



Benefits of Writing an Explanation During Pauses in Multimedia Lessons

Alyssa P. Lawson¹ · Richard E. Mayer¹

Accepted: 6 January 2021 / Published online: 23 March 2021

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2021

Abstract

Generative learning theory posits that learners engage more deeply and produce better learning outcomes when they engage in selecting, organizing, and integrating processes during learning. The present experiments examine whether the generative learning activity of generating explanations can be extended to online multimedia lessons and whether prompts to engage in this generative learning activity work better than more passive instruction. Across three experiments, college students learned about greenhouse gasses from a 4-part online lesson involving captioned animations and subsequently took a posttest. After each part, learners were asked to generate an explanation (write-an-explanation), write an explanation using provided terms (write-a-focused-explanation), rewrite a provided explanation (rewrite-an-explanation), read a provided explanation (read-an-explanation), or simply move on to the next part (no-activity). Overall, students in the write-an-explanation group (Experiments 2 and 3), write-a-focused-explanation group (Experiment 2), and rewrite-an-explanation group (Experiment 3) performed significantly better on a delayed posttest than the no-activity group, but the groups did not differ significantly on an immediate posttest (Experiment 1). These results are consistent with generative learning theory and help identify generative learning strategies that improve online multimedia learning, thereby priming active learning with passive media.

Keywords Explanations · generative learning theory · delayed testing · multimedia lesson · learning strategy

Author note: Requests for the materials, scoring keys, and raw data may be sent to Alyssa Lawson at Alyssa.lawson@psych.ucsb.edu.

✉ Alyssa P. Lawson
Alyssa.lawson@psych.ucsb.edu

¹ Department of Psychological and Brain Sciences, University of California, Santa Barbara, Santa Barbara, CA 93106, USA

Objective and Rationale

How can we improve the instructional effectiveness of online multimedia lessons, such as an annotated animation on greenhouse gases, like the lesson show in Fig. 1? This question is particularly relevant in light of the increasing role of remote learning for education around the world due to the coronavirus pandemic. The goal of this intervention study is to examine and understand the effectiveness of prompts to engage in generative learning embedded in online multimedia lessons. Traditionally, instructional design principles have focused on how best to present instructional material (Mayer, 2014). However, even the best-designed lessons will not be maximally effective if students do not process them appropriately. Therefore, researchers and practitioners have increasingly examined the instructional effectiveness of prompts to engage in generative learning activities during learning (Fiorella & Mayer, 2015, 2016).

Generative learning activities are behaviors that an individual engages in during learning with the goal of fostering deeper learning that leads to improved understanding (Fiorella & Mayer, 2015, 2016). Based on generative learning theory (Fiorella & Mayer, 2015, 2016; Mayer, 2020; Wittrock, 1974, 1989), deep learning occurs when learners engage in appropriate cognitive processing during learning, such as attending to relevant incoming information, mentally organizing it into a coherent structure, and integrating it with relevant prior knowledge activated from long-term memory. Generative learning activities, such as generating explanations during a lesson, are intended to prime each of these processes.

There are some generative learning strategies that focus on verbal activities include summarizing, explaining, paraphrasing, question answering, and elaborating in one's own

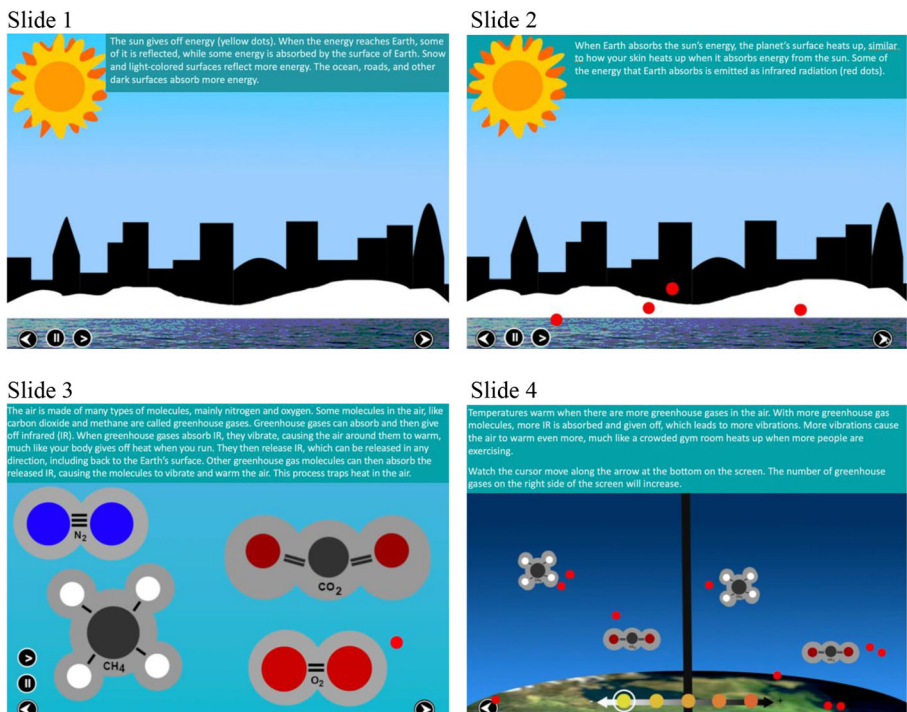


Fig. 1 Images of the four animated slides

words, whereas some generative learning strategies that focus on visual or spatial activities include mapping, outlining, drawing, imagining, or tracing. In the present set of three experiments, we focus on verbal learning strategies in written form because they are easiest to implement in a computer-based learning situation. In particular, across all three experiments, we ask students to write an explanation after each segment of a four-part animated lesson, which we call the write-an-explanation activity, using prompts shown in the first column of Table 1 to promote the verbal strategy of explanation. We focus on prompting students to generate explanations during an online multimedia lesson because this generative activity prompt appears to be powerful enough to cause learners to engage in several verbal activities and because of its potential to help engage students in an online lesson that might otherwise not foster sufficient student engagement. This approach is based on distinguishing between the external activity prompt (e.g., “please write an explanation”) and the internal cognitive process that it may prime (e.g., explaining, summarizing, paraphrasing, etc.). We focus on explaining during a pause after an animation rather than during an animation because an animation is transitory and subject to creating split attention when presented in conjunction with another task (Leahy & Sweller, 2011).

The first row in Table 1 indicates the conditions included in each experiment. In the first two studies, we also examine a version of explanation that we call write-a-focused-explanation, in which the prompt specifies the names of terms that should be included in the written explanation as a form of scaffolding, using prompts shown in the second column of Table 1. In the third study, we ask some students to rewrite a provided written explanation in their own words, which we call rewrite-an-explanation, using prompts shown in the third column of Table 1. In all three studies, we compare learning outcomes of the study activity groups with two control groups that are shown an instructor-provided explanation (which we call the read-an-explanation group), as shown in the fourth column of Table 1, or are not given a study activity prompt (which we call the no-activity group), as shown in the final column of Table 1.

Literature on Learning by Explaining

Learning by explaining occurs when students generate a written or oral explanation of instructional material they are reading or viewing (Fiorella & Mayer, 2015, 2016). In a review, Fiorella and Mayer (2015) reported that in 44 of 54 experimental tests, students who were prompted to explain what they were reading or viewing performed better on a posttest than students who were not, yielding a median effect size of $d = 0.61$. This research includes studies involving reading static text lessons and studies involving viewing dynamic multimedia presentations. Research on learning by explaining dates back to Chi et al.’s (1989) early work on what they called *self-explanation* in which students spontaneously engaged in generating oral explanations for themselves as they read a text. This line of research morphed somewhat into research on prompts to write explanations during paper-based lessons and, more recently, during computer-based, online lessons. It should be noted that explaining to oneself (e.g., in the content of making sense of a lesson) may require different cognitive activities than explaining to others (e.g., in the context of being prompted to write an explanation). This paper focuses on a form of learning by explaining in which learners are prompted to write an explanation.

Inspired by the results of the 1989 study, Chi et al. (1994) tried to understand this phenomenon in an experimental study. They had students read through a text on the

Table 1 Prompts given throughout lesson

Exp. #	Write-an-explanation	Write-a-focused-explanation	Rewrite-an-explanation	Read-an-explanation (control)	No-activity (control)
Exp. 1, 2, 3 After Slide 1	Exp. 1, 2, 3 “Please explain what you learned in the last clip.”	Exp. 1, 2 “Please explain what you learned about absorbed energy and reflected energy.”	Exp. 3 “Energy that comes through the atmosphere from the sun can either be reflected off the surface of the Earth (as reflected energy) or absorbed into the surface of the Earth (as absorbed energy). Please rewrite this explanation in your own words on the sheet provided.”	Exp. 1, 2, 3 “Energy that comes through the atmosphere from the sun can either be reflected off the surface of the Earth (as reflected energy) or absorbed into the surface of the Earth (as absorbed energy).”	Exp. 1, 2, 3 “Press the right arrow key on the keyboard to move forward.”
After Slide 2	“Please explain what you learned in the last clip.”	“Please explain what you learned about absorbed energy and infrared radiation.”	“Some energy from the sun gets absorbed by the Earth (as absorbed energy), which causes the Earth’s surface to heat up. As the Earth’s surface becomes warmer, it then gives off infrared radiation. Please rewrite this explanation in your own words on the sheet provided.”	“Some energy from the sun gets absorbed by the Earth (as absorbed energy), which causes the Earth’s surface to heat up. As the Earth’s surface becomes warmer, it then gives off infrared radiation.”	“Press the right arrow key on the keyboard to move forward.”
After Slide 3	“Please explain what you learned in the last clip.”	“Please explain what you learned about how greenhouse gases and non-greenhouse gases interact with infrared radiation.”	“Greenhouse gases, such as carbon dioxide and methane, are the only types of gases that can absorb and interact with infrared radiation, they vibrate, warm up, and then release the infrared radiation back into the atmosphere. Please rewrite this explanation in your own words on the sheet provided.”	“Greenhouse gases, such as carbon dioxide and methane, are the only types of gases that can absorb and interact with infrared radiation, they vibrate, warm up, and then release the infrared radiation back into the atmosphere.”	“Press the right arrow key on the keyboard to move forward.”

