Learning With Retrieval-Based Concept Mapping

Janell R. Blunt and Jeffrey D. Karpicke
Purdue University

Students typically create concept maps while they view the material they are trying to learn. In these circumstances, concept mapping serves as an elaborative study activity—students are not required to retrieve the material they are learning. In 2 experiments, we examined the effectiveness of concept mapping when it is used as a retrieval practice activity. In Experiment 1, students read educational texts and practiced retrieval either by writing down as many ideas as they could recall in paragraph format or by creating a concept map (retrieval-based concept mapping). In Experiment 2, we factorially crossed the format of the activity (paragraph vs. concept map) and the presence or absence of the text (i.e., whether the activity involved repeated studying or retrieval practice). On a final test 1 week later that assessed verbatim knowledge and inferencing, both paragraph and concept map retrieval practice formats produced better performance than additional studying, but the 2 retrieval formats themselves did not differ. The results demonstrate the effectiveness of concept mapping when it is used as a retrieval practice activity and show that retrieval itself, rather than merely the act of writing, drives the benefits of retrieval-based learning activities.

Keywords: retrieval practice, concept mapping, learning, writing, study strategies

Learning is often viewed as a process that occurs primarily when people encode or study new material, and the best learning is thought to occur when students elaborate on what they are studying by forming meaningful connections and creating enriched knowledge structures. Retrieval, which occurs when students take tests, is viewed as an assessment of learning that occurred in prior study experiences but is not thought to create learning itself. In contrast to the latter assumption, a great deal of recent research has shown that practicing retrieval creates long-term, meaningful learning, sometimes even more learning than elaborative encoding activities (see Karpicke, 2012; Karpicke & Blunt, 2011). Our purpose in this article was to examine the effectiveness of two different retrieval practice formats: retrieving by writing information in paragraph format (a common way to induce retrieval practice; Roediger & Karpicke, 2006), and retrieving by creating what we refer to as *retrieval-based concept maps*.

The exact mechanisms underlying the effects of retrieval practice have not yet been specified, but the idea that retrieval practice

effects stem from elaborative study processes has recently been called into question. If elaboration were responsible for retrieval practice effects, then engaging in repeated retrieval should produce the same or similar effects as engaging in repeated elaborative studying. Recent research, however, has shown that repeated retrieval consistently produces greater levels of long-term learning than elaborative studying. For example, Karpicke and Smith (2012) found that retrieval practice produced superior long-term retention relative to imagery-based and verbal elaborative study methods (see too Karpicke & Blunt, 2011). Instead of elaborative study processes, the benefits of retrieval practice are currently thought to stem from processes involved recollecting the context of a prior learning episode (Karpicke, Lehman, & Aue, in press). Remembering what occurred at a particular place and time is not necessary during a semantic elaboration task, but it is inherent to retrieval practice. As evidence for this account of retrieval-based learning, Karpicke and Zaromb (2010) showed that having people intentionally retrieve a prior event (in their experiments, the completions to word fragments) led to greater subsequent retention relative to asking people to generate knowledge without thinking back to the past (e.g., completing fragments with the first words that came to mind). Therefore, an essential component of retrieval-based learning is what Tulving (1983) called being in an *episodic retrieval mode*, which refers to the act of thinking back to what occurred in a particular place and time in the past. Retrieval-based learning activities should be aimed at guiding students to intentionally recollect prior experiences.

In the experiments reported here, we examined the effectiveness of using one popular learning task, *concept mapping*, as a retrieval-based learning activity. Concept mapping is a graphic organizational technique in which students create node-and-link diagrams, where nodes represent concepts and links connecting the nodes represent relations among the concepts (see Figure 1; Novak & Gowin, 1984). Typically, students construct concept maps while
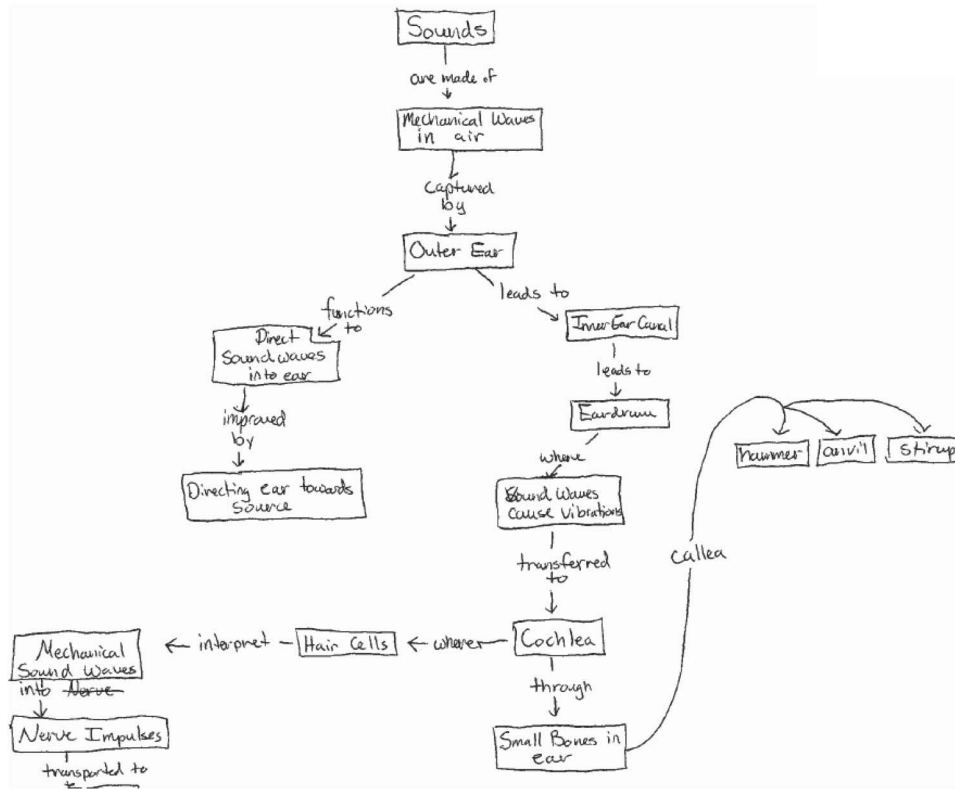
Figure 1. An example of a concept map created by a student in Experiment 1.

they view the materials they are learning. Although this presumably allows students to enrich the material by encoding meaningful relationships among concepts, when students create concept maps while viewing the to-be-learned materials, they are not required to practice retrieving the materials.

