American Psychological Association

Journal of Educational Psychology

ISSN: 0022-0663

2021, Vol. 113, No. 5, 986–997
http://dx.doi.org/10.1037/edu0000486

Comparing and Combining Retrieval Practice and Concept Mapping

Garrett M. O'Day and Jeffrey D. Karpicke
Purdue University

Retrieval practice enhances the learning of educational materials, and prior work has shown that practicing retrieval can enhance learning as much as or more than creating concept maps. Few studies have combined retrieval practice with other learning activities, and no prior work has explored whether concept mapping and retrieval practice might produce especially robust effects when the two activities are combined. In two experiments, students studied educational texts and practiced retrieval (by freely recalling the texts), created concept maps, or completed both activities. In the combined-activity condition, students studied and created concept maps prior to practicing retrieval. On a 1-week delayed assessment, practicing retrieval enhanced learning relative to creating concept maps. Surprisingly, combining concept mapping and retrieval practice failed to produce any benefit over retrieval practice without concept mapping, even though students in the combined condition spent substantially more time engaged with the materials than did students in single-activity conditions.

Educational Impact and Implications Statement

Can students benefit even more from retrieval-based learning if they first learn the material by using an elaborative study activity? To investigate this question, students read educationally relevant biology texts and then practiced retrieval by writing down everything they could remember from the text, constructed concept maps, or constructed a concept map before practicing retrieval. Surprisingly, combining concept mapping with retrieval practice failed to produce any benefit beyond retrieval practice alone, calling the alleged learning benefits of concept mapping into question.

Keywords: retrieval practice, concept mapping, learning strategies

Supplemental materials: http://dx.doi.org/10.1037/edu0000486.supp

Practicing retrieval is a potent way to enhance learning and long-term retention. Retrieval practice is most effective when learners successfully retrieve desired knowledge and when the retrieval task is also challenging. The challenging nature of retrieval is often described as retrieval effort and has been explained more formally as the processes involved in reinstating prior episodic contexts of learning (Karpicke, Lehman, & Aue, 2014). Retrieval success and retrieval effort often work in opposition: Some activities might make retrieval more successful but less effortful, whereas activities that increase retrieval effort are likely to produce lower levels of retrieval success. Therefore, a continuing challenge for researchers is to identify ways to promote retrieval success while preserving the effortful nature of retrieval. The present experiments examined one possible way to accomplish this by combining an elaborative study method, known as concept mapping, with free recall retrieval practice.

Elaborative encoding refers to processes that add features to memory representations. Elaboration enhances learning by promoting the encoding of relational information about similarities within a set of materials as well as item-specific information about the distinctiveness of individual ideas and events (Craik & Tulving, 1975; Hunt, 2012; Hunt & McDaniel, 1993; Nairne, 2006). Elaborative study methods are those that require learners to organize new material and distinguish unique features of terms and concepts in the service of future retrieval. There are two general reasons to expect elaborative study methods to work well when combined with retrieval practice. One reason is that elaborative encoding and retrieval practice might produce separate effects that sum together and enhance performance on a criterial test. Another possibility is that elaborative encoding might enhance initial retrieval success, thereby increasing the effectiveness of retrieval

This article was published Online First June 18, 2020.

Garrett M. O'Day and Jeffrey D. Karpicke, Department of Psychological Sciences, Purdue University.

This research was supported in part by grants from the James S. McDonnell Foundation (220020483), the National Science Foundation (DRL-1149363), and the Institute of Education Sciences in the U.S. Department of Education (R305A150546). The opinions expressed are those of the authors and do not represent the views of the National Science Foundation, the Institute of Education Sciences, or the U.S. Department of Education.

We thank Sofia Draper, Samantha Wagner, Samuel Carroll, Cassidy Walsh, Julianna Stevenson, and Breanne Kline for help collecting and scoring the data.

Materials and data are available at https://garrettoday.info/projects/CMRP_analyses.zip.

Correspondence concerning this article should be addressed to Garrett M. O'Day, Department of Psychological Sciences, Purdue University, 703 Third Street, West Lafayette, IN 47907-2081. E-mail: goday@purdue.edu

practice. Both possibilities provide strong reasons to expect that combining elaborative encoding and retrieval practice would enhance learning.

