Applied Cognitive Psychology, *Appl. Cognit. Psychol.* (2016)
Published online in Wiley Online Library (wileyonlinelibrary.com) **DOI**: 10.1002/acp.3227

Does Providing Prompts During Retrieval Practice Improve Learning?

MEGAN A. SMITH[1]*, JANELL R. BLUNT[2], JOSHUA W. WHIFFEN[2] and JEFFREY D. KARPICKE[2]

[1]*Department of Psychology, Rhode Island College, Providence, USA*
[2]*Department of Psychological Sciences, Purdue University, West Lafayette, USA*

Summary: The purpose of this investigation was to identify ways to prompt retrieval practice to make recall even more effective at producing meaningful learning. In two experiments, subjects read educational texts and practiced retrieval across two periods. During prompted retrieval, subjects were cued to explain and describe concepts from the text, whereas during free recall, subjects recalled as much of the material from the text as they could. A reading control condition was also included. Learning was assessed using both verbatim and higher-order questions 1 week later. Practicing retrieval improved learning relative to the control on both types of questions; however, whether subjects practiced free or prompted retrieval did not matter for learning. Subjects rated prompted retrieval as less enjoyable and interesting than the other retrieval conditions. Results demonstrate practicing retrieval promoted meaningful learning, and that subjects' initial retrieval success was highest when they used their own retrieval strategies during recall. Copyright © 2016 John Wiley & Sons, Ltd.

Practicing retrieval, or actively reconstructing knowledge, is a powerful way to promote student learning. More importantly, retrieving information promotes meaningful learning and transfer, not just rote memorization (e.g., Butler, 2010; Carpenter, 2012; Jensen, McDaniel, Woodard, & Kummer, 2014; Karpicke & Aue, 2015), increasing the value of retrieval practice as a learning strategy for educational purposes. Much of the past research on retrieval practice has examined learning benefits from taking tests. For this reason benefits from retrieval practice have been (and often still are) referred to as the testing effect. Yet it is not the act of taking a test that is important, but rather processes that tend to occur during testing. Specifically, it is the act of retrieving knowledge from memory that promotes learning. Retrieving knowledge can occur during testing, but can also be implemented in a wide range of activities (e.g., Blunt & Karpicke, 2014; Karpicke, Blunt, Smith, & Karpicke, 2014) that we refer to as *retrieval-based learning activities*. In order for students to use retrieval practice as a learning tool, they simply need to set their materials aside and practice reconstructing them.

There are many different ways to design retrieval-based learning activities, but practicing free recall of information is an especially effective method. During free recall, students set aside their study materials and freely reconstruct as much of the material from memory as possible. Practicing free recall allows learners to construct their own organizational structure and then use that structure during retrieval practice (see Raaijmakers & Shiffrin, 1981), and past research has suggested practicing free recall may improve learning more than other forms of retrieval practice such as answering short-answer questions (e.g., Glover, 1989). In addition to promoting student learning, free recall is a relatively practical way to engage in retrieval practice because neither students nor teachers need to prepare additional materials. Further, students do not seem to need training to engage in free recall. Students can simply set aside their textbooks or notes and practice freely recalling information.

*Correspondence to: Megan A. Smith, Department of Psychology, Rhode Island College, 600 Mount Pleasant Ave, Providence, RI 02908, USA.
E-mail: msmith1@ric.edu

There are a few things that need to occur for retrieval practice to be effective (Karpicke, Lehman, & Aue, 2014). First, students need to successfully retrieve information. If students are unable to reconstruct material during retrieval practice, then the activity is unlikely to promote much learning (see Karpicke et al., 2014). Similarly, if students miss a fair amount of important material during recall, they are not likely to benefit from recall as much, compared to when retrieval practice is more complete. It is true that unsuccessful retrieval attempts have been shown to improve learning when feedback is provided (e.g., Kornell, Hays, & Bjork, 2009). However, differences in initial retrieval success can lead to different amounts of learning across retrieval-based learning conditions (Butler, Marsh, Goode, & Roediger, 2006; Kang, McDermott, & Roediger, 2007; Marsh, Agarwal, & Roediger, 2009), and feedback is not always enough to overcome differences in initial success (see Smith & Karpicke, 2014). Therefore, ensuring retrieval success is important even if feedback is provided.

The second thing that needs to occur for retrieval practice to be effective is that students need to reinstate the prior context during retrieval (Karpicke & Zaromb, 2010; Karpicke et al., 2014; Lehman, Smith, & Karpicke, 2014). In other words, students need to think back to a previous time when they learned information and retrieve what they remember from the context. There is some evidence that reinstating the prior context is the mechanism by which retrieval practice improves learning (Lehman et al., 2014). If a retrieval-based learning activity is to be effective it must include this important element.

Ensuring successful retrieval and context reinstatement can be somewhat challenging. For example, there are ways in which one could ensure successful retrieval, such as recalling a sentence immediately after reading it. However, massed retrieval, or removing all spacing between the learning episode and retrieval, does not necessitate reinstating the prior context, reducing learning benefits from retrieval practice (e.g., Carpenter & DeLosh, 2005; Karpicke & Bauernschmidt, 2011). Therefore, it may be difficult to keep retrieval success high while also requiring students to reinstate the prior context. Given this challenge, it is important to design and empirically test retrieval-based learning

strategies that enhance student learning and are applicable by teachers and students in academic settings. Using free recall as a learning strategy has been shown to promote learning but involves some risk because free recall provides very little support to help ensure successful retrieval. Further, improving success above typical levels achieved during free recall may make retrieval practice even more beneficial and potentially more efficient for students. However, if too much support is provided, students may be very successful but may not actively reinstate the prior context, and this scenario could harm the effectiveness of retrieval practice.

Providing prompts that require students to describe and explain portions of material may bolster the effectiveness of retrieval practice. Providing students with prompts for retrieval would likely offer support during retrieval without compromising context reinstatement. King (1992) found evidence that students benefited from answering open-ended 'describe' and 'explain' questions, relative to rereading their notes. Guiding college students to describe and explain parts of their study materials during retrieval may help them reconstruct the material more completely. In addition, prompts may be provided during initial retrieval but then subsequently removed so that students may practice freely reconstructing information on their own. Designing a retrieval-based learning activity in this way utilizes the idea of scaffolding (see Meyer, 1993; Wood, Bruner, & Ross, 1976).