Table 1 (continued)

Write-an-explanation	Write-a-focused-explanation	Rewrite-an-explanation	Read-an-explanation (control)	No-activity (control)
<p>After Slide 4</p> <p>“Please explain what you learned in the last clip.”</p>	<p>“Please explain what you learned about how the number of greenhouse gases in the atmosphere relates to both the vibration of greenhouse gases and to the temperature of the atmosphere.”</p>	<p>“As the number of greenhouse gases in the atmosphere increase, the more likely it is that infrared radiation will interact with them. When more greenhouse gases interact with infrared radiation, this leads to more vibration and therefore more heat being created in the atmosphere. Please rewrite this explanation in your own words on the sheet provided.”</p>	<p>“As the number of greenhouse gases in the atmosphere increase, the more likely it is that infrared radiation will interact with them. When more greenhouse gases interact with infrared radiation, this leads to more vibration and therefore more heat being created in the atmosphere.”</p>	<p>“Press the right arrow key on the keyboard to move forward.”</p>

circulatory system. While reading through the text, students were either prompted to create an explanation after reading each sentence or were told to simply read through the text twice. The students who were asked to create explanations while reading had bigger gains in understanding, especially when asked difficult questions, compared to those that had not explained. This research demonstrated the benefits of prompting students to engage in what they called *self-explanation* while learning (and we call learning by explaining).

From these early studies on learning by explaining, many more studies have tried to understand the when, how, and why asking students to generate explanations during learning is an effective learning tool (i.e., Atkinson et al., 2003; Johnson & Mayer, 2010; Margulieux & Catrambone, 2018; McEldoon et al., 2013). Much of the research investigating learning by explaining has found it beneficial to student learning across different instructional materials, different topics, and different students. Additionally, a meta-analysis found that asking students to generate explanations during learning resulted in better test performance as compared to direct instruction, regardless of timing, content, and format of the explanations (Bisra et al., 2018).

Learning Using Multimedia Material

Multimedia learning materials, particularly those presented on computers, are different from more traditional paper-based learning materials. The multimedia learning hypothesis suggests that people learn more deeply when information is given in the form of words and pictures rather than just from words alone (Mayer, 2014). This would suggest that learning from explaining in a dynamic multimedia context could present learners with more opportunities to engage in deep learning. However, this type of medium also introduces new challenges based on the transient nature of the presented material. Thus, it is vital to understand how different types of learning strategies can be incorporated into dynamic multimedia lessons using new technology in order to understand if the benefits of learning by explaining persist.

Recently, there has been an increase in research aimed at investigating the use of learning from explanations with animated lessons. Unlike previous research mentioned above, there is more discrepancy in this data. Consistent with much of the prior literature on learning by explaining, some research suggests that being asked to generate explanations during learning can be beneficial to learning from animated lessons (i.e., De Koning et al., 2011; Lin & Atkinson, 2013; Ryoo & Linn, 2014; Wouters et al., 2008). The effect is found with inference and transfer questions, but not always with retention questions, suggesting that the act of generating explanations may help learners process the information more deeply (De Koning et al., 2011). Yet, some research finds that prompts to explain, when used in conjunction with an animation, are no better than providing direct instruction without prompts to explain (De Koning et al., 2010; Lin et al., 2016). Other research finds that prompting students to explain a dynamic presentation in pairs can be more beneficial than giving students an already developed explanation (Ryoo & Linn, 2014). Overall, research is mixed concerning the effects of prompts to generate explanations with animations.

The argument for why prompts to explain may not be beneficial when used with an animated lesson is based on the idea that explaining can create too much cognitive load (Lin et al., 2016). A cause of the cognitive load increase from using animations in a computer-based lesson is that the information is not permanently onscreen (Leahy & Sweller, 2011; Spanjers et al., 2010; Wong et al., 2012). The transitory nature of the information of an animated lesson

causes more processing load to the learner than a static picture. On top of that, having learners also engage in a difficult task, like generating an explanation, requires more cognitive load capacity to be used and can cause cognitive overload rather than a beneficial effect (Lin et al., 2016). We attempt to overcome this potential problem in the present study by prompting learners to explain during pauses in a multimedia lesson.

Novelty of This Research

This research is beneficial to our understanding of using explanation in learning for several reasons. First, this set of studies investigates how prompts to generate explanations can be added into an online lesson. Especially now, online lessons are becoming more prominent and a struggle many learners are facing is the lack of engagement that the technology promotes. This research aims to understand how to add a constructive component to learning that is often more passive and lacks an engaging pull. Second, this research uses a multimedia lesson that is commercially available, in an attempt to understand how to improve educational material that is already being used by learners. Much of the research on learning by explaining has not focused directly on multimedia presentations involving animation and the ones that have examined lessons with animation have shown conflicting results (De Koning et al., 2010; De Koning et al., 2011; Lin & Atkinson, 2013; Lin et al., 2016; Wouters et al., 2008). Lastly, this research involves delayed testing, which is often not used when investigating the effects of learning by explaining.

Theoretical Framework and Predictions

Generative learning activities, such learning by explaining, are inspired by generative learning theory, which posits that meaningful learning occurs when learners engage in appropriate cognitive processing during learning (Fiorella & Mayer, 2015, 2016; Mayer, 2020; Wittrock, 1974, 1989). These cognitive processes for building structure in working memory include the following:

Selecting—paying attention to the relevant incoming information for further processing in working memory

Organizing—arranging the selected information into a coherent cognitive structure in working memory

Integrating—connecting the selected information with relevant existing knowledge activated from long-term memory

Generative learning activities such as writing or rewriting explanations are intended to prime these three cognitive processes during learning. First, the students must select the material to include in their written explanation. Second, students must mentally organize the material into a coherent statement. Third, by putting the statement in their own words, students must use their relevant prior knowledge.

Generative learning theory dates back to the work of Wittrock on learning as a generative activity, which began in the 1970s (1974, 1989; Doctorow et al., 1978; Grabowski, 2004; Lee et al., 2008; Mayer, 2010; Tobias, 2010), and is related to earlier conceptions of meaningful

learning proposed by the Gestalt psychologists in the first half of the 20th century (e.g., Katona, 1942; Wertheimer, 1959). Generative learning theory is central to current conceptions of cognitive constructivism, which holds that learning is an active process of building cognitive structures in working memory (Mayer, 1992, 2020).

Additionally, the ICAP framework (Chi, 2009; Chi et al., 2018; Chi & Wylie, 2014) provides a complementary approach to the role of learner-generated explanations during learning. The ICAP framework proposes a taxonomy that differentiates four types of engagement during learning and orders them from lowest to highest level of engagement. Passive, the lowest level of engagement, occurs when there is no overt action occurring during learning, such as simply viewing a multimedia lesson. Active, the second level, occurs when the learner engages in a shallow level of overt action during learning, such as reading printed words aloud in each segment of a multimedia lesson. Constructive, the third level, occurs when learners generate information that goes beyond the given lesson, such as writing an explanation for each segment of a multimedia lesson. Interactive, the highest level of engagement, occurs when there is some sort of interaction with another person or computer agent during a multimedia lesson such as discussing the meaning of each segment of a multimedia lesson with another student. According to the ICAP framework, learning by explaining (or what can be called *self-explanation*) as implemented in the present study falls under the constructive level (Chi & Wylie, 2014). This is because learners are generating an explanation that goes beyond the given information but are not interacting with others.

We draw on generative learning theory and the ICAP framework to derive predictions concerning the effectiveness of prompting students to engage in generative learning activities such as writing or rewriting an explanation versus not prompting students to engage in generative learning activities. We measure learning outcomes with a broad posttest that involves writing answers to open-ended questions about the processes involved in greenhouse gases. Classic theory and research on meaningful learning suggests that learning in a meaningful way creates a more durable memory representation that persists in a coherent form longer than learning in a rote way, which is reflected in some evidence that the effects of meaningful learning show up better on a delayed test than on an immediate test (Katona, 1942; Mayer & Wittrock, 2006; Wertheimer, 1959). Importantly, this pattern has been observed in modern research on generative learning strategies such as learning by testing (also called retrieval practice) in which the effects of studying and self-testing are superior to studying and studying again when students take a delayed test but not when they take an immediate test (Brown et al., 2014; Dunlosky et al., 2013; Fiorella & Mayer, 2015, 2016; Roediger III & Karpicke, 2006). Additionally, the ICAP framework suggests that deeper learning occurs when higher levels of engagement are met; accordingly, activities that encourage more engagement should engender higher delayed test performance than those who have lower levels of engagement. Thus, in the present study, we examine whether effects of generative activities are stronger on delayed tests than on immediate tests.

