xssr0203_4

Graesser, A. C., McNamara, D. S., Louwerse, M. M., & Cai, Z. (2004). Coh-Metrix: Analysis of text on cohesion and language. *Behavior Research Methods, Instruments, & Computers*, *36*, 193–202. https://doi.org/10.3758/BF03195564

Griffin, T. D., Wiley, J., & Thiede, K. W. (2019). The effects of comprehension-test expectancies on metacomprehension accuracy. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *45*(6), 1066–1092. https://doi.org/10.1037/xlm0000634

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*(2), 151–164. https://doi.org/10.1016/j.jml.2013.03.002

Hoogerheide, V., Loyens, S. M. M., & van Gog, T. (2014). Effects of creating video-based modeling examples on learning and transfer. *Learning and Instruction*, *33*, 108–119. https://doi.org/10.1016/j.learninstruc.2014.04.005

Ishiwa, K., Sanjosé, V., & Otero, J. (2013). Questioning and reading goals: Information-seeking questions asked on scientific texts read under different task conditions. *British Journal of Educational Psychology*, *83*, 502–520. https://doi.org/10.1111/j.2044-8279.2012.02079.x

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. https://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*(3), 621–629. https://doi.org/10.1037/a0015183

King, A., Staffieri, A., & Adelgais, A. (1998). Mutual peer tutoring: Effects of structuring tutorial interaction to scaffold peer learning. *Journal of Educational Psychology*, *90*(1), 134–152. https://doi.org/10.1037/0022-0663.90.1.134

Kintsch, W. (1994). Text comprehension, memory, and learning. *American Psychologist*, *49*(4), 294–303. https://doi.org/10.1037/0003-066X.49.4.294

Kintsch, W., Welsch, D., Schmalhofer, F., & Zimny, S. (1990). Sentence memory: A theoretical analysis. *Journal of Memory and Language*, *29*(2), 133–159. https://doi.org/10.1016/0749-596X(90)90069-C

Kobayashi, K. (2019). Learning by preparing to teach and teaching: A meta-analysis. *Japanese Psychological Research*, *61*(3), 192–203. https://doi.org/10.1111/jpr.12221

Koh, A. W. L., Lee, S. C., & Lim, S. W. H. (2018). The learning benefits of teaching: A retrieval practice hypothesis. *Applied Cognitive Psychology*, *32*(3), 401–410. https://doi.org/10.1002/acp.3410

Luo, L., Kiewra, K. A., & Samuelson, L. (2016). Revising lecture notes: How revision, pauses, and partners affect note taking and achievement. *Instructional Science*, *44*, 45–67. https://doi.org/10.1007/s11251-016-9370-4

Martin, A. J. (2007). Examining a multidimensional model of student motivation and engagement using a construct validation approach. *British Journal of Educational Psychology*, *77*(2), 413–440. https://doi.org/10.1348/000709906X118036

Mayer, R. E. (1989). Models for understanding. *Review of Educational Research*, *59*(1), 43–64. https://doi.org/10.3102/00346543059001043

McDaniel, M. A., Blischak, D. M., & Challis, B. (1994). The effects of test expectancy on processing and memory of prose. *Contemporary Educational Psychology*, *19*(2), 230–248. https://doi.org/10.1006/ceps.1994.1019

McNamara, D. S., Graesser, A. C., McCarthy, P. M., & Cai, Z. (2014). *Automated evaluation of text and discourse with Coh-Metrix*. Cambridge University Press. https://doi.org/10.1017/CBO9780511894664

McNamara, D. S., Kintsch, E., Songer, N. B., & Kintsch, W. (1996). Are good texts always better? Interactions of text coherence, background knowledge, and levels of understanding in learning from text. *Cognition and Instruction*, *14*(1), 1–43. https://doi.org/10.1207/s1532690xci1401_1

Nestojko, J. F., Bui, D. C., Kornell, N., & Bjork, E. L. (2014). Expecting to teach enhances learning and organization knowledge in free recall of text passages. *Memory & Cognition*, *42*, 1038–1048. https://doi.org/10.3758/s13421-014-0416-z

Ozuru, Y., Best, R., Bell, C., Witherspoon, A., & McNamara, D. S. (2007). Influence of question format and text availability on the assessment of expository text comprehension. *Cognition and Instruction*, *25*(4), 399–438. https://doi.org/10.1080/07370000701632371

Palinscar, A. S., & Brown, A. L. (1984). Reciprocal teaching of comprehension-fostering and comprehension-monitoring activities. *Cognition and Instruction*, *1*(2), 117–175. https://doi.org/10.1207/s1532690xci0102_1