Although there has been a surge in research on retrieval practice during the last decade (Karpicke, 2017), the literature on the effects of combining any elaborative encoding activity with retrieval practice is quite small. Two studies examined the effects of combining retrieval practice with an elaborative encoding technique known as the keyword method (McDaniel & Pressley, 1984). The keyword method is a technique for learning vocabulary words that involves extracting part of the novel word and forming a mental image of the extracted keyword interacting with the translation (e.g.,"handsel" sounds like "hand" and means "payment"; learners would form an image of "hand" and "payment" to remember the meaning of "handsel"). Fritz, Morris, Acton, Voelkel, and Etkind (2007) had children, ages 12–13 years, learn English-German vocabulary words in a keyword method condition, a retrieval practice condition, or a condition that combined the two. On a criterial test 1 week after initial learning, practicing retrieval enhanced performance relative to studying with the keyword method, but there was little or no benefit of combining the two methods relative to retrieval practice alone. Karpicke and Smith (2012) had college students learn vocabulary words in a multitrial learning procedure, in which students studied and recalled the words until they had recalled each one at least once in the learning phase. Students repeatedly retrieved or repeatedly studied the words and learned them with or without the keyword method. Although the keyword method improved initial learning phase performance, it did not enhance retention on a 1-week delayed test. Practicing retrieval enhanced long-term retention, but echoing the results of Fritz et al., combining the keyword method with retrieval practice produced no benefit beyond practicing retrieval alone.

One study examined the effects of combining an elaborative study method and retrieval practice on learning from educational texts. Rummer, Schweppe, Gerst, and Wagner (2017) had college students study a text and take notes while studying. In one condition the students only studied the text and took notes. In a retrieval practice condition, students read the text, without taking notes, and then freely recalled the text two times, prior to rereading the text. In a combined condition, students read the text and took notes, recalled the text one time, and then reread their notes. Retrieval practice led to better performance relative to note taking on a final free recall test 2 weeks after the learning phase. The combined condition, however, produced no additional benefit beyond the retrieval practice condition.

The literature on combining retrieval practice and elaborative study methods is woefully sparse, but the combined conditions examined in previous research have not enhanced learning beyond retrieval practice alone. Limitations of these prior experiments, however, hinder the ability to know why previous studies failed to find benefits of combining retrieval practice and elaborative study methods. A chief reason to expect that combining elaboration and retrieval practice would enhance learning is that elaborative encoding would increase initial retrieval success, thereby enhancing the effectiveness of retrieval practice because learners would recall more items. Fritz et al. (2007) and Rummer et al. (2017), however, did not report initial retrieval success, so it is impossible to know whether elaborative encoding activities enhanced initial recall in their experiments. Rummer et al. sought to hold total time constant across their retrieval practice and combined note taking plus retrieval practice condition. Consequently, their combined condition engaged in less retrieval practice, recalling the material once, than did their retrieval practice condition, which recalled twice. Rummer et al.'s design failed to determine the effect of combining elaboration with retrieval practice relative to an equivalent dose of retrieval practice. It is also possible that the elaborative encoding tasks examined in prior work were not especially potent or that the benefits of those tasks were overpowered by the robust benefits of retrieval practice (as argued by Karpicke & Smith, 2012). The current literature does not provide definitive conclusions about the effectiveness of combining elaborative study and retrieval practice, evidence about possible ways to combine elaboration and retrieval practice that might prove effective, or explanations about why elaborative study and retrieval practice should hinder or promote learning.

The present experiments examined the effects of combining retrieval practice with concept mapping, an elaborative strategy that can be used when learning from educational texts. Concept mapping is an activity in which students create node-and-link diagrams that represent the key terms and relations among the terms within a set of material (Karpicke, 2018; Novak & Cañas, 2008). Concept mapping activities are used in a variety of ways in a range of educational settings. In general, concept mapping is thought to enhance learning because it requires learners to engage deeply with the material, to focus on the organizational structure of a set of material, and to produce elaborative connections among concepts (Nesbit & Adesope, 2006; Schroeder, Nesbit, Anguiano, & Adesope, 2018).