Writing an explanation, writing a focused explanation, and rewriting an explanation all are generative activities that are at higher levels of the ICAP framework (i.e., constructive or active) than passively reading the material (i.e., passive) or reading an instructor-generated explanation (i.e., passive). According to generative learning theory, in concert with the ICAP framework, our primary predictions are that the write-an-explanation group will outperform the no-activity group on the posttest in Experiments 1, 2, and 3 (hypothesis 1a), the write-a-focused-explanation group will outperform the no-activity group on the posttest in

Experiments 1 and 2 (hypothesis 1b), and the rewrite-an-explanation group will outperform the no-activity group on the posttest in Experiment 3 (hypothesis 1c).

Research on effectiveness of instructional interventions depends both on the intervention and on the control group. In the present study, the difference between the treatment groups and the no-activity groups include learner activity and engagement as well as access to a written explanation. In order to control for access to a statement of explanation, we included a control group that simply reads an instructor-provided explanation after each of the four segments of the lesson but does not engage in activity (i.e., read-an-explanation group). Similar to hypothesis 1, writing an explanation and rewriting an explanation both require generative activity that is not required for reading an explanation and they are at a higher level of the ICAP framework than passively reading an explanation. Accordingly, we predict that the write-an-explanation group will outperform the read-an-explanation group on the posttest in Experiments 1, 2, and 3 (hypothesis 2a), the write-a-focused-explanation group will outperform the read-an-explanation group on the posttest in Experiments 1 and 2 (hypothesis 2b), and the rewrite-an-explanation group will outperform the read-an-explanation group on the posttest in Experiment 3 (hypothesis 2c).

The ICAP framework (Chi, 2009; Chi et al., 2018; Chi & Wylie, 2014) suggests that activities at the same level (such as write-an-explanation and write-a-focused explanation) should promote the same level of deep learning, and thus there should be no differences in learning outcome. Based on this analysis, we predict that in Experiments 1 and 2, the write-an-explanation group should perform similarly to the write-a-focused-explanation group in (hypothesis 3). Additionally, this framework suggests that engaging in a learning activity at the constructive level, such as writing an explanation, should lead to better learning than a learning activity at the active level, such as rewriting an explanation. Based on this analysis, we predict that the write-an-explanation group should outperform the rewrite-an-explanation group on the posttest in Experiment 3 (hypothesis 4). We also are interested in whether the effects are strong on delayed tests and immediate tests, because prior work suggests that interventions aimed at deep learning may show stronger effects on delayed tests.

Experiment 1

The main goal of Experiment 1 is to assess whether adding prompts to write-an-explanation (or write-a-focused explanation) in a computer-based multimedia lesson affects performance on an immediate test of learning.

Method

Participants and Design

The participants were 126 undergraduates recruited from a university in Southern California through a psychology subject pool, in which they fulfilled a course requirement by participating. A power analysis based on $\alpha = 0.05$, effect size = 0.65, and power = 0.80 demonstrated that a sample of this size would be sufficient. The mean age of the participants was 18.67 years ($SD = 0.82$), the mean prior knowledge score was 4.55 ($SD = 2.18$) which is considered low, and 88 of them were women. The experiment used a between-subjects design with 4 levels (explanation type: write-an-explanation, write-a-focused-explanation, read-an-explanation, no-

activity) between-subjects design with 33 participants in the write-an-explanation group, 32 participants in the write-a-focused-explanation group, 30 participants in the read-an-explanation group, and 31 participants in the no-activity group.

Materials

The paper-based materials consisted of a prequestionnaire and a postquestionnaire. The computer-based materials consisted of 4 versions of a self-paced multimedia lesson on greenhouse gases and a posttest consisting of 7 questions.

P questionnaire The prequestionnaire solicited demographic information (including gender and age), had a five-point scale asking students to rate their knowledge of how greenhouse gases work from “very low” to “very high,” and included 11 statements relating to knowledge about greenhouse gases and global warming that students were asked to place a check mark next to if the statement applied to them. A sample of these statements include: “I have taken a class that has discussed greenhouse gases,” “I consider myself to be an environmentalist,” and “I know how a greenhouse works.”

The correlation between the self-reported prior knowledge rating and the more objective checklist assessment (i.e., the number of checked statements) was moderate, $r(124) = 0.46$, $p < 0.001$. For this reason, our assessment of prior knowledge is based solely on the more objective measure of number of items checked on the prior knowledge checklist measure. We chose to use the checklist as our measure of prior knowledge rather than a pretest on material from the lesson because we want to avoid a testing effect in which the act of taking a test is an instructional even that causes learning (Brown et al., 2014; Roediger III & Karpicke, 2006) and we wanted to avoid guiding learners’ attention to lesson material that could be primed by the pretest questions. The Cronbach’s alpha for the objective checklist assessment (with possible scores from 0 to 11 based on the number of checked items) is 0.64. An explanation for the sub-optimal internal consistency is that the checklist was designed to tap diverse situations across a variety of sub-domains that were intended to assess students’ broad background knowledge about the environment rather than to assess specific knowledge of greenhouse gases with a set of similar statements.

Multimedia lessons The instructional materials consisted of four versions of a computerized multimedia lesson that described how greenhouse gases warm up the atmosphere. All versions of the lesson had the same four animated slides with onscreen text (ranging from 39 words to 104 words at the top of the screen) describing the different steps of how heat from the sun interacts with greenhouse gases and becomes trapped in the atmosphere. Screenshots of the slides are shown in Fig. 1. The lesson animations came from a KQED public television lesson on greenhouse gases, found here: <https://ww2.kqed.org/quest/2014/12/12/how-do-greenhouse-gases-work/>. In the slides, students learned about how the Earth absorbs light from the sun, which is then expelled from the Earth as infrared radiation. This radiation interacts with greenhouse gases and causes the gases to vibrate and warm up the air around them. When there are more greenhouse gases in the atmosphere, the temperature rises quickly because there are more interactions occurring causing more vibrations.

After each animation slide, participants saw a slide with one of four prompts, corresponding to the explanation group the participant was in. Participants were prompted by the system to do the activities listed for the one column in Table 1 that corresponded to the condition they were randomly assigned to. The participants in the write-an-explanation group were given prompts that asked them to explain what they had learned in the previous slide (shown in the first column of Table 1). The participants in the write-a-focused-explanation group were given prompts that asked them to explain specific aspects they had learned in the previous slide (shown in the second column of Table 1). The specific terms in the prompt were intended to help scaffold the learner's generative activity. Participants in these two groups wrote their explanations on a separate sheet of paper that was provided. The participants in the read-an-explanation group were given an onscreen printed explanation about the material in the previous slide which was designed as the answer to the question prompted by the write-a-focused-explanation group prompt (shown in the fourth column of Table 1). This served as a control group that provided access to an explanation of the presented material (as in the write-an-explanation and write-a-focused explanation groups) without prompting learners to engage in a generative learning activity. The explanations provided to participants in the read-an-explanation group were created by the experimenters based on the material and were not verbatim text reproduced from the lesson. The participants in the no-activity group were given a prompt to move to the next slide (shown in the fifth column of Table 1). This was intended as the primary control group.

Posttest The posttest consisted of 7 open-ended questions intended to assess participants' ability to apply what was presented in the lesson at a variety of levels of transfer: (1) "Based on the lesson you saw, please explain how greenhouse gases work." (2) "What prevents infrared radiation from leaving the Earth's atmosphere?" (3) "How would planting more trees/plants affect the temperature of the atmosphere?" (4) "What is a reason that temperatures on Earth might decrease, on average?" (5) "Why does your skin feel warm when you step out into the sunlight?" (6) "How would Earth's atmosphere be different if the atmosphere contained only nitrogen and oxygen?" (7) "How could we decrease the Earth's temperature without changing the amount of greenhouse gases in the atmosphere?" Cronbach's alpha for the posttest was 0.56. A reason for this sub-optimal reliability is that the posttest questions were designed to assess learning at a variety of levels of transfer ranging from remembering to understanding to analyzing and to assess a variety of explanative mechanisms described in the lesson, rather than to have all questions similar in terms of level and content. The posttest questions were presented one question at a time, with 60 seconds allowed per question. The time limit was imposed to standardize the testing experience and to determine which ideas were most easily accessible to learners. Each question was presented on the computer screen and participants wrote their answers on a sheet of paper that was provided.

Postquestionnaire The postquestionnaire included six subjective questions intended to assess learners' experience with the lesson: (1) "I enjoyed this lesson." (2) "The topic of this lesson was interesting to me." (3) "I would like to learn from more lessons like this." (4) "I felt as though the way this lesson was taught was effective for me." (5) "How difficult was this lesson for you?" and (6) "How much effort did you exert during this lesson?" These items were rated on a five-point scale ranging from 1 (strongly disagree) to 5 (strongly agree) for items 1–4

(which were intended to assess enjoyment, interest, motivation, and satisfaction), from 1 (very easy) to 5 (very difficult) for item 5 (which was intended to measure perceived intrinsic cognitive load), and from 1 (very little effort) to 5 (very much effort) for item 6 (which was intended to measure perceived germane cognitive load). Overall, the postquestionnaire was included to provide preliminary and exploratory information about the learners' learning process and perceived load that might be useful in future research but was not the main focus of the study.