There are reasons to expect prompted recall to produce more learning than standard free recall. Prompting students to describe and explain important material is likely to support successful retrieval. If students are prompted to retrieve information about a specific concept, they may be less likely to omit information that they know about the concept. During free recall, it is reasonable to expect that students will omit information that they would have been able to recall simply because they forget to reconstruct the information (i.e., information is available but not accessible without a cue; Tulving & Pearlstone, 1966). Further, a scaffolded approach may be particularly effective if prompted retrieval increases initial retrieval success before students transition to actively reconstructing the material on their own without prompts during later free recall practice.

However, it is also possible that prompted retrieval may not improve learning or may hurt learning relative to free recall. If prompts fail to guide students to describe and explain important aspects of the material and restrict students' ability to reconstruct the material on their own, then prompting may hurt performance. Providing prompts during recall first and allowing students to practice free recall during a second retrieval practice period may alleviate issues associated with restricted retrieval during prompted recall. Still, if prompting students to describe and explain during retrieval restricts the reconstruction of knowledge, instructors will want to instruct their students to use free recall instead of providing prompts.

The current article reports two large experiments investigating the effect of providing students with prompts during retrieval practice. Experiment 1 examined whether prompting students to explain or describe the material during retrieval practice would improve initial retrieval success and enhance learning on a one-week delayed test. Experiment 1 also asked whether scaffolding students' retrieval, by first providing prompts as support and then removing the prompts so that students freely recall on their own, would be a particularly effective way to promote student learning. Experiment 2 used a variant of the prompted retrieval task used in Experiment 1 and tested subjects on the Internet.

In Experiment 1, undergraduate students learned educationally relevant materials by reading a text, engaging in a learning activity, rereading the text, and engaging in a second activity. This particular retrieval practice technique is easy to instruct students to do on their own. Whether the learning activity involved free recall (a standard free recall activity; Karpicke & Blunt, 2011), prompted recall, or a combination of free and prompted recall was manipulated across conditions. A reading control condition was also included. Learning was measured on a final short-answer assessment given in a second session 1 week after the first session. The final test included verbatim questions, which assessed content stated directly in the texts, and higher-order questions that required students to make inferences and apply the knowledge they learned from the texts. Both types of questions were included to replicate previous work showing retrieval practice enhances performance on both verbatim and higher-order questions (e.g., Blunt & Karpicke, 2014; Butler, 2010; Johnson & Mayer, 2009; Karpicke & Blunt, 2011; McDaniel, Howard, & Einstein, 2009; Smith & Karpicke, 2014). In addition, students were asked to rate several aspects of the learning activities in the first session, including how much they thought they had learned (judgments of learning), how much they enjoyed the activity, how difficult the activity was, and how interesting the activity was.

If prompts improve initial retrieval success without sacrificing retrieval effort or context reinstatement, then engaging in prompted retrieval during both retrieval periods should enhance learning on the final assessment one-week later relative to practicing free recall during both periods. Furthermore, scaffolding retrieval by first providing prompts and then having students practice free recall may be particularly effective at producing learning. Alternatively, if prompts reduce initial retrieval, then the prompted conditions may produce less learning relative to standard free recall activities. Regardless, we expected all forms of retrieval practice would improve student learning relative to the reading control condition.

EXPERIMENT 1

Method

Subjects

Subjects were 150 Purdue University undergraduate students. All students participated for partial credit for an Introductory Psychology course.

Materials

Two science texts were adapted from Basca, Burke, Campbell, and Sherman (2012) and Basca, Burke, Garcia, and Sullivan (2012): *Energy Transfer* (252 words, Flesch-Kincaid reading

level of 7.8) and *How we Breathe* (219 words, Flesch-Kincaid reading level of 5.3). An example text is provided in the Appendix. Six prompts were created for each text. The prompts cued students to describe or explain main concepts from the text. Each prompt was intended to help the students to recall a portion of the text and referred to a roughly equal portion of the text. Together the six prompts were intended to cue the student to recall the entire text, and were always presented in the order that concepts appeared in the text. Retrieval prompts for one text are provided in the Appendix. We created 12 short-answer questions for the final assessment. There were two types of questions: verbatim and higher-order. Verbatim questions had answers taken directly from the text (see Smith & Karpicke, 2014). Higher-order questions required students to go beyond what was presented. In some cases, students needed to integrate facts from the text, and in others students needed to take what was learned and apply it to a new situation. These questions assess meaningful learning (see Airasian et al., 2001; King, 1992). Example questions of both types are provided in the Appendix.

Design

The experiment used a between-subjects design with five conditions. Free recall and prompted recall were factorially manipulated across two consecutive retrieval periods. Thus, four experimental retrieval practice conditions were created: free–free, prompted–prompted, free–prompted, and prompted–free. A fifth reading control condition was also included in which students read twice but did not practice retrieval. Thirty students were assigned to each of the five conditions. In all five conditions, students completed their procedure for two texts, and the order of text presentation was fully counterbalanced across students.

Procedure

Students were tested in small groups in two sessions spaced 1 week apart. During the initial learning phase, students read one text for 5 min, engaged in a recall activity for 7 min, made ratings about the previously completed learning activity, reread the text for 5 min, completed another recall activity for 7 min, and then made ratings about the previously completed recall activity. In some retrieval periods, students practiced free recall by typing as much information as they could remember. In other retrieval periods, students practiced prompted retrieval: they were given prompts to describe or explain parts of the text to help them recall the full text. Students then repeated this procedure for a second text. In the reading control condition, students played a video game instead of the recall activities, thus they engaged in two spaced reading periods.

During reading periods, students read the text on the computer for 5 min. They were instructed to study the text so that they could remember it later. During recall activities, students practiced retrieval on the computer for 7 min. Retrieval practice was manipulated based on the students' assigned condition. During free recall periods, students recalled as much information as they could from the text. During prompted recall periods, students were given six prompts to cue them to retrieve the full text. Each prompt asked students to explain or describe a section of the text. Each prompt was presented one at a time on the screen for 70 s, and students were told to type as much information as they could remember from the passage that was related to the prompt. When 70 s passed, the screen cleared, the next prompt appeared on the screen, and students repeated the procedure. This continued until students recalled with all six prompts. In the free–free condition, students practiced the free recall procedure during both retrieval practice periods. In the prompted–prompted condition, students practiced the prompted recall procedure during both periods. In the free–prompted condition, students practiced free recall during the first period and prompted recall during the second period. In the prompted–free condition, students practiced prompted recall during the first period and free recall during the second period. The prompted–free condition represents a scaffolded format of retrieval practice because students first retrieve with support (prompted recall) and then the support is removed (free recall).