Preacher, K. J., & Hayes, A. F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior Research Methods*, *40*, 879–891. https://doi.org/10.3758/BRM.40.3.879

Rawson, K. A., & Kintsch, W. (2005). Rereading effects depend on time of test. *Journal of Educational Psychology*, *97*(1), 70–80. https://doi.org/10.1037/0022-0663.97.1.70

Renkl, A. (1995). Learning for later teaching: An exploration of mediational links between teaching expectancy and learning results. *Learning and Instruction*, *5*(1), 21–36. https://doi.org/10.1016/0959-4752(94)00015-H

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

Roscoe, R. D., & Chi, M. T. H. (2008). Tutor learning: The role of explaining and responding to questions. *Instructional Science*, *36*, 321–350. https://doi.org/10.1007/s11251-007-9034-5

Rosseel, Y. (2012). Lavaan: An R package for structural equation modeling and more. Version 0.5–12 (BETA). *Journal of Statistical Software*, *48*, 1–36. https://doi.org/10.18637/jss.v048.i02

Salomon, G., & Perkins, D. N. (1989). Rocky roads to transfer: Rethinking mechanism of a neglected phenomenon. *Educational Psychologist*, *24*(2), 113–142. https://doi.org/10.1207/s15326985ep2402_1

Schmalhofer, F., & Glavanov, D. (1986). Three components of understanding a programmer's manual: Verbatim, propositional, and situational representations. *Journal of Memory and Language*, *25*(3), 279–294. https://doi.org/10.1016/0749-596X(86)90002-1

Simmons, D. C., Fuchs, L. S., Fuchs, D., Mathes, P., & Hodge, J. P. (1995). Effects of explicit teaching and peer tutoring on the reading achievement of learning-disabled and low-performing students in regular classrooms. *The Elementary School Journal*, *95*(5), 387–408. https://doi.org/10.1086/461851

Tapiero, I., & Otero, J. (2002). Situation models as retrieval structures: Effects on the global coherence of science texts. In J. Otero, J. A. Leon, & A. C. Graesser (Eds.), *The psychology of science text comprehension* (pp. 179–198). Routledge.

The College Board. (2018). *Guide to the 2018 ACT/SAT concordance*. https://collegereadiness.collegeboard.org/pdf/guide-2018-act-sat-concordance.pdf

Thiede, K. W. (1996). The relative importance of anticipated test format and anticipated test difficulty on performance. *The Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology*, *49*(4), 901–918. https://doi.org/10.1080/713755673

Thiede, K. W., Griffin, T. D., Wiley, J., & Anderson, M. C. M. (2010). Poor metacomprehension accuracy as a result of inappropriate cue use. *Discourse Processes*, *47*(4), 331–362. https://doi.org/10.1080/01638530902959927

Vallerand, R. J., Pelletier, L. G., Blais, M. R., Briere, N. M., Senecal, C., & Vallieres, E. F. (1992). The Academic Motivation Scale: A measure of intrinsic, extrinsic, and amotivation in education. *Educational and Psychological Measurement*, *52*(4), 1003–1017. https://doi.org/10.1177/0013164492052004025

van den Broek, P., Lorch, R. F., Linderholm, T., & Gustafson, M. (2001). The effects of readers' goals on inference generation and memory for texts. *Memory & Cognition*, *29*, 1081–1087. https://doi.org/10.3758/BF03206376

Voss, J. F., & Silfies, L. N. (1996). Learning from history text: The interaction of knowledge and comprehension skill with text structure. *Cognition and Instruction*, *14*(1), 45–68. https://doi.org/10.1207/s1532690xci1401_2

Webb, N. M. (2013). Information processing approaches to collaborative learning. In C. E. Hmelo, C. A. Chinn, C. K. K. Chan, & A. M. O'Donnell (Eds.), *The international handbook of collaborative learning* (pp. 19–40). Routledge.

Wiley, J., Griffin, T. D., & Thiede, K. W. (2005). Putting the comprehension in metacomprehension. *Journal of General Psychology*, *132*(4), 408–428. https://doi.org/10.3200/GENP.132.4.408-428

Wiley, J., & Guerrero, T. A. (2018). Prose comprehension beyond the page. In K. Millis, J. Magliano, D. Long, & K. Wiemer (Eds.), *Deep comprehension: Multi-disciplinary approaches to understanding, enhancing, and measuring comprehension* (pp. 3–15). Routledge. https://doi.org/10.4324/9781315109503-2

Wiley, J., & Voss, J. F. (1999). Constructing arguments from multiple sources: Tasks that promote understanding and not just memory for text. *Journal of Educational Psychology*, *91*(2), 301–311. https://doi.org/10.1037/0022-0663.91.2.301