Recently, we carried out two experiments in which we directly compared the effectiveness of retrieval practice and elaborative studying with concept mapping (Karpicke & Blunt, 2011). Students studied educational texts on various science topics and either practiced retrieval or created concept maps of the texts. In the concept map conditions, students created concept maps while viewing the texts, whereas in the retrieval practice conditions, students read the texts and practiced retrieval by writing down as much of the material as they could recall without viewing the texts (a standard way of implementing retrieval practice for educational texts, which we refer to in this article as *paragraph format*; see Karpicke & Roediger, 2010; Roediger & Karpicke, 2006). The effects of these activities were assessed on final tests 1 week after the original learning phase. Practicing retrieval produced better long-term learning than elaborative concept mapping on final short-answer questions that assessed verbatim knowledge (items stated directly in the original text) and inferential knowledge (questions that required students to connect multiple concepts in the text). Furthermore, the benefits of retrieval practice were observed not only on short-answer questions but also on final assessments that involved creating a concept map of the material (Experiment 2 in Karpicke & Blunt, 2011). Thus, practicing retrieval produced more learning than creating concept maps when

the concept mapping activity was used as an elaborative study method.

Concept mapping could be used as technique to implement retrieval practice, and there are reasons to expect that concept mapping might serve as an effective retrieval-based learning activity. Specifically, concept mapping requires students to identify the main concepts in a text (Hay, Kinchin, & Lygo-Baker, 2008; Stewart, Van Kirk, & Rowell, 1979) and then identify how the concepts are related to each other, which helps focus students on the organizational structure of the material (Vanides, Yin, Tomita, & Ruiz-Primo, 2005). It is also assumed that creating concept maps helps student use their own prior knowledge to identify how concepts might be related (Novak, 1976). Thus, concept mapping is thought to promote not only students' verbatim knowledge and comprehension but also students' abilities to make inferences about what they are learning (Novak & Gowin, 1984).

Alternatively, there are also reasons to expect that concept mapping might not serve as an effective retrieval-based learning activity. When students freely recall material, they must adopt a retrieval strategy to guide their recall output (e.g., when recalling texts in paragraph format, students tend to recall in serial order, presumably to preserve the text structure; see Karpicke & Roediger, 2010). Concept mapping might require students to adopt an ineffective retrieval strategy or might disrupt students' default strategies, which could weaken the benefits of retrieval practice. It is also possible that asking students to retrieve knowledge in concept map format could introduce additional cognitive load during the process of retrieval, or the mapping task might function

as a secondary task that divides students' attention. Either factor could reduce the effectiveness of retrieval practice. Finally, retrieval-based concept mapping might produce learning that is equivalent to the learning afforded by practicing retrieval in paragraph format. This outcome would be expected if the organizational processing thought to occur during concept mapping were redundant with relational processing already afforded by paragraph recall and if both retrieval formats effectively allowed students to recollect the prior episodic context, which is the mechanism considered central to retrieval-based learning (Karpicke et al., in press; Karpicke & Zaromb, 2010).

In the present experiments, students read brief educational texts and practiced retrieval by writing in paragraph format or by creating concept maps. The effects of these retrieval practice activities were examined on a delayed short-answer test 1 week after the original learning phase. We also examined students' subjective experiences of the different activity formats. We were especially interested in students' judgments of learning (their predictions of how well they would perform in the future), but we also examined students' ratings of how interesting, difficult, and enjoyable the activities were. The inclusion of these metacognitive judgments allowed us to examine the correspondence between students' actual learning and their predicted performance, which is especially important to examine in light of claims that concept mapping represents "the most important metacognitive tool in science education" (Mintzes, Wandersee, & Novak. 1997, p. 424).

Experiment 1

Experiment 1 was a conceptual replication of Karpicke and Blunt (2011) with one important change: Rather than having students create concept maps while viewing texts, we had them create concept maps in the absence of the texts. Thus, we directly compared two different retrieval practice formats: concept mapping and paragraph recall. Students read and practiced retrieval of brief science texts. During concept map retrieval practice, students retrieved the material by creating a concept map, whereas during paragraph retrieval practice, students wrote as much of the material as they could recall in paragraph format. The students then made a series of metacognitive judgments (judgments of learning, interest, difficulty, and enjoyment). The effects of the two retrieval practice formats were assessed on a final test 1 week after the original learning phase that included both verbatim and inference short-answer questions.

Method

Subjects. Thirty-two Purdue University undergraduates participated in partial fulfillment of course requirements.

Materials. Two brief texts were selected from Cook and Mayer (1988, as described and used by Karpicke & Blunt, 2011). One text, "The Human Ear," had a sequential structure (Meyer, 1975), which means the text described a connected series of events and steps in a process (the sequence of events involved in the process of hearing). The other text, "Make-Up of Human Blood," had an enumeration structure, which means that the text listed and described a series of concepts (the properties of different blood components). The texts were 259 and 236 words in length, respectively.

Design. The two retrieval formats (concept map vs. paragraph) were manipulated within subject. Each student studied two texts and practiced retrieval of one text in concept map format and the other in paragraph format. The order of the two texts and the order in which students performed the two learning activities were counterbalanced across students.

Procedure. Students were tested in small groups in two sessions. During the learning phase (Session 1), students read one text for 5 min, recalled it for 10 min, reread it for 5 min, and recalled it again for 10 min in one of the two retrieval practice conditions. They then repeated this procedure for the other text and other retrieval practice condition.

Before completing the concept mapping retrieval practice condition, the students were instructed about the nature of the concept mapping activity. They were told that a concept map is a diagram in which concepts are represented as nodes that are linked together with words and phrases. The students were shown an example of a concept map selected from Novak (2005). Then, during recall periods, they were given a sheet of paper and told to recall the text by creating a concept map. Students were allowed to refer to the example concept map, but not to the text, throughout each 10-min recall period. Pilot testing showed that this was enough time for students to reach asymptotic levels of recall under these conditions.