Apparatus

The apparatus consisted of 4 iMac computer systems, with 20-inch color screens, each housed in an individual cubicle that blocked visual contact among participants.

Procedure

Participants were randomly assigned to one of the four groups and tested at individual cubicles in a lab setting with up to four participants in each session. First, the participants completed the prequestionnaire at their own pace. Next, the experimenter provided oral instructions for the study and began the lesson, which was self-paced, on each participant's computer. Participants could stay on any of the lesson or explanation slides as long as they wanted, but they were not allowed to move backwards in the lesson. The amount of time spent on the lesson was measured for each participant. After the lesson, participants completed the posttest. Each posttest question was individually displayed on the screen for 60 seconds. After the 60 seconds, participants were told to move on to the next question. We imposed a time limit in order to standardize testing conditions for all participants and to ensure that participants could not skip through the questions. Once finished with the posttest, participants completed the postquestionnaire at their own rate. The entire experiment took no longer than 30 minutes. We obtained Institutional Review Board (IRB) approval and adhered to guidelines for ethical treatment of human subjects.

Results and Discussion

Scoring the Posttest

One point was awarded for each of the key points participants included across the 7 transfer questions, yielding a total possible score of 36. Two researchers graded each posttest independently, and then all disagreements were resolved through discussion until 100% agreement was reached. A Pearson correlation was run to determine the relationship between the two researchers' grading and found a strong correlation between the two researchers' point assignment, $r = 0.87$, $p < 0.001$.

Do the Groups Differ on Basic Characteristics?

A preliminary issue concerns whether random assignment produced groups that were equivalent on basic characteristics. Concerning background knowledge score, there

were no statistically significant differences based on explanation prompt, $F(3, 122) = 0.59$, $p = 0.623$. Concerning age, there were no statistically significant differences based on explanation prompt, $F(3, 122) = 1.59$, $p = 0.196$. Concerning gender, a chi-square test showed that there was not a significant difference among the four groups based on explanation prompt, $\chi^2(3, N = 126) = 0.78$, $p = 0.855$. We conclude that participants in each group were equivalent in the basic characteristics of prior knowledge, age, and gender composition.

Does Writing Explanations During Pauses in a Multimedia Lesson Improve Learning?

Generative learning theory and the ICAP framework predict that the write-an-explanation group would do better than the no-activity group (hypothesis 1a) and the write-a-focused-explanation group would do better than the no-activity group (hypothesis 1b). Table 2 shows the posttest means and standard deviations for each group. A one-way (explanation type) ANOVA showed there was not a significant main effect of explanation prompt, $F(3, 122) = 1.36$, $p = 0.258$. In order to test this first set of a priori predictions, we conducted a Tukey post-hoc test. In contrast to hypotheses 1a and 1b, there were no significant differences found between the write-an-explanation group ($M = 10.12$, $SD = 3.62$) and the no-activity group ($M = 8.74$, $SD = 3.46$, $p = 0.409$) nor between the write-a-focused-explanation group ($M = 10.03$, $SD = 3.89$) and the no-activity group ($p = 0.341$). Additionally, there was no significant difference between the read-an-explanation group ($M = 8.87$, $SD = 3.16$) and the no-activity group ($p = 0.999$).

Generative learning also predicts that the write-an-explanation group would do better than the read-an-explanation group (hypothesis 2a) and the write-a-focused-explanation would do better than read-an-explanation group (hypothesis 2b). The same Tukey post-hoc test was used to test this second set of a priori predictions. In contrast to the hypotheses 2a and 2b, there were no significant differences found between the write-an-explanation group and the read-an-explanation group ($p = 0.361$) nor between the write-a-focused-explanation group and the read-an-explanation group ($p = 0.426$). Based on ANOVAs with post-hoc tests, the differences did not reach statistical significance although they were in the predicted direction.

Table 2 Means and standard deviations on learning posttest for groups in all experiments

	Write-an-explanation			Write-a-focused-explanation			Rewrite-an-explanation			Read-an-explanation (control)			No-activity (control)	
	<i>M</i>	<i>SD</i>	<i>d</i>	<i>M</i>	<i>SD</i>	<i>d</i>	<i>M</i>	<i>SD</i>	<i>d</i>	<i>M</i>	<i>SD</i>	<i>d</i>	<i>M</i>	<i>SD</i>
Experiment 1 (immediate)	10.12	3.62	0.39	10.03	3.89	0.35	-	-	-	8.87	3.16	0.04	8.74	3.46
Experiment 2 (delayed)	8.06	4.02	0.68	7.27	3.71	0.50	-	-	-	4.47	2.84	0.28	5.41	3.78
Experiment 3 (delayed)	9.13	2.99	1.31	-	-	-	7.79	3.40	0.85	6.81	4.04	0.51	4.90	3.43

Note. Bolded scores are significantly greater than the no-activity group at $p < 0.05$

ICAP predicts that the write-an-explanation group would perform similarly to the write-a-focused-explanation group (hypothesis 3). The same Tukey post-hoc test was used to test this third a priori prediction. In agreement with hypothesis 3, there was not a significant difference between the write-an-explanation group and the write-a-focused-explanation group ($p = 1.00$).

To summarize the data, we conducted contrast analyses using two models: a basic model in which the explaining groups each have a weight of +1 and the two control groups each have a weight of -1 (based on the idea that explaining is better than not explaining), and a weighted model in which the write-an-explanation group has a +2 weight, the write-a-focused-explanation group has a weight of +1, the read-an-explanation group has a weight of -1, and the no-activity group has a weight of -2 (based on the idea that certain kinds of explaining are more effective than others). In Experiment 1, the basic model was the best fitting model with R-square = 0.996, and $p = 0.004$. There is evidence for a model in which groups that engage in explaining during learning perform better than those that do not engage in any generative learning activities during learning, as predicted by generative learning theory.

Lastly, as we are interested in understanding how using generative learning strategies compared to non-generative learning strategies in this lesson, we also ran a t test that compared the combination of the two generative strategies to the combination of the two control groups. The t test was significant, $t(124) = 2.03$, $p = 0.045$, supporting the observation that participants who engaged in generative strategies ($M = 10.08$, $SD = 3.73$) outperformed participants who did not ($M = 8.80$, $SD = 3.29$). We did not use a Bonferroni correction because we only conducted one t test. Based on these summary analyses, there is some evidence for the power of the generative activity of writing explanations during pauses in a multimedia lesson.

Overall, these results suggest that, although there were no significant differences among all four groups on an immediate test based on an ANOVA, the more fine-grained analyses based on contrast analysis and the t test for combined explanation and combined control groups suggest an emerging pattern that should be explored further.

Are There Effects of the Amount of Time Spent on the Lesson?

A potential difference between the write explanations groups and the control groups is that the write explanations groups were required to spend more time on the lesson. First, an ANOVA was run to determine if there were significant differences in time spent on the lesson between the groups. A one-way ANOVA showed there was a significant main effect of explanation type, $F(3, 122) = 132.93$, $p < 0.001$. A Tukey post-hoc test revealed that both the write-an-explanation group ($M = 9$ min, 24 sec, $SD = 2$ min, 2 sec) and the write-a-focused-explanation group ($M = 10$ min, 27 sec, $SD = 2$ min, 40 sec) took significantly longer than both the read-an-explanation group ($M = 3$ min, 34 sec, $SD = 57$ sec) and the no-activity group ($M = 3$ min, 24 sec, $SD = 51$ sec, $ps < 0.001$). There was no significant difference in the time between the two writing groups ($p = 0.099$) nor between the two control groups ($p = 0.984$).

To better understand the relationship between lesson time and score on the posttest, we ran a Pearson correlation. There was a nonsignificant positive, weak correlation between the time taken on the lesson and score on the posttest, $r(124) = 0.15$, $p = 0.086$. Although nonsignificant, this correlation is confounded in that generative learning

activities take longer than no activity and generative learning activities also encourage deep learning more so than non-generative activities. Thus, the benefits of using generative learning could be due to the extra time or due to the active engagement prompted by the generative activity. To help differentiate the effects of more time and active engagement, we also ran the correlation within each group. There were no significant correlations (at $p < 0.05$) between study time and posttest performance for the write-an-explanation group ($r = -0.12$, $p = 0.516$), write-a-focused-explanation group ($r = 0.09$, $p = 0.607$), read-an-explanation group ($r = -0.04$, $p = 0.852$), and no-activity group ($r = -0.02$, $p = 0.915$). This suggests that self-selected study time is not a good predictor of test success.

Is There an Effect of Explanation Prompt on Perception of the Lesson?