A few studies have directly examined the relative effectiveness of retrieval practice and concept mapping when each activity is done on its own. Karpicke and Blunt (2011) had students read an educational text and create a concept map of the text while viewing it, as an elaborative study activity, or practice retrieval by freely recalling it. Retrieval practice not only enhanced long-term retention relative to concept mapping on final short answer test that included verbatim and inference questions but also improved performance on a final assessment that required concept mapping. Blunt and Karpicke (2014) had students create concept maps or write responses in paragraph format while viewing texts, as elaborative study tasks, or complete the same activities without the texts present, as retrieval practice tasks. Practicing retrieval in either format, by creating concept maps or by writing paragraphs, enhanced retention on a 1-week delayed test relative to completing the same activities as elaborative study tasks, with the texts available during the tasks. Other studies have replicated these findings (Lechuga, Ortega-Tudela, & Gómez-Ariza, 2015), but no studies have examined whether combining retrieval practice and elaborative studying with concept mapping would represent an especially potent approach to improving long-term learning and retention.

There are multiple reasons to expect combining concept mapping with retrieval practice to enhance learning. Concept mapping is thought to enhance a learner's organizational representation of material (Novak & Cañas, 2008). Organizational or relational encoding is essential for free recall because it helps learners create a retrieval structure, a set of retrieval cues that guides search processes during recall (Hunt, 2012; Kahana, 2017; Raaijmakers & Shiffrin, 1981). Concept mapping might therefore support free recall, and increasing retrieval success during initial free recall should enhance long-term retention. A combination of concept mapping and retrieval practice might also enhance learning by promoting variable practice. Practicing skills or experiencing new materials in variable ways is known to enhance learning relative to practicing or experiencing material repeatedly in the same manner or format (Carpenter, 2017; Kang, 2017). Learners might benefit from combining concept mapping with retrieval practice because they would process the material in two different ways: by creating a visual-spatial representation of the terms and concepts as a concept map and then by actively reconstructing the material during retrieval practice.

Alternatively, combining concept mapping with retrieval practice might fail to improve learning or might harm learning relative to when students complete only one activity. The prior studies described above did not observe benefits of combining elaborative study methods with retrieval practice. These outcomes may have occurred because the elaborative studies methods examined in prior work were not especially effective, but it is also possible that the benefits of retrieval practice outweigh those of elaborative encoding. If that were generally true, then concept mapping may not confer additional benefit when combined with retrieval practice. Concept mapping might also disrupt students' default retrieval strategies during recall. When people freely recall texts, they tend to preserve serial order by recalling ideas in the order in which they had occurred in the texts (Karpicke & Roediger, 2010). Creating a concept map might emphasize a relational structure that differs from the serial order of ideas in the text, and this may harm initial free recall by interfering with students' default retrieval strategies.

In two experiments, students studied educational texts and created concept maps, practiced retrieval, or completed both activities in a combined condition. Concept mapping was done as an elaborative encoding activity, with students creating concept maps while they viewed the texts. During retrieval practice, students freely recalled as much of the information from the text as they could and then restudied the text and recalled it a second time. In the combined condition, students created a concept map prior to practicing retrieval. The effects of these learning activities were assessed in a second session 1 week after the initial learning phase. Specifically, students completed a short answer test that included verbatim and inference questions and then freely recalled the text. The experiments also examined students' metacognitive assessments of how much they believed they had learned (judgments of learning), which were solicited during the initial learning phase. The combined condition might produce the best learning if concept mapping encourages elaborative encoding that increases initial retrieval success, thereby making retrieval practice more effective. The combined condition might be best, even if concept mapping does not increase initial retrieval success, if elaborative studying and retrieval practice produce separate enhancements that sum together or if the two activities promote variable practice that enhances learning. It is also possible that the combined condition would confer no benefit relative to retrieval practice alone. This outcome would adhere to prior results in the literature and may occur if the benefits of retrieval practice overpower those of concept mapping. Additional analyses were carried out to explore the effects of combining retrieval practice and concept mapping in more depth and detail. Specifically, students' recall protocols were analyzed for their correspondence to the original order of the text following the technique developed by Asch and Ebenholtz (1962). We also conducted an analysis of the recallability of idea units across two recalls during retrieval practice following the method from Tulving (1964).