During rating periods, students answered a series of questions: How well do you think you will remember this material in 1 week? How much did you enjoy this task? How difficult was this task? How interesting was this task? Questions were answered on a scale from 0 to 100. A 0 indicated they would not remember anything, the task was not enjoyable, not difficult, or not interesting. A 100 indicated they would remember all of the material, the task was extremely enjoyable, extremely difficult, or extremely interesting. Students saw ratings from 0 to 100 in increments of 10 on the screen, and they selected their response by clicking a corresponding button.

After students completed the entire procedure for one text, they repeated the procedure for a second text. Students completed the same condition for both texts. When students had finished the second text, they were dismissed and returned for a final assessment 1 week later. On the final test, students answered 12 questions for each text, six verbatim, and six higher-order questions. The order of questions was held constant across students, and students were tested over the texts in the order in which they originally learned them. Students responded by typing their answers on the computer, and they were given unlimited time to answer each question. After a question had been presented for 20 s, a button labeled 'Next' appeared on the screen, and students clicked the button to advance to the next question. This procedure ensured that students spent at least 20 s attempting to answer each question. Once students answered each question for both texts they were debriefed and thanked for their participation.

Scoring

All scorers were unaware of which student produced each response, and to which condition each student belonged. The texts were divided into 30 idea units. All recall protocols were scored by giving 1 point for each correctly recalled idea unit (Karpicke & Blunt, 2011). Each idea was marked as either recalled or not recalled. Two independent raters scored all initial recall protocols, and a third rater settled any disagreements.

For short-answer questions, 1 point was given for each correct response, .5 points were given for partially correct response, and 0 points were given for incorrect responses

or no response. Two independent raters scored all short-answer tests, and the raters agreed on 94% of items. Because the two sets of scores were extremely similar, the scores were averaged in cases of disagreement.

Results

Initial performance

The top panel of Table 1 shows proportion of correct idea units recalled during initial retrieval. Overall, performance improved from period 1 ($M = 0.46$) to period 2 ($M = 0.56$). However, there were performance differences based on retrieval activity. In period 1, students retrieved more idea units when they practiced free recall ($M = 0.50$) than when they practiced prompted recall ($M = 0.42$, $d = 0.54$, 95% CI [0.18, 0.91]).[1] The same pattern held for period 2; students retrieved more during free recall ($M = 0.62$) than during prompted recall ($M = 0.50$, $d = 0.78$, 95% CI [0.41, 1.15]). Prompted recall led to lower retrieval success relative to free recall. However, lower levels of retrieval success may have been because of lower output as opposed to lower levels of overall learning. Examining performance in the prompted–free and free–prompted conditions provides support for this idea. Although students in the prompted–free group produced fewer ideas than those in the free–free group during period 1 (.42 vs. .49), performance was the same during period 2 when both groups practiced free recall (.62 for both groups). Similarly even though students in the free–prompted group produced more ideas than those in the prompted–prompted group during period 1 (.51 vs. .41), performance was the same during period 2 when both groups practiced prompted recall (.50 vs. .51). These results suggest it is the retrieval activities themselves that are restricting what students recall, and it is possible that this occurred without altering learning across conditions.

Table 1. Proportion of ideas recalled during initial retrieval practice, and proportion of idea units recalled during initial retrieval practice based only on ideas needed to answer final short-answer questions (critical idea units) in Experiments 1 and 2

	Proportion recall		Proportion critical idea units	
	Period 1	Period 2	Period 1	Period 2
Experiment 1				
Free – Free	.49 (.16)	.62 (.18)	.50 (.18)	.64 (.20)
Prompted – Prompted	.41 (.13)	.51 (.12)	.50 (.17)	.59 (.13)
Prompted – Free	.42 (.18)	.62 (.19)	.50 (.22)	.67 (.20)
Free – Prompted	.51 (.14)	.50 (.09)	.56 (.17)	.57 (.11)
Experiment 2				
Free Recall	.35 (.16)	.47 (.20)	.41 (.21)	.54 (.24)
Prompted Recall	.36 (.15)	.43 (.18)	.45 (.18)	.52 (.20)

Note: Standard deviations are reported in parentheses.

[1] The data were analyzed by calculating effect sizes (d) between two conditions on the various dependent variables (see Cummings, 2012) and 95% confidence intervals around d (shown in brackets after the effect size d value), which were calculated using the Methods for the Behavioral, Educational, and Social Science (MBESS) package for R (Kelley, 2007). Analyzing the data in this way allowed us to examine not only the size of the difference between two conditions, but also the precision with which the effect was measured in our experiments.

Final assessment performance

Figure 1 shows the proportion correct on final verbatim questions and final higher-order questions by condition. Practicing retrieval improved performance on the final assessment 1 week later. Students in all four retrieval practice conditions performed better than students in the reading control condition on verbatim questions, all Fs $(1, 58) > 4.18$, and higher-order questions, all Fs $(1, 58) > 6.03$. The top panel of Table 2 shows effect sizes (d) comparing each retrieval group to the reading control group. An ANOVA was performed on the four retrieval conditions to see whether different types of retrieval practice influenced later performance on the assessment. However, for both verbatim and higher-order final assessment questions there were no differences between the different retrieval-based learning activities, both $Fs < 1$. Providing describe and explain prompts during retrieval practice did not lead to increased performance relative to other conditions. Further, scaffolding the prompts during retrieval practice (i.e., the prompted–free condition) did not lead to improved performance on the final assessment relative to other forms of practicing retrieval.

Additional analysis of initial recall

We conducted additional analysis to further explore the content of initial retrieval. The right column of Table 1 shows initial retrieval performance including only idea units that were required to answer the final short-answer questions, or 'critical idea units'. Idea units were coded based on whether the final short-answer questions required access to that particular idea unit. For example, the verbatim question 'What tiny, hair-like structures cover the inside of the airway?' corresponded to the idea unit 'Cilia cover the inside of your airways.' This resulted in 15 critical idea units for *Energy Transfers* and 16 for *How We Breathe*. Initial performance was then reanalyzed using only these idea units.