In the paragraph retrieval practice condition, students saw a response box on a computer screen and were told to recall as much of the information from the text as they could by typing their responses on the computer during each 10-min recall period (see Karpicke & Roediger, 2010). Overall, the total amount of learning time was identical in the elaborative concept mapping and retrieval practice conditions.

At the end of each learning activity, the students were asked to predict how much of the material they would remember in 1 week (an aggregate judgment of learning) and to rate the enjoyment, difficulty, and interestingness of the activities. Students made their ratings on a scale from 0% to 100% in increments of 10 (0, 10, 20, . . . 80, 90, 100). At end of Session 1, after completing both activities, students indicated which retrieval practice format they preferred.

The students were dismissed and returned to the laboratory 1 week later for the final short-answer test, which included 10 verbatim questions and four inference questions per text. Examples of questions are shown in the Appendix. During the final test, each question remained on the screen for at least 20 s; at that time, a button labeled "Next" appeared on the screen, and students pressed the button to proceed to the next question. Students were encouraged to take as much time as needed to answer the questions. At the end of the second session, the students were debriefed and thanked for their participation.

Results

An initial analysis indicated that there were no differences among the counterbalancing orders, so the results have been collapsed across orders. There was a difference between texts such that initial and final performance was better on the "Make-Up of Human Blood" text than on the "Human Ear" text. However, text did not interact with any other factors in the experiment, so the results have been collapsed across texts.

Scoring. The texts were divided into 30 idea units for scoring purposes. Both the paragraph and concept map protocols were scored using the same criteria: Students were given 1 point for each idea unit recalled (Karpicke & Blunt, 2011; Karpicke & Roediger, 2010). On the final short-answer test, correct responses were given 1 point, and partially correct responses were given partial credit (e.g., .75, .50, or .25 points, depending on completeness of the response). Two independent raters scored all recall protocols and short-answer tests, and a third rater resolved all discrepancies to reach 100% agreement.

Learning performance. The left side of Table 1 shows performance during the learning phase in Experiment 1 (the proportion of idea units recalled in each condition). Collapsed across retrieval formats, the proportion of ideas recalled increased from Period 1 to Period 2 (.39 vs. .55), $t(31) = 10.34$, $d = 1.83$, 95% confidence interval (CI) [1.25, 2.39].[1] Students recalled more ideas in paragraph format than in concept map format. This pattern occurred in Period 1, $t(31) = 3.77$, $d = 0.66$, 95% CI [0.28, 1.04], and in Period 2, $t(31) = 5.52$, $d = 0.98$, 95% CI [0.55, 1.39]. We examined the differences in initial recall in the concept map and paragraph conditions in a post hoc analysis reported in a later section.

Final short-answer performance. Figure 2 shows performance on the final short-answer test that occurred 1 week after the initial learning phase. Performance was essentially equivalent in the concept map and paragraph retrieval practice format conditions. There were only small differences, slightly favoring the paragraph format over the concept map format, on the verbatim questions (.68 vs. .62), $t(31) = 1.07$, $d = 0.19$, 95% CI [−0.16, 0.54], and on the inference questions (.84 vs. .82), $t(31) = 0.41$, $d = 0.07$, 95% CI [−0.28, 0.42].

Subjective ratings. The right panel of Figure 2 shows students' judgments of learning, and Table 2 shows students' additional ratings of their experiences during the learning tasks. There were very small differences in students' judgments of learning, $t(31) = 0.26$, $d = 0.05$, 95% CI [−0.30, 0.39]; ratings of enjoyment, $t(31) = 0.31$, $d = 0.05$, 95% CI [−0.29, 0.40]; ratings of task difficulty, $t(31) = 0.33$, $d = 0.06$, 95% CI [−0.29, 0.40]; and ratings of the interestingness of the tasks, $t(31) = 1.04$, $d = 0.18$, 95% CI [−0.17, 0.53]. However, at the end of the initial learning phase, when students were asked to indicate which format they preferred, the majority of students preferred the paragraph format (20/32 students = 63%) to the concept map format (12/32 students = 37%).

Table 1
Proportion of Idea Units Produced in Each Learning Period in Experiments 1 and 2

Learning activity	Experiment 1		Experiment 2	
	Period 1	Period 2	Period 1	Period 2
Retrieval practice (no text)				
Concept map	.33 (.02)	.45 (.03)	.24 (.03)	.39 (.03)
Paragraph	.44 (.03)	.64 (.03)	.27 (.03)	.48 (.04)
Repeated study (text)				
Concept map	—	—	.48 (.02)	.58 (.03)
Paragraph	—	—	.53 (.04)	.62 (.03)

Note. Standard errors of the means are shown in parentheses.

Conditional analysis. In the next two sections, we report two sets of analyses aimed at exploring differences in recall in the concept map and paragraph conditions. The left portion of Table 3 shows the results of an analysis of the relationship between initial learning performance and final short-answer performance (collapsed across question type) in Experiment 1. In order to analyze the fate of idea units on the final test, we coded short-answer questions based on the idea unit or units required to answer the questions. Verbatim questions typically required access to a single idea unit (collapsed across texts, $M = 1.3$ idea units per verbatim question). For example, the question "What happens when hemoglobin combines with oxygen?" corresponded to the idea unit "Hemoglobin releases oxygen to the lungs." Inference questions required access to multiple idea units (collapsed across texts, $M = 2.3$ idea units per inference question). For example, the question "What would happen if blood did not contain white blood cells, and bacteria were introduced to the body?" relies on the following idea units: (a) "White blood cells are mainly disease fighters"; (b) "White blood cells digest bacteria and other foreign material"; and (c) "When there is an infection somewhere in the body, white blood cells move toward it."