After completing the posttest, all participants filled out a questionnaire that assessed their perception of the lesson itself. We ran multiple ANOVAs on postquestionnaire questions assessing differences in how the participants' perceived the lesson in enjoyment, interest, motivation, effectiveness, difficulty, and required effort. There were no significant differences ($ps > 0.05$) among the four groups on any of these questions. Overall, the different groups did not appear to affect how learners perceived the lesson, including their self-assessment of perceived intrinsic and germane cognitive load. This shows that the extra effort required for generative activities did not affect student perceptions of the lesson.

Experiment 2

Given the promising, but nonsignificant results of Experiment 1 with an immediate test, the main goal of Experiment 2 is to assess whether adding prompts to write an explanation (or write a focused explanation) in an online multimedia lesson affects performance on a delayed test of learning. As described in the "Introduction" section, delayed tests may be better able to detect meaningful learning outcomes, which is the proposed outcome of generative learning activities such as learning by explaining. Additionally, some previous research has demonstrated that the benefits of generative activities are not necessarily seen on an immediate test but do show up strongly on a delayed test (Brown et al., 2014; Dunlosky et al., 2013; Fiorella & Mayer, 2015, 2016). Thus, Experiment 2 is designed to examine this idea by replicating Experiment 1 but using a delayed test.

Method

Participants and Design

The participants were 131 undergraduates recruited from a university in Southern California through a psychology subject pool, in which they fulfilled a course requirement by participating. A power analysis of $\alpha = 0.05$, effect size = 0.65, and power = 0.80 demonstrated that a sample of this size would be sufficient. The mean age of the participants was 18.97 years ($SD = 1.64$), their average prior knowledge score was 4.89 ($SD = 2.01$) which is considered low, and 106 of them were women. The experiment used a one-way between-subjects design with 34 participants in the write-an-explanation group, 33 participants in write-a-focused-

explanation group, 32 participants in the read-an-explanation group, and 32 participants in the no-activity group.

Materials

The materials were the same as Experiment 1.

Prequestionnaire The prequestionnaire was the same as Experiment 1. The correlation between the self-report prior knowledge and more objective assessment (the number of checked statements) was moderate, $r(129) = 0.48$, $p < 0.001$. As in Experiment 1, our assessment of prior knowledge is based solely on the more objective prior knowledge items. The Cronbach's alpha for the checklist was 0.58. As in Experiment 1, an explanation for the sub-optimal internal consistency is that the checklist was designed to tap diverse situations in order to assess students' broad background knowledge rather than specific knowledge of greenhouse gases.

Multimedia lessons The lessons were the same as Experiment 1.

Posttest The posttest was the same as Experiment 1. Cronbach's alpha for the posttest test was 0.65. As in Experiment 1, the low internal consistency was probably due to the fact that the posttest assessed for knowledge about different aspects of the lesson and different levels of instructional objectives, not uniform knowledge of a singular point.

Postquestionnaire The postquestionnaire was the same as Experiment 1.

Apparatus

The apparatus was the same as in Experiment 1.

Procedure

The procedure was the same as Experiment 1 with one difference. In Experiment 2, participants completed the prequestionnaire and the lesson in the first session. After completing the lesson, participants left the lab and were instructed to come back exactly a week later. In session 2, a week later, participants completed the posttest and the postquestionnaire. The whole experiment took no more than 30 minutes total. We obtained IRB approval and adhered to guidelines for ethical treatment of human subjects.

Results and Discussion

Scoring the Posttest

The scoring of the posttest was done in the same way as Experiment 1. A Pearson correlation was run to determine the relationship between the two researchers' grading and found a strong correlation between the two researchers' point assignment, $r = 0.87$, $p < 0.001$.

Do the Groups Differ on Basic Characteristics?

A preliminary issue concerns whether random assignment produced groups that were equivalent on basic characteristics. There were no statistically significant differences among the groups for prior knowledge score, $F(3, 127) = 0.57, p = 0.572$; or age, $F(3, 127) = 0.10, p = 0.961$. However, there was a significant difference in gender composition, $\chi^2(3, N = 130) = 8.53, p = 0.036$, so gender was included as a covariate in subsequent analyses of posttest and postquestionnaire data in Experiment 2.

Does Writing Explanations During Pauses in a Multimedia Lesson Improve Learning?

As in Experiment 1, generative learning theory and the ICAP framework predict that the write-an-explanation group would do better than the no-activity group (hypothesis 1a) and the write-a-focused-explanation group would do better than the no-activity group (hypothesis 1b). Table 2 shows the posttest means and standard deviations for each group on the posttest. A one-way ANCOVA, with gender as a covariate, showed there was a significant main effect of explanation prompt, $F(3, 126) = 6.13, p = 0.001$. Additionally, gender did not have a significant effect in this ANCOVA, $F(1, 126) = 2.04, p = 0.156$, indicating that gender was not related to posttest performance. For clarity of the post-hoc tests, a Tukey post-hoc test was run using the ANOVA without the covariate of gender, $F(3, 127) = 6.83, p < 0.001$. The Tukey post-hoc test revealed that the write-an-explanation group ($M = 8.06, SD = 4.02$) scored significantly higher on the posttest than the no-activity group ($M = 5.41, SD = 3.78, p = 0.018$), in line with hypothesis 1a. However, the write-a-focused-explanation group ($M = 7.27, SD = 3.71$) did not perform significantly better than the no-activity group ($p = 0.167$), in contrast to hypothesis 1b. There was no significant difference between the read-an-explanation group ($M = 4.47, SD = 3.78$) and the no-activity group ($p = 0.729$).

Generative learning theory and the ICAP framework also predict that the write-an-explanation group would do better than the read-an-explanation group (hypothesis 2a) and the write-a-focused-explanation group would do better than the read-an-explanation group (hypothesis 2b). The same Tukey post-hoc test was used to test the second set of a priori predictions. The Tukey test revealed that the write-an-explanation group scored significantly higher than the read-an-explanation group ($p = 0.001$), in line with hypothesis 2a. Additionally, the write-a-focused-explanation group scored significantly higher than the read-an-explanation group ($p = 0.012$), in line with hypothesis 2b.

The ICAP framework also predicted that there should not be differences between the write-an-explanation group and the write-a-focused-explanation group (hypothesis 3), as they are both constructive activities. The same Tukey post-hoc test revealed that there were no differences between these two groups ($p = 0.811$), consistent with hypothesis 3.

As a summary, we conducted contrast analyses using the two models from Experiment 1. Again, the basic model was the best fitting with R-square = 0.909, $p = 0.047$, suggesting that the two explaining groups outperformed the two control groups.

Once again, we ran a t test to investigate the combined generative activity groups to the combined control activity groups to understand how generative strategies may benefit learners. There was a significant effect, $t(129) = 4.32, p < 0.001$, confirming the observation that participants in the generative activity groups ($M = 7.67, SD = 3.86$) outperformed participants in the control activity groups ($M = 4.94, SD = 3.35$).

These results suggest that engaging in generative activities, such as writing explanations during pauses in a multimedia lesson, is beneficial to learning compared to control activities. Additionally, this study displayed the importance of using a delayed test to assess the benefits of using generative learning strategies. A delayed test may serve as a better measure of deep understanding than an immediate test, as discussed further in the “General Discussion” section.

Are There Effects of the Amount of Time Spent on the Lesson?

A potential reason that participants may have done better in the explanation groups may have been that those groups required the students to spend more time on the lesson. First, an ANOVA was run to determine if there were significant differences in time spent on the lesson among the groups. A one-way ANCOVA showed there was a significant main effect of explanation type, $F(3, 125) = 136.76, p < 0.001$. The covariate, gender, did not have a significant effect in this ANCOVA, $F(1, 125) = 0.14, p = 0.709$, indicating that gender was not related to time taken on the lesson. For clarity of the post-hoc tests, Tukey tests were run using the ANOVA without the covariate of gender, $F(3, 126) = 138.02, p < 0.001$. Tukey tests revealed that the write-an-explanation group ($M = 9$ min, 42 sec, $SD = 2$ min, 45 sec) and the write-a-focused-explanation group ($M = 10$ min, 10 sec, $SD = 2$ min, 11 sec) both took significantly longer ($p < 0.001$) than the read-an-explanation group ($M = 3$ min, 15 sec, $SD = 44$ sec) and the no-activity groups ($M = 3$ min, 5 sec, $SD = 1$ min). There were no differences between the two write-an-explanation groups ($p = 0.757$) nor between the two control groups ($p = 0.985$). It appears that participants spent more time on the lesson when tasked with writing an explanation than when they were not asked to write an explanation, which is expected as generative activities take longer to complete. We conclude that added study time is part of the explanation writing treatments.

To better understand the relationship between lesson time and score on the posttest, we ran a Pearson correlation. There was a positive correlation between the time taken on the lesson and score on the posttest, $r(128) = 0.36, p < 0.001$. However, this correlation is confounded in that generative learning activities take longer than no activity and generative learning activities encourage deep learning more so than non-generative activities. Thus, the benefits to learning from generative learning could potentially be due to the extra time or due to the active engagement prompted by the activity. To help differentiate the effects of more time and active engagement, we also ran the correlation within each group. There were no significant correlations (at $p < 0.05$) between study time and posttest performance for the write-an-explanation group ($r = 0.05, p = 0.719$), write-a-focused-explanation group ($r = 0.116, p = 0.358$), read-an-explanation group ($r = 0.10, p = 0.412$), and no-activity group ($r = 0.10, p = 0.458$). This suggests that self-selected study time is not a good predictor of test success; however, it does not rule out time as a possible factor in the treatment effects in this or subsequent experiments.