If during prompted recall students tended to retrieve only the critical idea units necessary to answer final questions, then analyzing initial performance as a function of only relevant idea units would reveal no differences in initial performance. However, if prompted recall restricted students' output, then the same pattern of results would be expected here as with the full set of idea units. Overall, the pattern of results with the critical idea units was the same as with the full set of idea units. Performance improved from period 1 ($M = 0.52$) to period 2 ($M = 0.62$). However, there were differences in performance based on retrieval activity. In period 1, students retrieved slightly more idea units when they practiced free recall ($M = 0.53$) than when they practiced prompted recall ($M = 0.50$, $d = 0.06$, 95% CI [−0.30, 0.42]). The same pattern held for period 2; students retrieved more during free recall ($M = 0.66$) than during prompted recall ($M = 0.58$, $d = 0.48$, 95% CI [0.11, 0.84]). Given these results, it is likely that prompted recall restricted students' output during retrieval.

Ratings of the learning activities

Table 3 shows judgments of learning, and Table 4 shows ratings of enjoyment, difficulty, and interest for each recall period across all five conditions. The computer failed to record ratings for one student in the reading control condition.

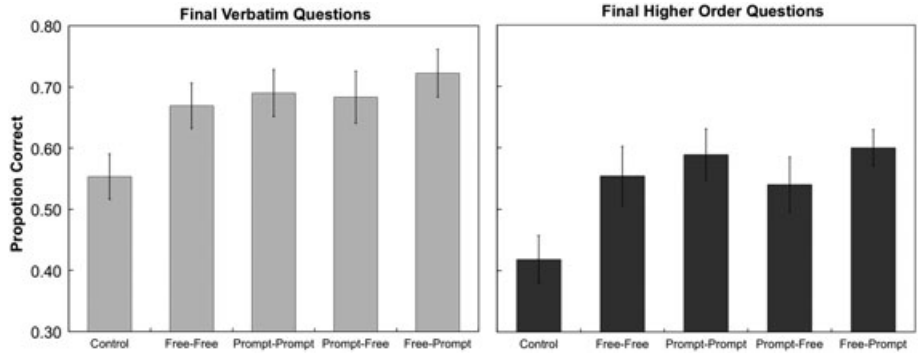

Figure 1. Proportion correct on final verbatim and higher-order short-answer questions in Experiment 1. Error bars represent standard errors of the mean

Table 2. Effect sizes (*d*) comparing retrieval practice conditions to the reading control conditions on the final short-answer assessment in Experiments 1 and 2

	Verbatim	Higher order
Experiment 1		
Free – Free	0.54 [0.02, 1.05]	0.76 [0.23, 1.28]
Prompted – Prompted	0.67 [0.14, 1.18]	1.02 [0.48, 1.56]
Prompted – Free	0.57 [0.05, 1.08]	0.64 [0.12, 1.16]
Free – Prompted	0.80 [0.27, 1.33]	1.03 [0.49, 1.57]
Experiment 2		
Free Recall	0.59 [0.32, 0.87]	0.50 [0.22, 0.77]
Prompted Recall	0.50 [0.22, 0.77]	0.47 [0.19, 0.74]

Note: 95% confidence intervals around *d* are shown in brackets.

Table 3. Mean judgments of learning in Experiments 1 and 2

	Period 1	Period 2
Experiment 1		
Reading Control	64.7 (19.9)	67.8 (17.0)
Free – Free	52.0 (21.7)	56.0 (25.4)
Prompted – Prompted	56.5 (21.0)	65.3 (22.5)
Prompted – Free	52.0 (24.2)	58.3 (21.6)
Free – Prompted	55.5 (22.3)	57.3 (23.0)
Experiment 2		
Reading Control	-----	70.6 (23.1)
Free Recall	-----	62.7 (23.7)
Prompted Recall	-----	66.5 (22.8)

Note: Standard deviations are reported in parentheses.

All four ratings were analyzed using separate 2 (period) × 5 (learning condition) factorial ANOVAs. The analysis of judgments of learning indicated there was a main effect of period, $F(1, 144) = 19.07$, $\eta_p^2 = .12$. Students increased their judgments of learning from period 1 ($M = 56.13$) to period 2 ($M = 60.95$), indicating they predicted learning from the additional reading and retrieval period. There was no effect of condition, $F(4, 144) = 1.69$, $\eta_p^2 = .05$, and no interaction, $F(4, 144) = 1.27$, $\eta_p^2 = .03$.

The analysis of ratings of enjoyment yielded a main effect of period, $F(1, 144) = 5.11$, $\eta_p^2 = .03$. Overall enjoyment of the learning activities decreased from period 1 ($M = 51.44$) to period 2 ($M = 49.03$). There was also a main effect of condition, $F(4, 144) = 5.01$, $\eta_p^2 = .03$. Overall, students found the reading control activity, which included playing Pac-Man as a distractor, more enjoyable ($M = 65.69$) than all four learning activities with retrieval practice, all Fs $(1, 57) > 7.57$. These effects were qualified by and a marginal interaction, $F(4, 144) = 2.42$, $p = .05$, $\eta_p^2 = .06$. On average, enjoyment decreased slightly from period 1 to period 2 in the reading control, recall–recall, and prompted–prompted conditions, mean difference (MD) = −1.38, $d = -0.14$, 95% CI [−0.50, 0.23] for the reading control group; $MD = -4.17$, $d = -0.24$, 95% CI [−0.60, 0.12] for the recall–recall group; $MD = -1.83$, $d = -0.16$, 95% CI [−0.51, 0.21] for the prompted–prompted group. The interesting conditions are the ones where students experience both free recall and prompted retrieval, because these indicate how students

Table 4. Mean ratings of enjoyment, difficulty, and interest of the initial learning activities in Experiments 1 and 2

	Enjoyment		Difficulty		Interest	
	Period 1	Period 2	Period 1	Period 2	Period 1	Period 2
Experiment 1						
Reading Control	66.4 (20.0)	65.0 (22.2)	22.4 (20.8)	21.2 (21.1)	60.2 (22.9)	57.4 (22.8)
Free – Free	51.3 (25.1)	47.2 (27.8)	42.7 (25.3)	38.2 (24.3)	52.2 (23.3)	48.3 (25.7)
Prompted – Prompted	42.3 (21.7)	40.5 (25.2)	30.8 (21.5)	28.0 (21.7)	40.7 (17.2)	35.8 (20.7)
Prompted – Free	44.5 (24.8)	47.2 (23.3)	38.5 (25.1)	40.3 (20.4)	45.7 (23.9)	45.7 (21.9)
Free – Prompted	53.2 (18.6)	45.8 (20.9)	39.8 (21.6)	43.8 (24.9)	52.5 (18.5)	50.7 (15.9)
Experiment 2						
Reading Control	-----	76.7 (24.9)	-----	26.4 (24.3)	-----	74.7 (25.9)
Free Recall	-----	67.5 (25.0)	-----	45.6 (25.6)	-----	69.7 (25.9)
Prompted Recall	-----	66.1 (20.9)	-----	49.0 (25.4)	-----	68.1 (23.3)