We followed Tulving's (1964) method to analyze the correspondence in recall of individual idea units across two tests (see also Karpicke & Zaromb, 2010). C_1 refers to idea units produced in either Period 1 or 2 in the initial learning phase, and N_1 refers to idea units that were not produced in the initial learning phase. C_2 refers to short-answer questions correctly answered on the final short-answer test, and N_2 refers to questions not correctly answered on the final test. As shown in Table 3, the joint probability of recalling an idea initially and correctly answering a final short-answer question (C_1C_2) was greater in the paragraph condition than in the concept map condition, $t(31) = 3.68$, $d = 0.65$, 95% CI [0.26, 1.03]. Likewise, the probability of not recalling an idea but then correctly answering a final question (N_1C_2) was greater in the concept map condition than in the paragraph condition, $t(31) = 2.45$, $d = 0.43$, 95% CI [0.07, 0.79]. Together, these results reflect the fact that students initially recalled more ideas in paragraph format than they did in concept map format, yet the conditions produced equivalent levels of final short-answer performance. (We explore this pattern further in the analysis reported in the next section.) There was a small difference in *intertest forgetting* (the probability of recalling an idea but then failing to answer a short-answer question; C_1N_2) across conditions, with the paragraph condition showing slightly less forgetting, $t(31) = 1.67$, $d = 0.30$, 95% CI [−0.06, 0.65]. Finally, proportion of ideas not recalled or expressed on either test (N_1N_2) was greater in the concept map condition relative to the paragraph condition, $t(31) = 2.32$, $d = 0.41$, 95% CI [0.05, 0.77].

Initial recall and normative importance. Students might have produced fewer ideas during initial concept map recall relative to initial paragraph recall because they adopted different output strategies in the two tasks. We reasoned that students might selectively produce only "important" ideas under concept map

---

[1] We report standardized mean differences (*d*s) and 95% confidence intervals around the effect size estimates (see Cumming, 2012), which were calculated using the Methods for the Behavioral, Educational, and Social Sciences (MBESS) package for R (Kelley, 2007).
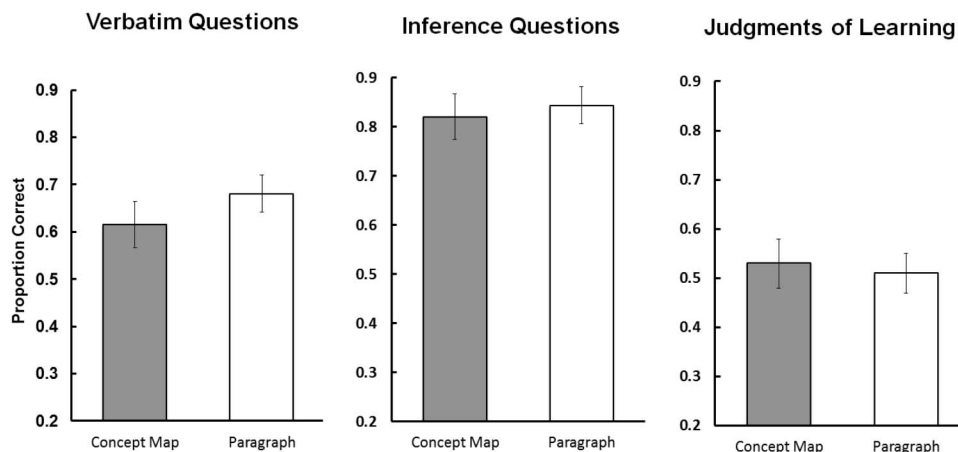
Figure 2. Final short-answer performance for verbatim questions and inference questions (left and middle panels), and judgments of learning (right panel) in Experiment 1. Error bars represent standard errors of the means.

conditions. To examine this possibility, we had 16 undergraduate students, who were not subjects in either experiment reported here, rate the importance of all 30 idea units from both texts, using a scale ranging from 1 (*not important at all*) to 5 (*very important*). The average importance rating was calculated for each idea unit, and the intraclass correlation among the average ratings was .78, indicating good interrater reliability (Shrout & Fleiss, 1979). If students selectively included important ideas in the concept map condition, then the average importance rating of recalled ideas should be greater in the concept map condition than in the paragraph recall condition. The results of our analysis confirmed this: Students tended to output ideas with higher normative importance ratings in the concept map condition ($M = 3.86$, $SE = 0.02$) than in the paragraph condition ($M = 3.76$, $SE = 0.02$), $t(31) = 3.20$, $d = 0.57$, 95% CI [0.19, 0.94]. Although the raw mean difference was small, the result was robust: for 28 of 32 students (88%), the mean normative importance of recalled ideas was greater in the concept map condition than in the paragraph condition. This analysis indicates that students might have covertly retrieved the same number of ideas in both retrieval practice conditions (which would still benefit learning; see Smith, Roediger, & Karpicke, 2013), but

students chose to include the relatively more important ideas when creating their concept maps.

Discussion

Experiment 1 showed that practicing retrieval in paragraph format or in concept map format produced approximately equivalent levels of performance on a delayed assessment of learning. Students also gave nearly identical subjective ratings to the two retrieval practice formats (judgments of learning and ratings of enjoyment, difficulty, and interestingness of the tasks), though students did tend to prefer the paragraph retrieval format relative to the concept map format. These results provide preliminary evidence that concept mapping may be an effective retrieval practice activity. Experiment 2 was carried out as a further investigation of the paragraph and concept map formats when used as either retrieval practice or repeated study activities.

Experiment 2

Experiment 2 was designed with two main purposes in mind. First, we sought to replicate Experiment 1 and generalize the

Table 2
Students' Ratings of Enjoyment, Difficulty, and Interestingness of the Learning Activities in Experiments 1 and 2

Learning activity	Experiment 1			Experiment 2		
	Enjoyment	Difficulty	Interest	Enjoyment	Difficulty	Interest
Retrieval practice (no text)						
Concept map	.49 (.04)	.46 (.04)	.49 (.04)	.39 (.06)	.55 (.05)	.42 (.06)
Paragraph	.51 (.04)	.47 (.04)	.53 (.04)	.40 (.04)	.54 (.06)	.49 (.04)
Repeated study (text)						
Concept map	—	—	—	.50 (.06)	.40 (.04)	.55 (.05)
Paragraph	—	—	—	.29 (.06)	.30 (.04)	.32 (.06)

Note. Students' ratings were indicated on a scale from 0 (*not at all*) to 100 (*totally*). Ratings were then converted to proportions. Standard errors of the means are shown in parentheses.