Yeari, M., van den Broek, P., & Oudega, M. (2015). Processing and memory of central versus peripheral information as a function of reading goals: Evidence from eye-movements. *Reading and Writing*, *28*, 1071–1097. https://doi.org/10.1007/s11145-015-9561-4

Ziegler, E., & Stern, E. (2014). Delayed benefits of learning elementary algebraic transformations through contrasted comparisons. *Learning and Instruction*, *33*, 131–146. https://doi.org/10.1016/j.learninstruc.2014.04.006

Zwaan, R. A. (1994). Effect of genre expectations on text comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*(4), 920–933. https://doi.org/10.1037/0278-7393.20.4.920

(*Appendices follow*)

Appendix A

Experiment 1 Instructions and Test Questions

Table A1
Instructions for Teach and Test Expectancy Groups

Initial instructions given to all participants	
Please read the following text in the same manner that you would read any text assigned in one of your college courses. Read and study it so that you have learned it as well as you can in a period of about 20 minutes. You may take notes on the lesson itself and/or use the blank sheet of paper provided to you by the experimenter. Use whatever methods are most natural and most beneficial to you for learning the material.	
Continuation of instructions	
Expecting to test	Expecting to teach
The purpose of studying and learning the text is so that you will score as high as possible on a test based on the text. During the time when you are testing you will not have access to the text. The test will be like a typical test based on a reading assignment. Again, use whatever study methods seem most appropriate for you.	The purpose of studying and learning the text is so that you **will be able to teach the contents to another student so they** will score as high as possible on a test based on the text. During the time when you are **teaching** you will not have access to the text. **The student to whom you teach the contents will then be given a test based on the text.** The test will be like a typical test based on a reading assignment. Again, use whatever study methods seem most appropriate for you.

Note. The bolded terms highlight the differences in the instructions between conditions.

(*Appendices continue*)

Table A2
Text-Based and Inference Multiple-Choice Test Questions for the Doppler Effect Text

Multiple-choice questions	
Text-based	Inference
1. As a fire truck passes you, what will happen to the pitch of the sounds from the siren? *It will get lower. It will get higher. It will become more frequent. It will stay the same.	1. Two ambulances are equally distant from you. One is moving away from you and one is approaching you. Which is true of how you perceive their sirens? *If they are the same distance from you, the sirens will sound the same in pitch. The ambulance siren that is approaching you will sound further away. The ambulance siren that is moving away from you will be louder. The ambulance siren that is approaching you will seem higher.
2. Which of the following most influences how we perceive sound? Depth perception Loudness of sound *Movement Pitch	2. If you are moving toward a sound wave, what will happen to the wave? *Wavelengths will become closer together. Frequency will become lower. Pitch will become lower. Sound will get louder.
3. Which has the shortest wavelength? *    	3. If two firetrucks are traveling toward you and one increases its speed, what will happen to the perceived sound of the siren by the observer from the faster moving truck? The pitch will be the same as the slower moving truck. *The wavelengths will become shorter than the slower moving truck. The frequency will get lower than the slower moving truck. The pitch will be lower than the slower moving truck.
4. Which is the lowest frequency sound wave? *   	4. If the pitch heard by an observer is higher than that of the emitted source frequency, then which statement is NOT correct? The observer is moving toward the stationary source. The source is moving toward the stationary observer. *The observer and source are moving at the same speed. The observer and source are approaching each other.
5. What is the term for how we perceive the pitch of a sound? Wavelength *Wave frequency Doppler Effect Sound waves	5. If you move away from a high pitched sound, what will happen to the sound wave? The frequency will remain the same. The wavelengths will decrease. The frequency will increase. *The pitch will decrease.

Note. Items marked with an asterisk (*) denotes the intended answer.

(*Appendices continue*)

Appendix B

Experiment 2 Text and Test Questions

The Role of Yeast in the Fermentation Process

Bread, Beer, and the Science of Yeast

Baking bread, brewing beer, producing biofuel, getting a gross skin infection. What do all four activities have in common? Yeast. Yeast are eukaryotic microorganisms that can reproduce asexually and metabolize sugars anaerobically (without oxygen) through a process called alcoholic fermentation. And they are everywhere—there are thousands of species of these single-celled fungi living in all corners of the globe, including all over your body.