Note: Standard deviations are reported in parentheses.

judged different activities from period 1 to period 2. For students in the prompted–free condition, enjoyment increased slightly from period 1 to period 2, $MD = -2.67$, $d = 0.18$, [−0.18, 0.54]. However, the interaction was driven by students in the free–prompted condition. On average enjoyment decreased from period 1 to period 2 for students in the free–prompted condition, $MD = -7.33$, $d = -0.74$, [−1.14,−0.33]. It appears that students enjoyed practicing free recall more than prompted retrieval.

The analysis of difficulty ratings revealed no main effect of period, $F < 1$. However, there was a main effect of condition, $F(4, 144) = 4.70$, $\eta_p^2 = .15$. Students rated the reading control activity, including playing Pac-Man, as less difficult than the free–free, prompted–free, and free–prompted retrieval activities, all Fs $(1, 57) > 9.94$. There was no interaction, $F(4, 144) = 1.87$, $\eta_p^2 = .05$.

The analysis of interest ratings revealed a main effect of period, $F(1, 144) = 7.16$, $\eta_p^2 = .05$. Students decreased their ratings of interest in the learning activities from period 1 ($M = 50.23$) to period 2 ($M = 47.58$). There was also a main effect of condition, $F(4, 144) = 4.02$, $\eta_p^2 = .10$. Students rated the prompted–prompted learning activity as less interesting than the reading control activity, $F(1, 57) = 15.11$, $\eta_p^2 = .21$. There was no interaction, $F < 1$.

Discussion

In Experiment 1, students who practiced retrieval learned more than those who did not practice retrieval (the reading control condition) for both verbatim questions and higher-order questions. However, prompted retrieval did not lead to an improvement in initial retrieval performance and therefore did not improve learning measured 1 week later. In fact, during period 1, prompted retrieval led to worse performance than free recall. However, there is some evidence that, even though prompted retrieval restricted students' output, it did not hurt learning. In the prompted–free condition, students' performance during free recall in period 2 was equivalent to those in the free–free condition (both recalled 62% of the ideas). This occurred even though students in the prompted–free condition retrieved fewer ideas than students in the free–free condition during period 1 (42% vs. 49%). An analysis of the critical idea units needed to answer the questions on the final assessment still showed students retrieved more during free recall than prompted recall in period 2 (.66 vs. .58). Given that performance was the same on the final assessment across all retrieval practice groups regardless of prompted or free recall, it is possible that during prompted retrieval students retrieved the important information covertly, providing a learning benefit similar to more overt retrieval observed during free recall (see Smith, Roediger, & Karpicke, 2013).

Students may have had a more difficult time overtly producing what they retrieved because of the particular way prompted retrieval was done in Experiment 1. During prompted retrieval, six retrieval prompts were provided to students one at a time for a set amount of time, and students were asked to recall the information. Each time a new prompt appeared the old prompt and the information the students already recalled disappeared. It is possible that this procedure disrupted students' retrieval strategy and caused them to forget what they already typed into the computer and what they did not. Therefore, a second experiment was conducted with a different way of implementing prompted retrieval. Three learning conditions were compared in Experiment 2: free recall, prompted recall, and reading control. In Experiment 2, when subjects practiced prompted recall they saw all six prompts at the top of the screen throughout a single recall period. In addition, Experiment 2 was conducted with subjects recruited online, and we were able to obtain a larger and more diverse population than the group of subjects from Experiment 1.

EXPERIMENT 2

Method

Subjects

Four hundred and eight people were recruited online through a Human Intelligence Task (HIT) posted on Amazon Mechanical Turk. Subjects were restricted to those who were located in the United States, had a 95% HIT acceptance rate, and had completed at least 1000 HITs. Out of 408 subjects, 333 subjects returned for the second session. Of those 333, 17 subjects were excluded for not completing the HIT, either because they indicated they cheated during the experiment or because they failed to comply with the instructions by not providing any responses during the recall task. Two subjects were excluded because of a computer error in recording their final data. The final sample contained 314 subjects for analysis. Subjects participated in two online sessions 1 week apart and were randomly assigned to one of the three between-subjects conditions at the beginning of the first session, resulting in 100 subjects in the reading control group, 108 subjects in the free recall group, and 106 subjects in the prompted recall group. There were 184 females and 130 males, ages ranging from 18 to 75 ($M = 36.7$, $SD = 11.5$). Subjects received $1.50 to complete the first session, which took approximately 20 min, and $2.50 to complete the second session 1 week later, which took approximately 12 min.

Materials and design

The materials from Experiment 1 were used in Experiment 2. The experiment used a between-subjects design with three conditions: prompted recall, free recall, and reading control. Subjects were randomly assigned to learn one of the two texts in Experiment 2.

Procedure

Subjects completed Session 1 online at the time and location of their choosing and were informed that a second session would take place exactly 1 week after the first session. During Session 1, subjects first electronically signed the informed consent, entered demographic information, and read a detailed set of instructions about the procedure. Subjects then read a text, engaged in a recall (or distracter) activity, reread the text, engaged in the recall (or distracter) activity again, and made ratings at the end of the session. The procedures for each task were very similar to those from

Experiment 1 with minor changes. The primary change was that in the prompted recall condition, the six prompts were shown simultaneously at the top of the screen instead of one by one. During study periods, subjects spent a minimum of 2 min reading the text, and then a 'Continue' button appeared on the screen allowing subjects to advance when they were ready. Subjects spent an average of 2 min 31 s during reading phases. During recall periods, subjects spent a minimum of 4 min recalling before the 'Continue' button appeared on the screen. Subjects spent an average of 5 min 19 s on recall tasks. In the reading control condition, subjects played a video game for 4 min in between study periods. During the rating task at the end of Session 1 subjects were asked if they cheated during the experiment in addition to information about the learning activities.