Table 3

Joint Probabilities Between Initial Performance and Final Short-Answer Performance in Experiments 1 and 2

Learning activity	Experiment 1				Experiment 2			
	C_1C_2	C_1N_2	N_1C_2	N_1N_2	C_1C_2	C_1N_2	N_1C_2	N_1N_2
Retrieval practice (no text)								
Concept map	.34 (.05)	.08 (.02)	.38 (.04)	.20 (.05)	.20 (.04)	.13 (.02)	.38 (.04)	.29 (.05)
Paragraph	.51 (.05)	.11 (.02)	.29 (.02)	.09 (.03)	.28 (.04)	.15 (.01)	.31 (.02)	.26 (.04)
Repeated study (text)								
Concept map	—	—	—	—	.28 (.04)	.26 (.03)	.24 (.02)	.21 (.03)
Paragraph	—	—	—	—	.25 (.04)	.31 (.02)	.18 (.02)	.26 (.04)

Note. Standard errors of the means are shown in parentheses. C_1 = items produced during the initial learning activity; N_1 = items that were not produced during the initial learning activity; C_2 = questions correctly answered on the final short answer test; N_2 = questions not correctly answered on the final short answer test.

results to a new set of text materials. Second, we included two new conditions to directly compare concept mapping and paragraph formats when they are used as retrieval practice activities (without the texts) with when they are used as repeated study activities (with the texts present; see Agarwal, Karpicke, Kang, Roediger, & McDermott, 2008; Agarwal & Roediger, 2011). Thus, in Experiment 2, we factorially crossed the presence of the material during the learning activity (text vs. no text) with the format of the learning activity (concept map vs. paragraph format). Our prediction was that the retrieval-based learning conditions would enhance long-term retention more than the repeated study conditions, even though students in the two conditions completed the exact same activities either with or without the materials in front of them. This result would support the idea that practicing retrieval, rather than the mere act of writing down the material in paragraph or concept map format, is the key to promoting long-term learning.

Method

Subjects. Eighty Purdue University undergraduates participated in partial fulfillment of course requirements. None of the students had participated in Experiment 1.

Materials. Two science texts were based on information in Stabler, Metz, and Gier (2011). One text, "Enyzmes," had a generalization structure (Meyer, 1975), which means the sentences in the passage provided clarification or examples of one main idea. The other text, "Domains of Life," had an enumeration structure (like the "Make-Up of Human Blood" text used in Experiment 1). The texts were 283 and 282 words in length, respectively.

Design. A 2 (activity format: concept map vs. paragraph) × 2 (learning condition: repeated study vs. retrieval practice) between-subjects design was used. There were four conditions, and 20 students were assigned to each condition. Each student completed the same activity for two texts, and the order in which the texts were presented was held constant across students.

Procedure. The procedure was similar to the one used in Experiment 1. Students were tested in small groups in two sessions, and each student was assigned to one of four learning conditions: (a) repeated study–concept map, (b) repeated study–paragraph, (c) retrieval practice–concept map, and (d) retrieval practice–paragraph. During the learning phase, students read one text for 5 min, engaged in a learning activity for 10 min, reread the text for 5 min, and completed the learning activity again for 10 min. Students then repeated the procedure for the second text. All

instructions were identical in the repeated study and retrieval practice conditions, and the total amount of learning time was equivalent in all conditions. The only difference was that in the repeated study conditions, students viewed the texts while they completed the learning activities, whereas in the retrieval practice conditions the students completed the activities without the texts (as in Experiment 1). Thus, students in the repeated study–concept map condition completed their concept maps while reading the texts (Karpicke & Blunt, 2011), and students in the repeated study–paragraph condition were instructed to write everything from the text on their paper in paragraph format (essentially copying the text). In both conditions, students were told to include all of the ideas from the texts. Texts were presented on the computer screen, and students completed the concept mapping or paragraph activities on paper. The subjective rating procedures and the final short answer test procedures were identical to those used in Experiment 1.

Results

An initial analysis indicated that there were no differences among the counterbalancing orders, and the levels of performance and patterns of results were the same for the two texts. Thus, the results have been collapsed across counterbalancing orders and texts.

Scoring. The texts were divided into 40 idea units for scoring purposes, and the scoring procedure used in Experiment 1 was used in Experiment 2. Two independent raters scored all recall protocols and short-answer tests, and a third rater resolved all discrepancies to achieve 100% agreement.

Learning performance. The right portion of Table 1 shows the mean proportion of idea units produced in each period in the initial learning phase in Experiment 2. Collapsed across conditions, the proportion of ideas produced increased from Period 1 to Period 2 (.38 vs. .52), $t(79) = 11.59$, $d = 1.33$, 95% CI [1.02, 1.62]. Students in the repeated study condition (who viewed the texts during the concept map and paragraph activities) produced more ideas than did students in the retrieval practice conditions. This was true for both activity formats in Period 1 (.50 vs. .25), $t(78) = 8.25$, $d = 1.85$, 95% CI [1.32, 2.36], and Period 2 (.60 vs. .44), $t(78) = 5.06$, $d = 1.13$, 95% CI [0.65, 1.60]. In the repeated study conditions, there were very small differences in the proportion of ideas produced in the concept map and paragraph formats in Period 1 ($M = 0.48$ vs. 0.53), $t(38) = 1.07$, $d = 0.34$, 95% CI

[−0.29, 0.96], or in Period 2 (M = 0.58 vs. 0.62), $t(38)$ = 0.70, d = 0.31, 95% CI [−0.31, 0.93]. However, as in Experiment 1, students tended to recall more ideas in the paragraph condition than in the concept map condition. There was a small difference in Period 1, (.27 vs. .24), $t(38)$ = 0.56, d = 0.18, 95% CI [−0.44, 0.80], and a larger difference in Period 2, (.48 vs. .39), $t(38)$ = 1.96, d = 0.62, 95% CI [−0.02, 1.25]. In a later section, we report an analysis of the role of idea unit importance in students' performance.

Final short-answer performance. Figure 3 shows performance on the final short-answer test 1 week after the initial learning phase. In general, students in the retrieval practice conditions (without the text available) performed better than students in the repeated study conditions (with the text available), but whether the activity was in concept map or paragraph format made little difference for long-term retention.