Is There an Effect of Explanation Prompt on Perception of the Lesson?

After completing the posttest, all participants filled out a questionnaire that assessed their perception of the lesson itself. We ran multiple ANCOVAs on postquestionnaire questions assessing differences in how the participants’ perceived the lesson in enjoyment, interest, motivation, effectiveness, difficulty, and required effort. There were no significant differences ($ps > 0.05$) between the four groups on any of these

questions. Overall, the explanation type did not appear to affect how learners perceive the lesson, including their perceived intrinsic cognitive load and perceived germane cognitive load. As in Experiment 1, the writing prompts did not affect how students perceived the lesson, including how hard they thought the lesson or how much effort they put in.

Experiment 3

Experiments 2 found that asking students to write an explanation led to better posttest performance as compared to a no-activity control group or read-an-explanation control group, and that this pattern was clearly found on a delayed test but not on an immediate test (in Experiment 1). Given the potential importance of this finding, we sought to determine whether we could replicate this finding in Experiment 3, specifically if writing an explanation would be better than reading an explanation and engaging in no activity. Additionally, a new aim of Experiment 3 was to examine the effects of engaging in a less demanding form of explanation writing—rewriting a provided explanation in one’s own words—which might be more suitable for remote learning with inexperienced learners. As discussed, generating explanations for multimedia lessons with animations can be a cognitively demanding task. We are interested in understanding how a less cognitively demanding task, like rewriting an already given explanation, may benefit learners. Additionally, we were interested in understanding how an activity that would theoretically fall at the active level of the ICAP framework would compare to our constructive activities.

Method

Participants and Design

The participants were 128 undergraduates recruited from a university in Southern California through a psychology subject pool. They fulfilled a course requirement by participating. A power analysis of $\alpha = 0.05$, effect size = 0.65, and power = 0.80 demonstrated that a sample size of this large would be sufficient. The mean age of the participants was 18.95 years ($SD = 1.26$), the mean knowledge score was 4.55 ($SD = 2.13$) which is considered low, and 89 of the participants were women. The experiment used a one-way between-subjects design with 32 participants in the write-an-explanation group, 33 participants in the rewrite-an-explanation group, 32 participants in the read-an-explanation group, and 31 participants in the no-activity group.

Materials

The materials were mostly the same from Experiments 1 and 2. The only difference was that the write-a-focused-explanation group was replaced with a new rewrite-an-explanation group. The rewrite-an-explanation group was exactly the same as the read-an-explanation group, but on the slides that had the explanations displayed, a prompt was added that told participants to “Please rewrite this explanation in your own words” (as shown in the third column of Table 1). As in Experiments 1 and 2, participants wrote their explanations on a separate sheet of paper.

Prequestionnaire The prequestionnaire was the same as Experiments 1 and 2. The Cronbach's alpha for the background knowledge items was 0.62, which can be explained as in Experiments 1 and 2.

Multimedia lessons The lessons were the same as Experiments 1 and 2 except the rewrite-an-explanation group was included and replaced the write-a-focused-explanation group. The rewrite-an-explanation group consisted of a given explanation about the material in the previous slide with a prompt that asked participants to write the explanation in their own words. As can be seen in Table 1, the explanations in the rewrite-an-explanation group were exactly the same as in the read-an-explanation group.

Posttest The posttest was the same as Experiments 1 and 2. Cronbach's alpha for the posttest was 0.63, subject to the same explanation as in Experiments 1 and 2.

Apparatus

The apparatus was the same as in Experiments 1 and 2.

Procedure

The procedure was the same as in Experiment 2. We obtained IRB approval and adhered to guidelines for ethical treatment of human subjects.

Results and Discussion

Scoring for the Posttest

The posttest questions were scored in the same way as in Experiments 1 and 2. A Pearson correlation was run to determine the relationship between the two researchers' grading and found a strong correlation between the two researchers' point assignment, $r = 0.91$, $p < 0.001$.

Do the Groups Differ on Basic Characteristics?

A preliminary concern was whether random assignment produced groups that were equivalent on basic characteristics. There were no statistically significant differences among the groups on prior knowledge score, $F(3, 124) = 0.10$, $p = 0.958$; age, $F(3, 124) = 0.31$, $p = 0.815$; or gender composition, $\chi^2(3, N = 128) = 4.36$, $p = 0.225$. We can conclude that the groups were equivalent in the basic characteristics of prior knowledge, age, and gender composition.

Does Writing or Rewriting Explanations During Pauses in a Multimedia Lesson Improve Learning?

Generative learning theory and the ICAP framework predict that the write-an-explanation group would do better than the no-activity group (hypothesis 1a) and that the rewrite-an-explanation group would do better than the no-activity group (hypothesis 1c). Posttest means

and standard deviations for each group are shown in Table 2. A one-way ANOVA showed there was a significant main effect of explanation type, $F(3, 124) = 8.18, p < 0.001$. Consistent with hypothesis 1a, a Tukey post-hoc test revealed that the write-an-explanation group ($M = 9.13, SD = 2.99$) significantly outperformed the no-activity group ($M = 4.90, SD = 3.43, p = 0.001$). Additionally, in line with hypothesis 1c, the rewrite-an-explanation group ($M = 7.79, SD = 3.40$) outperformed the no-activity group, ($p = 0.007$). There was no significant difference between the read-an-explanation group ($M = 6.81, SD = 4.04$) and the no-activity group ($p = 0.136$).

Generative learning theory and the ICAP framework also predict that the write-an-explanation group would do better than the read-an-explanation group (hypothesis 2a) and the rewrite-an-explanation group would do better than the read-an-explanation group (hypothesis 2c). The same Tukey post-hoc test also revealed that the write-an-explanation group outperformed the read-an-explanation group ($p = 0.044$), supporting hypothesis 2a. However, the rewrite-an-explanation group did not significantly outperform the read-an-explanation group ($p = 0.673$), not supporting hypothesis 2c.

ICAP framework predicts that the write-an-explanation condition would do better than the rewrite-an-explanation group (hypothesis 4), as writing an explanation is at a higher level of engagement (constructive) rewriting an explanation (active). In contrast to this hypothesis, the same Tukey post-hoc test revealed that there was no difference between these groups ($p = 0.413$).

As a summary, we conducted contrast analyses using the same two models as in Experiment 1 and 2, except the weight given to the rewrite-an-explanation group replaced the weight given to the write-a-focused-explanation group. In Experiment 3, the basic model was the best fitting model with R-square = 0.937, $p = 0.032$. Again, this pattern suggests that the generative activity groups outperformed the control groups with the strongest difference for the write-an-explanation group over the no-activity group.

A t test was also used to compare the effect of the combined generative activities to the combined non-generative activities. The t test was significant, $t(126) = 4.09, p < 0.001$, supporting the observation that participants who engaged in generative activities ($M = 8.45, SD = 3.25$) outperformed participants who did not ($M = 5.87, SD = 3.85$).

This study displays how generative activities—even ones that help scaffold a student's learning such as rewriting an explanation—are effective in bolstering learners' understanding of material compared to activities that are not generative. We replicated the findings that writing an explanation was better than reading an explanation and doing nothing (control) on a delayed test. Additionally, this research expanded upon Experiment 2 by finding that rewriting an explanation was better than only being given the lesson. This research demonstrated that rewriting an explanation, typically considered at the active level of the ICAP model, has similar benefits as writing an explanation, typically considered at the constructive level of the ICAP model. The implications of this are discussed in the “General Discussion” section.

Are There Differences in the Amount of Time Spent on the Lesson?

In Experiments 1 and 2, the write explanation groups required more time than the control groups. To understand if this also occurred in this experiment, a one-way ANOVA was run to determine if there were significant differences in the time spent on the lesson between the groups. A one-way ANOVA showed that there was a significant main effect of explanation prompt, $F(3, 124) = 184.98, p < 0.001$. A Tukey post-hoc test showed that the write-an-

explanation group ($M = 10$ min, 17 sec, $SD = 2$ min 28 sec) and the rewrite-an-explanation group ($M = 11$ min 5 sec, $SD = 2$ min 17 sec) took significant longer ($p < 0.001$) than the read-an-explanation group ($M = 3$ min 32 sec, $SD = 54$ sec) and the no-activity group ($M = 2$ min 58 sec, $SD = 44$ sec). There was no significant difference between the write-an-explanation group and the rewrite-an-explanation group ($p = 0.227$) nor between the read-an-explanation group and the no-activity group ($p = 0.598$). Again, it took participants more time to complete the groups that had generative activities than the groups that did not have those activities. We conclude the need for more time is part of the treatment when students are asked to engage in generative activities during learning at their own pace.

As in Experiment 2, we explored the issue of whether study time was related to posttest performance. The overall Pearson correlation was positive, $r(126) = 0.26$, $p = 0.003$. As in Experiment 2, we also ran the same Pearson correlations within each of the four groups to understand if time is a variable that would have a significant impact on learning without the confound of the generative learning activity. There were no significant correlations (at $p < 0.05$) between study time and posttest performance for the write-an-explanation group ($r = -0.23$, $p = 0.212$), rewrite-an-explanation group ($r = -0.15$, $p = 0.397$), read-an-explanation group ($r = 0.042$, $p = 0.819$), and no-activity group ($r = -0.037$, $p = 0.845$). As in Experiments 1 and 2, this suggests that self-selected study time is not a good predictor of test success.