Exactly 1 week after Session 1, subjects received an email through Mechanical Turk reminding them to complete Session 2 by logging on to the Mechanical Turk website. The short-answer procedure in Session 2 was identical to the one used in Experiment 1, except that subjects completed the task online and were asked at the end of the session whether they cheated. Subjects could complete Session 2 at any time and location within 24 h of receiving the email reminder.

Scoring

The procedures for scoring the data from Experiment 2 were similar to those used in Experiment 1. For the recall data, two independent raters scored 10% of the initial recall protocols. The raters agreed on 92% of scores. Given the relatively high inter-rater agreement, one of the raters scored the rest of the recall data. For the short-answer data, two independent raters scored 33% of the responses. The raters agreed on 82% of the answers. A third rater settled the disagreements and created a method for scoring the remaining data. One rater then scored the remaining short-answer data according to this scoring system.

Results

Initial performance

The bottom panel of Table 1 shows proportion of idea units recalled during initial retrieval. Again, performance generally improved from period 1 ($M = 0.36$) to period 2 ($M = 0.45$). However, the prompted recall group did not recall much more than the free recall group during period 1, $d = 0.11$, 95% CI [−0.16, 0.37], and recalled slightly less than the free recall group during period 2, $d = -0.18$, 95% CI [−0.44, 0.09].

Final assessment performance

Figure 2 shows the proportion correct on the final verbatim and higher-order questions. Once again, practicing retrieval improved performance on the final assessment 1 week later. Subjects in the two retrieval practice conditions performed better than those in the reading control group on both verbatim and higher-order questions, all Fs > 11.79. The bottom portion of Table 2 shows the effect sizes (d) comparing the free recall and prompted recall conditions to the reading control group. While differences were found between the retrieval practice groups and the reading control group, once again the differences between the retrieval practice groups were close to zero for both the verbatim, $d = 0.10$, 95% CI [−0.17, 0.37], and higher-order, $d = 0.06$, 95% CI [−0.20, 0.33], final assessment questions.

Additional analysis of initial recall

Table 1 shows the proportion of critical idea units recalled (those that were directly related to final short-answer questions). Replicating the results of Experiment 1, the pattern of results using only the critical idea units was the same as when the full set of idea units were used. Performance generally improved from period 1 ($M = 0.43$) to period 2 ($M = 0.53$), but performance did not greatly differ based on the type of retrieval practice implemented. During period 1, those in the free recall group recalled 0.41 of the idea units while those in the prompted recall group recalled 0.45 of the idea units, $d = 0.21$, 95% CI [−0.06, 0.48]. During period 2, those in the free recall group recalled 0.54 of idea units while those in the prompted recall group recalled 0.52 of idea units, $d = 0.09$, 95% CI [−0.18, 0.36].

Ratings of the learning activities

Table 3 shows judgments of learning, and Table 4 shows ratings of enjoyment, difficulty, and interest across all three conditions. Because of an error writing data to the database, ratings were not recorded for two subjects in the reading control condition, one subject in the free recall condition, and one subject in the prompted recall condition. For

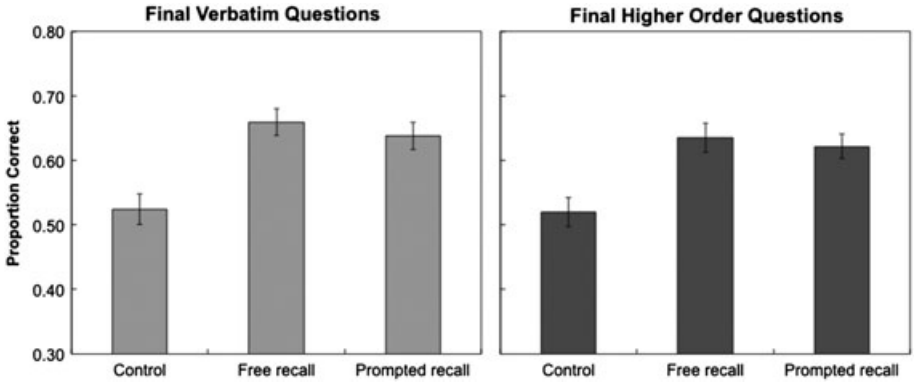

Figure 2. Proportion correct on final verbatim and higher-order short-answer questions in Experiment 2. Error bars represent standard errors of the mean

judgments of learning, subjects in the reading control group ($M = 70.6$) believed they learned more than those in the free recall group ($M = 62.7$, $d = 0.34$, 95% CI [0.06, 0.62]). Judgments of learning in the reading control condition were higher than those in the prompted recall condition ($M = 66.5$), but the confidence interval around the effect included zero, $d = 0.18$, 95% CI [−0.10, 0.46]. The difference between the two retrieval practice groups was close to zero, $d = 0.03$, 95% CI [−0.24, 0.30]. Analysis of the enjoyment ratings indicated that, as in Experiment 1, subjects found the reading control activity (which included playing a video game) more enjoyable ($M = 76.7$) than the free recall activity ($M = 67.5$, $d = 0.37$, 95% CI [0.09, 0.65]) and the prompted recall activity ($M = 66.1$, $d = 0.46$, 95% CI [0.18, 0.74]). Again the difference between the two retrieval practice groups was close to zero, $d = 0.09$, 95% CI [−0.18, 0.36]. Analysis of the difficulty ratings indicated that, as in Experiment 1, subjects reported the reading control activity to be less difficult ($M = 26.4$) than the free recall activity ($M = 45.6$, $d = 0.77$, 95% CI [0.48, 1.05]) and the prompted recall activity ($M = 49.0$, $d = 0.90$, 95% CI [0.61, 1.19]). The difference between the two retrieval practice groups again was close to zero, $d = 0.05$, 95% CI [−0.22, 0.32]. Finally, analysis of the interest ratings indicated that subjects reported the reading control activity to be slightly more interesting ($M = 74.7$) than the free recall activity ($M = 69.7$, $d = 0.19$, 95% CI [−0.08, 0.47]) and the prompted recall activity ($M = 68.1$, $d = 0.27$, 95% CI [−0.01, 0.54]). However, both confidence intervals around d included zero. The difference between the two retrieval practice groups was close to zero, $d = 0.07$, 95% CI [−0.20, 0.34].