For verbatim questions, collapsed across activity formats, students in the retrieval practice (no text) conditions outperformed students in the repeated study (with text) conditions (.48 vs. .38), $t(78)$ = 2.22, d = 0.50, 95% CI [0.05, 0.94]. In the retrieval practice condition, there was a small difference between the paragraph and concept map formats, favoring the paragraph format, as was the case in Experiment 1 (.49 vs. .46), $t(38)$ = 0.50, d = 0.16, 95% CI [−0.46, 0.78]. However, in the repeated study condition, there was a larger difference between activity formats, favoring the concept map format over the paragraph format (.43 vs. .33), $t(38)$ = 1.59, d = 0.50, 95% CI [−0.13, 1.13]. This result supports the idea that creating a concept map while studying a text afforded elaborative encoding, as concept mapping enhanced long-term retention relative to essentially copying the text in the repeated study–paragraph condition.

The pattern of results was similar for the inference questions. Collapsed across activity formats, students in the retrieval practice conditions outperformed students in the repeated study conditions (.39 vs. .31), $t(38)$ = 2.07, d = 0.46, 95% CI [0.02, 0.91]. In the retrieval practice condition, there was almost no difference between the paragraph and concept map formats (.40 vs. .39), $t(38)$ = 0.27, d = 0.09, 95% CI [−0.53, 0.70]. Likewise, there was almost no difference between activity formats in the repeated study condition (.30 vs. .32), $t(38)$ = 0.30, d = 0.09, 95% CI [−0.52, 0.71], a result that is somewhat surprising in light of the advantage of concept mapping seen in the verbatim questions, as reported earlier.

Subjective ratings. The right panel of Figure 3 shows students' judgments of learning, which were made at the end of each task in the learning phase. Collapsed across activity formats, judgments of learning were higher in the repeated study conditions relative to the retrieval practice conditions (.56 vs. .48), $t(78)$ = 1.32, d = 0.30, 95% CI [−0.15, 0.74]. Although the effect was small in the present experiment, the finding that students believed they had learned more after repeatedly studying than after practicing retrieval is consistent with a wealth of prior work (e.g., Agarwal et al., 2008; Karpicke & Blunt, 2011; see Karpicke, 2012, for review). In the repeated study condition, students' judgments of learning were higher in the concept map condition than in the paragraph condition (.61 vs. .50), $t(38)$ = 1.46, d = 0.46, 95% CI [−0.17, 1.09]. In the retrieval practice condition, the opposite pattern occurred: students' judgments of learning were higher in the paragraph condition than in the concept map condition (.53 vs. .43), $t(38)$ = 1.27, d = 0.51, 95% CI [−0.13, 1.13].

Table 2 shows students' additional ratings of their subjective experiences in the learning tasks, and here we highlight a few findings displayed in the table. Students rated the repeated study–concept map condition as most enjoyable and the repeated study–paragraph task as least enjoyable (.50 vs. .29), $t(38)$ = 2.47, d = 0.78, 95% CI [0.13, 1.42], which is likely due to boredom associated with simply copying the text in the latter condition. The enjoyment ratings of the two retrieval practice conditions fell in between the ratings of the two repeated study conditions. A similar pattern was observed in the interest ratings: students rated the repeated study–concept map task as most interesting and the repeated study–paragraph task as least interesting (.55 vs. .32), $t(38)$ = 2.99, d = 0.94, 95% CI [0.28, 1.59], and the interest ratings of the two retrieval practice conditions fell in between the ratings of the two repeated study conditions. Finally, collapsed across activity formats, the retrieval practice tasks were rated as more difficult than the repeated study tasks (.54 vs. .35), $t(78)$ = 3.93, d = 0.89, 95% CI [0.42, 1.34].

Conditional analysis. The right portion of Table 3 shows the results of an analysis of the relationship between initial learning performance and final short-answer performance in Experiment 2. As in Experiment 1, short-answer questions were coded based on the idea unit or units required to answer the questions. Verbatim questions typically required access to a single idea unit (M = 1.5 idea units per verbatim question). For example, the question "What
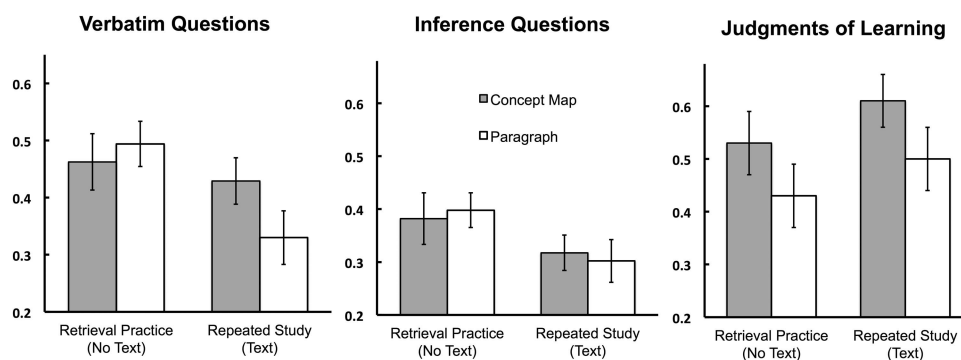

Figure 3. Final short-answer performance for verbatim questions and inference questions (left and middle panels), and judgments of learning (right panel) in Experiment 2. Error bars represent standard errors of the means.

do proteins lose at high temperatures?" corresponded to the idea unit "Proteins lose their structure at high temperatures." Inference questions required access to multiple idea units ($M = 2.9$ idea units per inference question). For example, the question "What happens to catalytic activity if temperature decreases?" relies on the following idea units: (a) "Catalytic activity is greatly affected by temperature"; (b) "Increasing temperature will also increase the amount of free energy"; (c) "This results in an increased rate of collision"; and (d) "[This] leads to a faster reaction time."

First, we analyzed the relationship between initial learning performance and final short-answer performance, collapsing across activity format (concept map vs. paragraph). As shown in Table 3, the probability of recalling an idea but then failing to answer a short-answer question (intertest forgetting; C_1N_2) was greater in restudy conditions than in the retrieval practice conditions, $t(78) = 7.25$, $d = 1.62$, 95% CI [1.11, 2.12]. Likewise, the probability of not recalling an idea but then correctly answering a final question (N_1C_2) was greater in retrieval practice conditions than in restudy conditions, $t(78) = 4.99$, $d = 1.12$, 95% CI [0.64, 1.58]. There were small differences across conditions in C_1C_2, $t(78) = 0.60$, $d = 0.13$, 95% CI [−0.31, 0.57], and N_1N_2, $t(78) = 0.88$, $d = 0.20$, 95% CI [−0.24, 0.64].