Despite being kind of gross, yeast are culinary powerhouses. In fact, they were probably one of the first creatures that humans domesticated. Even today we use them as little fermentation factories—feed them sugar, deprive them of air, and they turn out alcohol and carbon dioxide at a mean rate. When yeast is added to bread dough, the carbon dioxide gas it produces causes the bread to rise. As the dough bakes, it eventually gets hot enough to kill the yeast cells. Yeast also provides the alcohol content in beer, wine, and liquor. The reaction by which alcoholic beverages are produced is generally referred to as alcoholic fermentation.

In general, alcoholic fermentation is a process where sugar is broken down into ethanol alcohol and carbon dioxide. In more specific terms, the process of alcoholic fermentation begins when glucose (sugar) comes into contact with the yeast. The glucose provides the yeast cells with the necessary energy to complete the process of fermentation. The glucose is then broken down to yield a molecule of energy (ATP) for the yeast cell. In addition to ATP, two molecules called pyruvate are made. The pyruvate is then converted by the yeast cell into the byproducts, carbon dioxide (CO_2) and alcohol (ethanol).

During the production of beer or wine, yeast converts the glucose, derived from grains or grape juice, to alcohol and carbon dioxide which are actually just byproducts of the biochemical process that yeast use to extract energy from sugar when it doesn't have access to oxygen. However, from the yeast's point of view, alcohol and carbon dioxide are waste products. Unfortunately, for the yeast, alcohol is toxic, so after the alcohol produced by the yeast reaches a critical concentration, the yeast cells die. As the yeast continues to grow and metabolize in the sugar solution, the accumulation of alcohol that is produced will become toxic, thereby killing the yeast cells. This typically happens when the solution reaches a concentration between 14–18%, which means that there is a limit to the alcohol content of beer and wine produced by normal fermentation. Hence, the reason why the percentage of alcohol in wine and beer is most often under 18%.

There are thousands of species of yeast which vary in the amount of total amount of sugar they can convert. Attenuation refers to the percentage of sugar the yeast can convert to carbon dioxide and ethanol. Yeast strains, on average, attenuate 65–80%, meaning that there is a small percentage of residual sugar that remains in the beverage. Champagne yeasts are a strain of yeasts that typically attenuate to a higher percentage, around 75%. When brewers or wine makers want to achieve a beverage with high alcohol content, they can use yeast with high attenuation like Champagne yeasts because more sugar will be converted to alcohol. This would result in a product with very little glucose left in the product.

What's the Difference Between Beer and Wine?

Generally, any beverage derived from fermented fruit juice is classified as wine. However, commercially speaking, "wine" is fermented grape juice from *vitis vinifera,* a grape vine. Beer on the other hand is usually derived from fermentation of malt derived from the digestion of germinated barley grains, in western cultures, but other grains may be utilized in other cultures.

There is also a difference between processes by which wines and beers are fermented. Beer making is almost a science. Compared with wine making, it is rather complex and there's a purpose for everything that is done in making beer. Beer makers can control just about everything that goes into beer. Wine making, on the other hand, is relatively simple. It's truly a natural drink and its origin probably preceded beer making. The yeast responsible for fermenting the sugars in the fruits are usually present on the grape skins. Fermentation will begin whenever there is a break in the skin in which the sugar inside comes into contact with the yeasts on the outside of the grape. So when human production of wine began, it involved collecting fruit, crushing them, and allowing them to ferment with their own natural yeast.

Beer Making

Despite the complex machinery that is used in brewing beer, it's still essentially the same procedure that has been used for hundreds of years. However, beer making has become very sophisticated because of the advances in knowledge that has resulted from advances in science. Prior to and even during the 1800s, there were many who knew how beer could be made, but no one knew of the science behind each step. It would not be until the 19th century that it would be known that yeasts were the organisms that actually were responsible for the fermentation process.

In the beginning of beer making, beer was an alcoholic beverage with the flavor of malt and grain. It was flat, slightly sweet and would spoil quickly. It would not be until the 8th century, that brewers in central Europe found that the addition of hop flowers preserved the beer and gave it a slightly bitter taste that made it more palatable. Yeast did not become an ingredient in beer until it was identified by Louis Pasteur in 1857.

(*Appendices continue*)

When yeast is deprived of oxygen and added to a sugar-rich solution, it immediately begins to consume the sugars and reproduce, creating more yeast. But from the brewer's point of view, the important thing is not the growth of more yeast, but the waste products of yeast metabolism, alcohol and carbon dioxide which gives beer its fizz. As the food supply, sugar, is digested and the alcohol levels rise, the environment becomes literally toxic to the yeast, which becomes dormant and dies.