GENERAL DISCUSSION

The two experiments reported here investigated whether providing prompts during retrieval practice would improve initial retrieval success and enhance the size of the retrieval practice effect on long-term retention. In Experiment 1, students were given six 'describe' or 'explain' prompts cueing main ideas from the text. Together, the six prompts were designed to cue students to recall the entire text. Students completed two prompted retrieval periods, two free recall periods (where no prompts were provided), or one prompted and one free recall period in the initial learning session. All retrieval practice conditions outperformed a reading control condition, but there were no discernible differences among the four retrieval practice conditions on long-term retention performance. In Experiment 2, prompted retrieval was implemented by providing all six prompts simultaneously on the screen throughout the recall periods. The results replicated Experiment 1: Both prompted and free recall conditions produced more learning than a reading control condition, but the two retrieval practice conditions did not differ.

Importantly, the present experiments also provide evidence that practicing retrieval improves higher-order learning. In both experiments, retrieval practice enhanced long-term retention of higher-order questions, although there were no discernible differences among the different forms of retrieval practice. Although some authors have questioned whether retrieval practice promotes complex learning (Van Gog & Sweller, 2015), there is now a great deal of evidence showing that retrieval practice enhances performance on long-term measures of higher-order meaningful learning (e.g., Blunt & Karpicke, 2014; Butler, 2010; Jensen, McDaniel, Woodard, & Kummer, 2014; Karpicke & Blunt, 2011; Smith & Karpicke, 2014; for a review see Karpicke & Aue, 2015). The present experiments contribute to the growing base of evidence demonstrating that retrieval practice enhances long-term complex learning.

Although there were no differences in learning among the different retrieval-based learning activities, there were differences in ratings among the conditions in Experiment 1. In general students reported enjoying the prompted retrieval periods less than the free recall periods in Experiment 1. The only condition during which students reported increased enjoyment from period 1 to period 2 was when they practiced prompted retrieval first and free recall second. Students rated the prompted–prompted retrieval condition as the least interesting. In Experiment 2 the rating differences between prompted recall and free recall were very close to zero. Comparing the ratings from Experiment 1 and Experiment 2, it seems that providing the prompts all at once on the top of the screen and allowing students to see all of their recall on the screen during prompted recall likely caused subjects to enjoy the prompted recall learning activity as much as they enjoyed the free recall learning activity. It seems that subjects simply did not enjoy answering the prompts one by one. However, the results being compared come from two separate experiments, and the populations of subjects were different (undergraduates and a more general population), so it is possible that the different pattern of results emerged for a different reason.

The finding that prompted retrieval did not lead to a greater learning benefit compared to free recall is surprising in light of the motivation behind the current investigation. Describe and explain prompts were originally expected to improve initial retrieval success. Describe and explain prompts should have provided support to bolster retrieval success without disrupting context reinstatement. Instead our prompts did not improve initial retrieval success. In fact, initial data suggest our describe and explain prompts restricted students' retrieval output in some situations, namely when the prompts were provided one at a time (Experiment 1). It seems that in Experiment 1 the prompts restricted students to only think about small pieces of information at once when presented one by one. If prompts restricted the way students retrieved information, then it is possible the prompts disrupted students' ability to create their own organizational structure and use that structure to retrieve information. In Experiment 2 we changed the prompted format so that students could see all prompts at the top of the screen during the full recall period. With this procedure, both prompted recall and free recall led to similar recall output. In fact, during the first recall period those in the prompted condition recalled slightly more than those in the free recall condition (.35 vs. .36 for the overall idea units, and .41 vs. .45 for the critical idea units). These data suggest that the prompted format used in Experiment 2 allowed subjects to create their own organizational structure and utilize it

during retrieval. Even still, this form of prompted retrieval did not lead to considerably higher levels of recall during the learning activities, and therefore did not greatly improve long-term learning.

Our prompts did not improve learning from retrieval practice beyond the levels found using the standard free recall method. However, we do not take these data to mean that all prompts will fail at increasing the power of retrieval practice as a learning strategy. An effective retrieval prompt should allow students to build upon their organizational structure. Most important for educational purposes, any form of support or prompting will not necessarily improve retrieval success. There are clear tradeoffs between initial retrieval success and students' ability to reinstate the prior context during retrieval. While it may seem intuitive to provide additional support by asking students to describe and explain different aspects of the material, support may not always have the intended effect.

Our results show that retrieval practice in many forms improved learning 1 week later. Most importantly, our retrieval-based learning activities improved both verbatim learning and higher-order meaningful learning as measured after a one-week retention interval. Therefore, practicing retrieval is a powerful learning tool that educators can encourage students to use both in the classroom and on their own.

ACKNOWLEDGEMENTS

Support during the writing of this article was provided by Grants from the National Science Foundation (DRL-1149363 and DUE-1245476) and the Institute of Education Sciences in the U.S. Department of Education (R305A110903). J.R.B. is supported by a National Science Foundation Graduate Research Fellowship (2012124747). The opinions expressed in this article are those of the authors and do not represent the views of the National Science Foundation, the Institute of Education Sciences, or of the U.S. Department of Education. We wish to thank Ian Barkdull, Cathrine Brattain, Tiniesha Jensen, Nikita Saoji, Michael Sikorski, and Caitlin Taylor for assistance with data collection and scoring, and Stephanie Karpicke for assistance with material creation. We also wish to thank the Learning Lab for constructive comments and feedback.

REFERENCES

Airasian, P. W., Cruikshank, K. A., Mayer, R. E., Pintrich, P. R., Raths, J., & Wittrock, M. C. (2001). In L. W. Anderson & R. Krathwohl (Eds.), *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. New York: Longman.

Basca, B., Burke, R., Campbell, L., & Sherman, D. (2012). In W. Gayle (Ed.), *Earth's changing surface: Student reference book*. Chicago, IL: Chicago Educational Publishing Company, LLC.

Basca, B., Burke, R., Garcia, D., & Sullivan, M. (2012). In W. Gayle (Ed.), *Human body in motion: Student reference book*. Chicago, IL: Chicago Educational Publishing Company, LLC.

Blunt, J. R., & Karpicke, J. D. (2014). Learning with retrieval-based concept mapping. *Journal of Educational Psychology*, *106*, 849–858.

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

Butler, A. C., Marsh, E. J., Goode, M. K., & Roediger, H. L. (2006). When additional multiple-choice lures aid versus hinder later memory. *Applied Cognitive Psychology*, *20*, 941–956.