The pattern of results within the retrieval practice conditions (comparing the concept map format with the paragraph format) replicated the results of Experiment 1. The joint probability of recalling an idea initially and correctly answering a final short-answer question (C_1C_2) was slightly greater in the paragraph condition than in the concept map condition, $t(38) = 1.46$, $d = 0.46$, 95% CI [−0.17, 1.09]. Likewise, the probability of not recalling an idea but then correctly answering a final question (N_1C_2) was slightly greater in the concept map condition than in the paragraph condition, $t(38) = 1.63$, $d = 0.52$, 95% CI [−0.12, 1.14]. There was a small difference in intertest forgetting (the probability of recalling an idea but then failing to answer a short-answer question; C_1N_2) across conditions, with those in the paragraph condition showing slightly less forgetting, $t(38) = 0.91$ $d = 0.29$, 95% CI [−0.34, 0.91]. Finally, the proportion of ideas not recalled or expressed on either test (N_1N_2) was slightly greater in the concept map condition relative to the paragraph condition, $t(38) = 0.93$, $d = 0.29$, 95% CI [−0.33, 0.92].

Initial recall and normative importance. As in Experiment 1, 16 independent raters, who had not served as raters or subjects in Experiments 1 or 2, rated the importance of each idea unit in the two texts used in Experiment 2, using a scale from 1 (*not important at all*) to 5 (*very important*). The average importance rating was calculated for each idea unit, and the intraclass correlation among the average ratings was .80. In the retrieval practice condition, the mean importance rating of the idea units that students recalled was greater in the concept map condition ($M = 3.60$, $SE = 0.02$) than in the paragraph condition ($M = 3.48$, $SE = 0.03$), $t(38) = 3.00$, $d = 0.95$, 95% CI [0.29, 1.60]. However, in the repeated study condition, there was a smaller difference between the mean importance ratings in the concept map ($M = 3.54$, $SE = 0.02$) and paragraph conditions ($M = 3.50$, $SE = 0.02$), $t(38) = 1.36$, $d = 0.43$, 95% CI [−0.20, 1.05]. Thus, as in Experiment 1, when students practiced retrieval, they tended to include ideas with higher normative importance ratings in the concept map conditions than in the paragraph conditions, though this difference was much

smaller when students completed the activities with the materials in front of them.

Discussion

Experiment 2 showed that actively retrieving material during learning, either by creating concept maps or by writing the material in paragraph format, enhanced long-term retention more than completing the same activities in the presence of the materials (as study activities). Practicing retrieval produced more learning than repeated studying even though students re-experienced the entire set of material in the repeated study conditions, whereas students only re-experienced what they could recall in the retrieval practice conditions. Indeed, the proportion of ideas recalled in the retrieval practice conditions was lower than the proportion of ideas produced on the concept map or paragraph protocols in the repeated study conditions. It is important to note that the concept map and paragraph formats were equally effective as retrieval practice activities. As in Experiment 1, there were no additional benefits conferred by retrieval-based concept mapping beyond practicing retrieval in paragraph format. There was a small cost to retrieval-based concept mapping in the initial recall periods, on which students recalled fewer ideas in the concept map condition than in the paragraph condition. However, this cost was not seen on the final delayed assessments of long-term retention. Together with Experiment 1, the results of Experiment 2 show that concept mapping can serve as an effective learning task when it is implemented as a retrieval-based learning activity.

General Discussion

The purpose of the present experiments was to examine the effectiveness of retrieval-based concept mapping. The results show that the critical factor in retrieval-based learning is requiring students to think back to and recall material, while the format in which information is retrieved (concept map or paragraph format) did not much matter. We review three important findings from the present experiments in light of hypotheses proposed in the introduction.

First, concept mapping and paragraph formats were equally effective retrieval-based learning activities. When students created retrieval-based concept maps of the materials, there were no practical differences, relative to recalling in paragraph format, on delayed short-answer performance in Experiment 1 or 2. Furthermore, Experiment 2 showed that both activity formats produced retrieval practice effects: Students performed better on a final test when the initial activities required retrieval (in the absence of the texts) rather than studying or elaborating on the material (in the presence of the texts). This advantage of retrieval practice occurred even though students in the retrieval conditions produced less material during the initial learning activities relative to students in the repeated study conditions.

Second, retrieving in paragraph format produced greater long-term performance relative to restudying and rewriting the material in paragraph format. It is reasonable to wonder whether the locus of retrieval practice effects rests in the act of writing itself, rather than in the mental activity of retrieving and reconstructing knowledge. If this were the case, the repeated study–paragraph condition in Experiment 2 should have produced long-term performance

similar to that produced by the retrieval practice–paragraph condition. Indeed, because students were able to re-experience the entire set of material in the repeated study condition, one might expect that condition to outperform the retrieval practice condition. However, the opposite result occurred in Experiment 2, confirming that the act of retrieving knowledge itself, rather than the act of writing, drives the benefits seen in retrieval-based learning activities.

Third, students generally believed they had learned more after repeatedly studying than after practicing retrieval. This result is consistent with a wealth of prior research (see Karpicke, 2012) and is also broadly consistent with a cue utilization approach to metacognitive judgments (e.g., Koriat, 1997). According to this view, students base their judgments of learning in part on the ease of processing they experience during a learning activity. When students complete activities with the text in front of them, processing is fluent and easy, whereas when students complete activities without the text, they base their judgments on the ease or difficulty with which the material can be brought back to mind during retrieval. Thus, repeated study activities tend to afford overconfident judgments of learning, whereas retrieval practice leads to underconfident judgments. In Experiment 2, students rated concept mapping as more interesting and enjoyable than studying by copying the text in paragraph form, but students' ratings did not differ among concept map and paragraph formats when completed as retrieval activities. Despite some speculation that concept mapping might somehow promote or improve metacognitive performance (e.g., Mintzes et al., 1997), the present experiments offer no evidence that this is true (see too Karpicke & Blunt, 2011).