Is There an Effect of Explanation Prompt on Perception of the Lesson?

After completing the posttest, all participants filled out a questionnaire that assessed their perception of the lesson itself. We ran multiple one-way ANOVAs on postquestionnaire questions assessing differences in how the participants' perceived the lesson in enjoyment, interest, motivation, effectiveness, difficulty, and required effort. For all but one question, effectiveness, there were no significant differences ($ps > 0.05$) between the four groups. For the effectiveness rating, there was a significant main effect of explanation type, $F(3, 124) = 3.53$, $p = 0.017$. A Tukey post-hoc test revealed that those in the rewrite-an-explanation group ($M = 3.70$, $SD = 1.08$) reported the lesson as more effective than the no-activity group ($M = 2.97$, $SD = 1.08$), with no other pairwise differences. Overall, we conclude that the groups did not show substantial differences in learners' perceptions of these lessons, indicating that generative activities did not cause students to like the lessons less or experience greater perceived cognitive load.

General Discussion

Empirical Contributions

These experiments illustrated the benefit of engaging in generative strategies during pauses in a multimedia lesson. Experiments 2 and 3 found writing an explanation led to better learning over a delay than not engaging in a generative activity (such as reading an explanation or engaging in no activity). Additionally, Experiment 3 found that adding a scaffolded generative activity (rewriting a provided explanation) benefitted learning as compared to not engaging in a generative activity.

Experiments 2 and 3, in comparison to Experiment 1, showed the importance of using a delayed test when trying to gauge the benefits of generative learning strategies. When given an

immediate test, there were no significant differences among the groups, but when tested with a delayed test, the benefits of generative learning strategies were displayed. This demonstrates that the benefits of engaging in generative learning strategies are more likely to be seen after a period of time. This is in line with prior literature that the benefits of generative learning show up on a delayed test (Brown et al., 2014; Dunlosky et al., 2013; Fiorella & Mayer, 2015, 2016). Overall, this finding is consistent with the idea that meaningful learning outcomes—created through generative learning activities—are more coherent and better linked to prior knowledge, which allows for greater durability over time.

These experiments add to the generative learning literature in several ways. First, they demonstrated that generative learning activities can be successfully incorporated into computer-based learning with animations. They may also provide an explanation for the differences in research findings about the benefits of prompting explanations with animations. An immediate test may not always be sensitive enough to show the benefits of generating an explanation and thus, previous studies using an immediate test show mixed results as described in the Introduction. In contrast, Experiments 2 and 3 showed consistent evidence for the benefits of learning by explaining on delayed tests. It should be noted that some previous research discussed in the Introduction reported significant effects for learning by explaining on an immediate test, so a more nuanced conclusion is warranted.

Additionally, these experiments showed that learning by explaining may appear to involve different levels of the ICAP framework, all of which benefit learners when compared to simply going through a lesson. However, this research suggests that rewriting an explanation could belong at the constructive level for some learners because it can involve going beyond the presented information (e.g., through putting the statement in one's own words), whereas an activity such as reading text aloud or copying an explanation word for word—which were not part of the present study—belong at the active level and would not be expected to benefit deeper learning as strongly.

Furthermore, these experiments compared prompting explanations and rewriting explanations not only to a control group, but also to a group where participants are asked to read an explanation of the material. This adds to the literature on learning by explaining by demonstrating how important it is for multimedia learners to be active participants in their own learning rather than passive recipients of information. This work extends previous research in which learners generate explanations concurrently while studying static material such as printed text and illustrations, to a new situation in which learners generate explanations during pauses after short segments of annotated animations that are transitory in nature.

Lastly, this set of experiments illustrated that generative learning strategies did not change how learners perceived the lesson. In all experiments, adding a generative learning strategy did not have much of an effect on how enjoyable, how interesting, how motivating, how effective, how difficult, or how much effort was used by learners compared to control activities. This suggests that adding generative learning strategies is beneficial to learning without a cost to students' perceptions of the lesson.

Theoretical Contributions

The findings in these three experiments relate to the generative learning theory, as represented in the select-organize-integrate (SOI) model of cognitive processing during multimedia learning (Fiorella & Mayer, 2015, 2016; Mayer, 2014, 2020). Generating an explanation is one of the strategies that can help prime the cognitive processes of selecting, organizing, and

integrating during learning (Fiorella & Mayer, 2015, 2016). The explanation prompt is intended to encourage learners to apply all the key aspects of generative learning theory—selecting, organizing, and integrating—into a single explanation and thus leads to deeper learning of the material, as demonstrated by these experiments.

These experiments add to generative learning theory by demonstrating how prompting learners to generate an explanation lead to better learning, specifically from a multimedia lesson. The previous literature using generative learning strategies, specifically learning by explaining, with animations has been mixed, but the present research helps demonstrate that lasting benefits of learning by explaining with a multimedia lesson may need to be investigated with a delayed test. Furthermore, they show how students can still benefit from generative explanations even if the student requires more support in developing an explanation. By rewriting the explanation in Experiment 3, learners did not have to select the most relevant knowledge from the lesson, but they still had to engage in organizing and integrating it with their previous knowledge. Rewriting, even with a scaffolded form of selecting, was still beneficial to participants.

Additionally, these findings map onto the ICAP framework (Chi, 2009; Chi & Wylie, 2014; Chi et al., 2018), which posits that engaging in activities that require a higher level of engagement (i.e., the constructive level) results in better learning than engaging in no activities during learning (i.e., the passive level). Engaging in activities at the constructive level (i.e., writing an explanation) or even the active level (i.e., rewriting an explanation) in the ICAP framework led to deeper learning that persisted over a delay.

Practical Implications

These studies extend our understanding of how to apply the generative learning principle (Mayer, 2020) to a computer-based multimedia lesson that students might encounter in an online learning scenario. Learning remotely can be an isolating experience that primes a passive stance in learners, so it is worthwhile to incorporate effective prompts to engage in generative activities—including writing or rewriting an explanation after each section of an onscreen lesson. Thus, this research demonstrates how beneficial it can be for instructors to incorporate explanation prompts into their more passive lessons, especially if the goal of presenting the material to help learners develop a deeper and persisting understanding of the material. However, if the goal of the instruction is to perform well on an immediate test, prompting explanations may not be the most effective strategy.

Limitations and Future Directions

One limitation of this study is that the material used was one short lesson. When learning, most individuals engage with multiple lessons that build on top of one another that require more than an hour of learning. This research provides a foundational understanding for how writing an explanation and rewriting an explanation may benefit learners in a short, simple lesson, which may not generalize to classroom learning. Future research should investigate how generating an explanation and rewriting an explanation may benefit students over a long-term course in a class or school-like environment.

The measure of cognitive load was based on self-reports, so it reflects the learner's perception of cognitive load rather than an objective measure of cognitive load. It would be

useful to include objective measures of cognitive load in future studies in order to determine its role as a mediating variable. Similarly, it would be useful to have objective measures of cognitive processing during learning, which could serve as mediators between generative learning activities and learning outcomes.

Another limitation is that the generative activity groups required more time than the control group. Although this may suggest that time is the crucial factor leading the differences in the groups, there is an alternate explanation, as discussed earlier. It could be the fact that additional time is an integral part of the activity of explaining because writing an explanation takes longer than simply reading through a lesson. Although we attempted to separate the effects of time and effects of the activity in our studies, there is still a possibility that the benefits are due to time. To truly understand if time and explanation prompt are confounded, future research should conduct the same experiment but require all participants to spend the same amount of time on each slide. Although this is not ecologically valid and therefore subject to criticism as well, controlling for study time in this way would allow for a straightforward test of the effects of explanation prompts on learning outcome. Additionally, in some situations oral explanations could be used in place of written explanations and the control group could be asked to read the material twice to minimize time differences across groups, similar to how Chi et al. (1994) conducted their study.

A potential limitation to this study is that students may not have been engaging fully in explanation and instead created summaries of the material. Although summarizing is a type of generative learning strategy, according to the ICAP framework (Chi, 2009; Chi et al., 2018; Chi & Wylie, 2014), summarizing would only reach the active level rather than the constructive framework. If students engaged more in summarizing than explaining, the benefits of the activity may not be as strong as possible. Future research should investigate how to differentiate the effects of explaining and summarizing.

Lastly, a potential limitation of this study was that there was a limit on the amount of time participants could spend answering each question on the posttest. This time limit was used to encourage participants to spend time thinking about each question and write more than a simple response. However, for some items, this may not have been enough time to process the question and write everything down. Future research should investigate if there are different effects of prompting explanation when learners are allowed to spend more time on the posttest.

Future research should investigate the effects of writing an explanation on a more difficult lesson. This lesson, although novel to many of the participants, was fairly simple to follow. It may have been easy enough for students to follow and thus, most students were able to benefit from less support than they might require in a more difficult lesson. It might be the case that a more difficult lesson would show the differences between the write-an-explanation, write-a-focused-explanation, and the rewrite-an-explanation groups as these groups provide different amounts of support to learners.