Carpenter, S. K. (2012). Testing enhances the transfer of learning. *Current Directions in Psychological Science*, *21*, 279–283.

Carpenter, S. K., & DeLosh, E. L. (2005). Application of the testing and spacing effects to name learning. *Applied Cognitive Psychology*, *19*, 619–636.

Cummings, G. (2012). *Understanding the new statistics: Effect sizes, confidence intervals, and meta-analysis*. New York, NY: Routledge Taylor & Francis Group.

Glover, J. A. (1989). The "testing" phenomenon: Not gone but nearly forgotten. *Journal of Educational Psychology*, *81*, 392–399.

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.

Johnson, C. I., & Mayer, R. E. (2009). A testing effect with multimedia learning. *Journal of Educational Psychology*, *101*, 621–629.

Kang, S. H. K., McDermott, K. B., & Roediger, H. L. (2007). Test format and corrective feedback modify the effects of testing on long-term retention. *European Journal of Cognitive Psychology*, *19*, 528–558.

Karpicke, J. D., & Aue, W. R. (2015). The testing effect is alive and well with complex materials. *Educational Psychological Review*, *27*, 317–326.

Karpicke, J. D., & Bauernschmidt, A. (2011). Spaced retrieval: Absolute spacing enhances learning regardless of relative spacing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*, 1250–1257.

Karpicke, J. D., & Blunt, J. R. (2011). Retrieval practice produces more learning than elaborative study with concept mapping. *Science*, *331*, 772–775.

Karpicke, J. D., Blunt, J. R., Smith, M. A., & Karpicke, S. S. (2014). Retrieval-based learning: The need for guided retrieval in elementary children. *Journal of Applied Research in Memory and Cognition*, *3*, 198–206.

Karpicke, J. D., Lehman, M., & Aue, W. R. (2014). Retrieval-based learning: An episodic context account. In B. H. Ross (Ed.), *The psychology of learning and motivation, vol 61* (edn, pp. 237 – 284). San Diego, CA: Elsevier Academic Press.

Karpicke, J. D., & Zaromb, F. M. (2010). Retrieval mode distinguishes the testing effect from the generation effect. *Journal of Memory and Language*, *62*, 227–239.

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

King, A. (1992). Comparison of self-questioning, summarizing, and notetaking-review as strategies for learning from lectures. *American Educational Research Journal*, *29*(2), 303–323.

Lehman, M., Smith, M. A., & Karpicke, J. D. (2014). Towards an episodic context account of retrieval-based learning: Dissociating retrieval practice and elaboration. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*, 1787–1794.

Marsh, E. J., Agarwal, P. K., & Roediger, H. L. (2009). Memorial consequences of answering SAT II questions. *Journal of Experimental Psychology: Applied*, *15*, 1–11.

McDaniel, M. A., Howard, D. C., & Einstein, G. O. (2009). The read-recite-review study strategy: Effective and portable. *Psychological Science*, *20*, 516–522.

Meyer, D. K. (1993). What is scaffolded instruction? Definitions, distinguishing features, and misnomers. *National Reading Conference Yearbook*, *42*, 41–53.

Raaijmakers, J. G. W., & Shiffrin, R. M. (1981). Search of associative memory. *Psychological Review*, *88*, 93–134.

Smith, M. A., & Karpicke, J. D. (2014). Retrieval practice with short-answer, multiple-choice, and hybrid tests. *Memory*, *22*, 784–802.

Smith, M. A., Roediger, H. L., & Karpicke, J. D. (2013). Covert retrieval practice benefits retention as much as overt retrieval practice. *Journal*

of Experimental Psychology: Learning, Memory, and Cognition, *39*, 1712–1725.
Tulving, E., & Pearlstone, Z. (1966). Availability versus accessibility of information in memory for words. *Journal of Verbal Learning and Verbal Behavior*, *5*, 381–391.
Van Gog, T., & Sweller, J. (2015). Not new, but nearly forgotten: The testing effect decreases or even disappears as the complexity of learning materials increases. *Educational Psychology Review*, *27*, 247–264.
Wood, D., Bruner, J. S., & Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry*, *17*, 89–100.

APPENDIX: EXAMPLE TEXT, PROMPTS, AND SAMPLE FINAL ASSESSMENT QUESTIONS.

The correct response to questions is typed in italics along with each question. (V) denotes a verbatim final assessment question; (H) denotes a higher-order question

Energy transfer

Energy transfers from the sun
There are many planets in the solar system. Some are further away from the Sun than others. Mercury is the closest to the Sun and Neptune is the furthest from the Sun. The Earth is the third planet from the Sun, and it receives a steady supply of energy. The transfer of energy from the Sun to Earth is responsible for many changes that take place around us.

Clouds and precipitation
As the Sun heats up the Earth's water, some water evaporates and rises into the atmosphere. Eventually, it cools and condenses on tiny dust particles to form clouds. The size of droplets grows until they are so large that they fall as precipitation.

Wind
The Sun does not heat all parts of the Earth equally. The areas around the equator, the tropics, receive more of the Sun's energy and are warmer than other parts of the Earth. Unequal heating leads to the movement of air from cooler, higher-pressure, regions to warmer, lower-pressure regions. This movement is called wind.

Storms
Storms such as hurricanes also result from the transfer of the Sun's energy to Earth. As large bodies of water are warmed by the Sun, more and more of their water evaporates and eventually condenses into the air above. A huge amount of energy is released into the air as this occurs. The released energy sets the air in motion, spinning faster and wider until a hurricane forms. Hurricane season is typically 1 June to 30 November.

Prompts for prompted retrieval practice

(1) Describe the distance between the Earth, Mercury, Neptune, and the Sun and what this means for energy transfer.
(2) Describe how clouds are made.
(3) Explain why it rains.
(4) Describe how different parts of the Earth are heated by the Sun.
(5) Explain air movement on Earth.
(6) Describe storms and how they are made.

Final assessment questions

(V) Which parts of Earth receive the most heat from the Sun? *The areas around the equator*
(V) Unequal heating to the Earth causes what type of air movement? *Wind*
(H) Does the wind blow toward or away from the equator and why? *Toward, because it moves from cooler areas toward warmer areas and the equator is a warmer area since it receives more energy from the Sun.*
(H) Imagine you live in an environment with no dust particles. What would not form? *Clouds*