The key finding from the present experiments was that retrieval practice was equally effective when done in concept map or paragraph format. Students did not gain additional benefits by retrieving knowledge in concept map format relative to retrieving in paragraph format. Concept mapping is assumed to promote organizational or relational processing that should improve learning, but our results are consistent with the possibility that such organizational processing may be redundant with the processing people already engage in when practicing retrieval in other ways. Furthermore, practicing retrieval in concept map format did not impair learning relative to practicing retrieval in paragraph format. This finding suggests that the concept mapping task did not introduce extra cognitive load or divide attention in ways that were detrimental to learning. When students retrieved in concept map format, they tended to recall fewer ideas than when they retrieved in paragraph format, because they selectively reported ideas that were rated as most important. However, this was not detrimental to long-term learning either. Thus, the present experiments support the conclusion that concept mapping can indeed function as an effective learning activity when it involves practicing retrieval.

Conclusion

Retrieval practice is a powerful way to enhance long-term meaningful learning of educationally relevant content. The present results show that practicing retrieval, either by creating concept maps or by writing down the material in paragraph format, enhanced long-term learning more than completing the same tasks as study activities. The locus of these learning effects was in the act of retrieving knowledge, rather than the mere act of writing down

the material in paragraph or concept map format. It is important to note that the results show that concept mapping can indeed serve as an effective task when it is implemented as a retrieval-based learning activity. The key element for promoting meaningful learning was not the format of the activity; it was the requirement to engage in active retrieval practice during learning.

References

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

Agarwal, P. K., & Roediger, H. L., III. (2011). Expectancy of an open-book test decreases performance on a delayed closed-book test. *Memory, 19,* 836–852.

Cook, L. K., & Mayer, R. E. (1988). Teaching readers about the structure of scientific text. *Journal of Educational Psychology, 80,* 448–456.

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

Hay, D., Kinchin, I., & Lygo-Baker, S. (2008). Making learning visible: The role of concept mapping in higher education. *Studies in Higher Education, 33,* 295–311. doi:10.1080/03075070802049251

Karpicke, J. D. (2012). Retrieval-based learning: Active retrieval promotes meaningful learning. *Current Directions in Psychological Science, 21,* 157–163. doi:10.1177/0963721412443552

Karpicke, J. D., & Blunt, J. R. (2011, February 11). Retrieval practice produces more learning than elaborative studying with concept mapping. *Science, 331,* 772–775. doi:10.1126/science.1199327

Karpicke, J. D., Lehman, M., & Aue, W. R. (in press). Retrieval-based learning: An episodic context account. In B. Ross (Ed.), *The psychology of learning and motivation.* New York, NY: Elsevier.

Karpicke, J. D., & Roediger, H. L., III. (2010). Is expanding retrieval a superior method for learning text materials? *Memory & Cognition, 38,* 116–124. doi:10.3758/MC.38.1.116

Karpicke, J. D., & Smith, M. A. (2012). Separate mnemonic effects of retrieval practice and elaborative encoding. *Journal of Memory and Language, 67,* 17–29. doi:10.1016/j.jml.2012.02.004

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. doi:10.1016/j.jml.2009.11.010

Kelley, K. (2007). Methods for the behavioral, educational, and social science: An R package. *Behavior Research Methods, 39,* 979–984. doi:10.3758/BF03192993

Koriat, A. (1997). Monitoring one's own knowledge during study: A cue-utilization approach to judgments of learning. *Journal of Experimental Psychology: General, 126,* 349–370. doi:10.1037/0096-3445.126.4.349

Meyer, B. J. (1975). *The organization of prose and its effects on memory.* Amsterdam, the Netherlands: North-Holland.

Mintzes, J. J., Wandersee, J. H., & Novak, J. D. (1997). Meaningful learning in science: The human constructivist perspective. In G. D. Phye (Ed.), *Handbook of academic learning* (pp. 405–447). Orlando, FL: Academic Press. doi:10.1016/B978-012554255-5/50014-4

Novak, J. D. (1976). Understanding the learning process and effectiveness of teaching methods in the classroom, laboratory, and field. *Science Education, 60,* 493–512. doi:10.1002/sce.3730600410

Novak, J. D. (2005). Results and implications of a 12-year longitudinal study of science concept learning. *Research in Science Education, 35,* 23–40. doi:10.1007/s11165-004-3431-4

Novak, J. D., & Gowin, D. B. (1984). *Learning how to learn.* New York, NY: Cambridge University Press. doi:10.1017/CBO9781139173469

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

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin, 86,* 420–428.

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.

Stabler, L. B., Metz, M., & Gier, P. (2011). *AP Biology 2012.* New York, NY: Kaplan.

Stewart, J., Van Kirk, J., & Rowell, R. (1979). Concept maps: A tool for use in biology teaching. *The American Biology Teacher, 41,* 171–175. doi:10.2307/4446530

Tulving, E. (1964). Intratrial and intertrial retention: Notes toward a theory of free recall verbal learning. *Psychological Review, 71,* 219–237. doi:10.1037/h0043186

Tulving, E. (1983). *Elements of episodic memory.* New York, NY: Oxford University Press.

Vanides, J., Yin, Y., Tomita, M., & Ruiz-Primo, M. A. (2005). Using concept maps in the science classroom. *Science Scope, 28,* 27–31.

Appendix

Examples of Verbatim and Inference Questions Used in Experiment 1[a] and Experiment 2

Experiment 1. Sample questions from text on "Make-Up of Human Blood":

Verbatim question:
"What happens when hemoglobin combines with oxygen?"
(Sample answer: Oxygen is released to cells in the body.)

Inference question:
"What would happen to blood flow from a wound if the body did not have fibrin?"
(Sample answer: Blood would not clot, because fibrin is needed to form a meshwork of fibers that trap blood cells and aid in clotting.)

Experiment 2. Sample questions from text on "Enzymes":

Verbatim question:
"What are two forms of free energy?"
(Sample answer: Heat and kinetic energy.)

Inference question:
"What happens to catalytic activity if temperature decreases?"
(Sample answer: Catalytic activity decreases because increasing temperature increases the rate of molecular collision, which leads to a faster reaction time.)

[a] For a complete set of questions, see Karpicke & Blunt (2011).