Future research could include enough subjects to be able to examine the quality of the explanations produced by students in the rewrite group. Perhaps, the effects of rewriting would be stronger for students who generated explanations that reflected more changes to the presented explanation, which could indicate performance at the constructive level of the ICAP framework.

Another future direction of this research is to understand how writing an explanation may be impacting learners' mental models of the material. As Chi (2000) explained, generating an explanation may aid learning due to the fact that it can help students repair their mental models, either by filling gaps in understanding or correcting misconceptions. These experiments did not directly investigate the mental models participants held about greenhouse gases

prior to participating in the study and thus cannot draw conclusions about how the learners' mental models are being affected.

Funding This project was supported by Grant N00014162046 from the Office of Naval Research.

Declarations

Conflict of Interest The authors report no conflicts of interest.

References

- Atkinson, R. K., Renkl, A., & Merrill, M. M. (2003). Transitioning from studying examples to solving problems: Effects of self-explanation prompts and fading worked-out steps. *Journal of Educational Psychology, 95*(4), 774–783. <https://doi.org/10.1037/0022-0663.95.4.774>.
- Brown, P. C., Roediger, H. L., & McDaniel, M. A. (2014). *Make it stick: The science of successful learning*. Harvard University Press.
- Bisra, K., Liu, Q., Nesbit, J. C., Salimi, F., & Winne, P. H. (2018). Inducing self-explanation: A meta-analysis. *Educational Psychology Review, 30*, 703–725. [s10648-018-9434-x](https://doi.org/10.10648-018-9434-x).
- Chi, M. T. H. (2000). Self-explaining expository texts: The dual process of generating inferences and repairing mental models. In R. Glaser (Ed.), *Advances in instructional psychology: Educational design and cognitive science* (pp. 161–238). Lawrence Erlbaum Associates, Inc..
- Chi, M. T. H. (2009). Active-constructive-interaction: A conceptual framework for differentiating learning activities. *Topics in Cognitive Science, 1*(1), 73–105.
- Chi, M. T. H., Adams, J., Bogusch, E. B., Bruchok, C., Kang, S., Lancaster, M., Levy, R., Li, N., McEldoon, K. L., Stump, G. S., Wylie, R., Xu, D., & Yaghmourian, D. L. (2018). Translating the ICAP Theory of Cognitive Engagement into practice. *Cognitive Science, 42*(6), 1777–1832.
- Chi, M. T. H., Bassok, M., Lewis, M. W., Reimann, R., & Glaser, R. (1989). Self-explanations: How students study and use examples in learning to solve problems. *Cognitive Science, 13*(2), 145–182. [https://doi.org/10.1016/0364-0213\(89\)90002-5](https://doi.org/10.1016/0364-0213(89)90002-5).
- Chi, M. T. H., de Leeuw, N., Chiu, M. H., & LaVancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science, 18*(3), 439–477. [https://doi.org/10.1016/0364-0213\(94\)90016-7](https://doi.org/10.1016/0364-0213(94)90016-7).
- Chi, M. T. H., & Wylie, R. (2014). The ICAP framework. Linking cognitive engagement to active learning outcomes. *Educational Psychologist, 49*(4), 219–243.
- De Koning, B. B., Tabbers, H. K., Rikers, R. M. J. P., & Paas, F. (2010). Learning by generating vs. receiving instructional explanations: Two approaches to enhance attention cueing in animations. *Computers and Education, 55*(2), 681–691. <https://doi.org/10.1016/j.compedu.2010.02.027>.
- De Koning, B. B., Tabbers, H. K., Riker, R. M. J. P., & Paas, F. (2011). Improved effectiveness of cueing by self-explanations when learning from a complex animation. *Applied Cognitive Psychology, 25*(2), 183–194. <https://doi.org/10.1002/acp.1661>.
- Doctorow, M., Wittrock, M. C., & Marks, C. (1978). Generative processes in reading comprehension. *Journal of Educational Psychology, 70*(2), 109–118. <https://doi.org/10.1037/0022-0663.70.2.109>.
- Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). What works, what doesn't. *Scientific American Mind, 24*(4), 47–53. <https://doi.org/10.1038/scientificamericanmind0913-46>.
- Fiorella, L., & Mayer, R. E. (2015). *Learning as a generative activity: Eight learning strategies that promote understanding*. Cambridge University Press.
- Fiorella, L., & Mayer, R. E. (2016). Eight ways to promote generative learning. *Educational Psychology Review, 28*(4), 717–741. <https://doi.org/10.1007/s10648-015-9348-9>.
- Grabowski, B. L. (2004). Generative learning contributions to the design of instruction and learning. In D. H. Jonassen (Ed.), *Handbook of research on educational communications and technology* (2nd ed., pp. 719–743). Erlbaum.
- Johnson, C. I., & Mayer, R. E. (2010). Applying the self-explanation principle to multimedia learning in a computer-based game-like environment. *Computers in Human Behavior, 26*(6), 1246–1252. <https://doi.org/10.1016/j.chb.2010.03.025>.
- Katona, G. (1942). *Organizing and memorizing*. Columbia University Press.
- Leahy, W., & Sweller, J. (2011). Cognitive load theory, modality of presentation and the transient information effect. *Applied Cognitive Psychology, 25*(6), 943–951.

- Lee, H. W., Lim, K. Y., & Grabowski, B. L. (2008). Generative learning: Principles and implications for making meaning. In J. M. Spector, M. D. Merrill, J. van Merriënboer, & M. P. Driscoll (Eds.), *Handbook of research on educational communications and technology* (3rd ed., pp. 111–124). New York: Erlbaum.
- Lin, L., & Atkinson, R. K. (2013). Enhancing learning from difference visualizations by self-explanation prompts. *Journal of Educational Computing Research*, 49(1), 83–110. <https://doi.org/10.2190/ec.49.1.d>.
- Lin, L., Atkinson, R. K., Savenye, W. C., & Nelson, B. C. (2016). Effects of visual cues and self-explanation prompts: Empirical evidence in a multimedia environment. *Interaction Learning Environment*, 24(4), 799–813. <https://doi.org/10.1080/10494820.2014.924531>.
- Margulieux, L. E., & Catrambone, R. (2018). Finding the best types of guidance for constructing self-explanations of subgoals in programming. *The Journal of the Learning Sciences*, 28(1), 1–44. <https://doi.org/10.1080/10508406.2018.1491852>.
- Mayer, R. E. (1992). Cognition and instruction: Their historic meeting within educational psychology. *Journal of Educational Psychology*, 84(4), 405–412. <https://doi.org/10.1037/0022-0663.84.4.405>.
- Mayer, R. E. (2010). Merlin C. Wittrock's enduring contributions to the science of learning. *Educational Psychologist*, 45(1), 46–50. <https://doi.org/10.1080/00461520903433547>.
- Mayer, R. E. (2014). The cognitive theory of multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (2nd ed., pp. 43–71). Cambridge University Press. <https://doi.org/10.1017/CBO9781139547369>.
- Mayer, R. E. (2020). *Multimedia learning* (3rd ed). Cambridge University Press.
- Mayer, R. E., & Wittrock, M. C. (2006). Problem solving. In P. Alexander, P. Winne, & G. Phye (Eds.), *Handbook of educational psychology* (pp. 287–303). Erlbaum.
- McEldoon, K. L., Durkin, K. L., & Rittle-Johnson, B. (2013). Is self-explanation worth the time? A comparison to additional practice. *British Journal of Educational Psychology*, 83(4), 615–632. <https://doi.org/10.1111/j.2044-8279.2012.02083.x>.
- Roediger III, H. L., & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological science*, 17(3), 249–255. <https://doi.org/10.1111/j.1467-9280.2006.01693.x>.
- Ryoo, K., & Linn, M. C. (2014). Designing guidance for interpreting dynamic visualizations: Generating versus reading explanations. *Journal of Research in Science Teaching*, 51(2), 147–174.
- Spanjers, I. A. E., van Gog, T., & van Merriënboer, J. J. G. (2010). A theoretical analysis of how segmentation of dynamic visualizations optimizes students' learning. *Educational Psychology Review*, 22(4), 411–423.
- Tobias, S. (2010). Generative learning theory, paradigm shifts, and constructivism in educational psychology: A tribute to Merl Wittrock. *Educational Psychologist*, 45(1), 51–54. <https://doi.org/10.1080/00461520903433612>.
- Wertheimer, M. (1959). *Productive learning*. Harper.
- Wittrock, M. C. (1974). Learning as a generative process. *Educational Psychologist*, 11(2), 87–95. <https://doi.org/10.1080/00461527409529129>.
- Wittrock, M. C. (1989). Generative processes of comprehension. *Educational Psychologist*, 24(4), 345–376. https://doi.org/10.1207/s15326985ep2404_2.
- Wong, A., Leahy, W., Marcus, N., & Sweller, J. (2012). Cognitive load theory, the transient information effect and e-learning. *Learning and Instruction*, 22(6), 449–457.
- Wouters, P., Paas, F., & van Merriënboer, J. J. G. (2008). How to optimize learning from animated models: A review of guidelines based on cognitive load. *Review of Educational Research*, 78(3), 645–675. <https://doi.org/10.3102/0034654308320320>.