Recently brewers have become interested in controlling the amount of carbohydrates in beer. What that amounts to is developing a strain of yeast that will digest more of the carbohydrates in barley malt, so beer drinkers won't have to. During fermentation, a yeast cell breaks down a glucose into carbon dioxide and ethanol—with a dozen or so chemical reactions in between, each catalyzed by a specific enzyme. Of the 500 species of yeast, many can break down more of the complex carbohydrates in barley malt than brewer's yeast, to ferment a beer with fewer calories.

Making Wine

Wine is made today much the same way that it was centuries ago. However, unlike beer, there is still a great deal that cannot be controlled in the production of wine. The grapes from which the wine is to be made are first separated from the stem and then crushed to release the juice. The combination of the skin, juice, and seeds is called the must. If the desired product is a white wine, the free juice is transferred to a fermentation tank and the peels and stems are removed and pressed again. If red wine is the desired product, the skins of the grape go into the fermentation tank with the juice. The red color of this wine is from the red pigment in the outer skin of the grape.

Various vessels may be used as the fermentation tank. Once the juice is in the fermentation tank, preferred strains of yeast can sometimes be added, but are not needed. The skin of the grapes already has adequate yeasts on them so this is not necessary. This is one of the reasons wine quality is more difficult to control. Because the yeasts that grow on the grapes vary in different vineyard, especially if they are in different countries, the quality of the finished wine will also vary.

Table B1
Text-Based and Inference Multiple-Choice Test Questions for Fermentation Text

Multiple-choice questions	
Text-based	Inference
1. What is fermentation? *A process where a sugar is broken down into ethanol and carbon dioxide. A process where a sugar is broken down into yeast and carbon dioxide. A process where yeast in broken down to ethanol and carbon dioxide. A process where sugar is broken down to ethanol and yeast.	1. Which is true of anaerobic fermentation? Oxygen is needed as a source of energy for the yeast. For sugars to be metabolized, yeast must have an attenuation rate above 60%. *Pyruvate molecules must be converted into CO_2 and ethanol. More ATP is required to produce CO_2 than alcohol.
2. In the fermentation process, what molecule is produced as yeast consume sugars? Lactic acid Glucose *ATP Carbon dioxide	2. What is special about the yeast involved in the wine making process? They can be strategically selected so that the quality of wine will be consistent. They can reach alcohol content of up to 50%. *They don't have to be added by the winemaker. They don't require as much glucose for the fermentation process.
3. What are the byproducts of fermentation? *Ethanol and CO_2 Glucose and ethanol Pyruvate and CO_2 Alcohol and pyruvate	3. What could be a reason for a wine to contain a large amount of residual sugars? Yeast with a high attenuation rate were used. All sugars have been metabolized by the yeast During fermentation yeasts with access to oxygen will not digest as much glucose. *Toxicity causes the yeast to die before the glucose has been broken down.
4. What causes wine or beer to become toxic environment for yeast? Increases in glucose Decreases in oxygen Decreases in pyruvate *Increases in alcohol	4. If yeast cells have a low attenuation, what is most likely to happen to the beer or wine? It will have a high alcohol content. *It will contain residual sugars. It will take a longer amount of time to ferment. It will create a toxic environment.
5. Crushing or breaking the skin of a grape is the first step in the fermentation process of wine. Why is this true? The sugar begins to produce alcohol. *The sugar comes into contact with the yeast. The sugar decreases carbon dioxide levels. The sugar mixes with the pyruvate.	5. Why is wine considered to be less complex to produce as compared with beer? *Wine has naturally occurring yeast. Grapes have naturally occurring sugars. Beer requires exact ratios of products to be added. Beer requires a preservative.

Note. Items marked with an asterisk (*) denotes the intended answer.

(*Appendices continue*)

Appendix C

Analyses Excluding Prior Knowledge as a Covariate

Table C1

Effect of Expecting to Teach on Multiple-Choice Test Performance Excluding Prior Knowledge as a Covariate

Experiment	F	p	η_p^2
Experiment 1			
Expectation	5.56	.02	.03
Time of test	9.48	.002	.05
Question type	2.82	.09	.01
Expectation × Time of Test	0.71	.40	.004
Expectation × Question Type	1.76	.19	.01
Time of Test × Question Type	1.24	.63	.001
Expectation × Time of Test × Question Type	1.70	.19	.01
Experiment 2			
Expectation	2.50	.12	.01
Time of test	29.74	<.001	.12
Question type	43.48	<.001	.17
Expectation × Time of Test	1.62	.20	.01
Expectation × Question Type	1.29	.26	.01
Time of Test × Question Type	0.85	.36	.004
Expectation × Time of Test × Question Type	0.01	.91	.00

Received February 12, 2020
Revision received December 5, 2020
Accepted December 8, 2020 ■