Experiment 1

Experiment 1 included three conditions. In a study only condition, students read an educational text in a single study period. The study only condition served as a baseline with which the other strategy conditions could be compared. In a retrieval practice condition, students read the text and then practiced retrieval by freely recalling it. In a concept map plus retrieval practice condition (CM + RP), students read the text and then, prior to practicing retrieval, they created a concept map as an elaborative study activity. Students indicated how well they thought they had learned the material (judgments of learning). One week later, students took a final short answer test, which included both verbatim and inference questions and then freely recalled the text.

Method

Subjects and design. Ninety undergraduate students (47 females, $M_{age} = 19.0$ years, $SD = 1.3$, range = 18–25) participated for partial fulfillment of course requirements. They were recruited from the introductory psychology subject pool at Purdue University. A between-subjects design was used with students randomly assigned to a study only condition ($n = 30$), a retrieval practice condition ($n = 30$), or a CM + RP condition ($n = 30$). The conditions did not significantly differ with respect to age, $F(2, 87) = 0.19$, $\eta_p^2 = 0.00$, or the proportions of men and women, $\chi^2(2) = 1.16$, $p = .56$. All subjects were fluent English speakers. This research was approved by the Purdue University's Institutional Review Board.

Materials and measures. All of the materials are provided in the online supplemental materials. In Experiment 1, students studied a text, titled "Homeostasis", in preparation for a memory test that would take place 1 week later during Session 2. This text was selected from Cook and Mayer (1988). The text was three paragraphs, 27 sentences, and 262 words in length with a Flesh-Kincaid ease of reading score of 49.1 and a reading grade level of 10.4. The text was presented on the computer screen during the study (5 min) and concept mapping (10 min) periods (see Table 1).

Students were asked to make judgments of learning in an effort to assess the metacognitive impact of the learning conditions. Specifically, they were asked to indicate how much of the material from the text they would remember in 1 week, and they made their ratings on a scale from 0% to 100% in increments of 10 (0, 10, 20, . . . , 80, 90, 100) by clicking a radio button that was displayed on the screen.

During the concept mapping period, students were told that a concept map is a diagram in which concepts are represented as nodes, and relationships among the concepts are represented as lines linking the nodes together. Following previous research (i.e., Blunt & Karpicke, 2014; Karpicke & Blunt, 2011), students were also provided an example of a concept map on paper from Karpicke and Blunt (2011) that they could view while they created their map of the Homeostasis text (see online supplemental mate-

Table 1
Procedures Used During the Learning Phase in Experiments 1 and 2

Conditions	Period 1	Period 2	Period 3	Period 4	Period 5	Total time
Study only	Study (5 min)					5 min
Concept mapping	Study (5 min)	Concept map (20 min)				25 min
Retrieval practice	Study (5 min)	Recall (10 min)	Restudy (5 min)	Recall (10 min)		30 min
CM + RP	Study (5 min)	Concept map (10 or 20 min)	Recall (10 min)	Restudy (5 min)	Recall (10 min)	40 or 50 min

Note. CM + RP = concept mapping plus retrieval practice. The concept mapping condition occurred only in Experiment 2. In the CM + RP condition, the time allotted for concept mapping was 10 min in Experiment 1 and 20 min in Experiment 2. Thus, total time in the CM + RP condition was 40 min in Experiment 1 and 50 min in Experiment 2.

rials for the example concept map students were provided). Students constructed their concept maps on a blank sheet of paper using a pen while the text was presented on the computer screen. Students were given 10 min to construct their concept maps and were encouraged to continue improving their concept maps for the entire 10-min period.

During the two free recall periods during Session 1 for the retrieval practice and CM + RP conditions and the final free recall during Session 2 for all conditions, students were asked to recall as much of the material as they could remember from the text. They were shown the title of the text (Homeostasis) and a response box below it, and they made their responses by typing on the computer. Students were given 10 min to recall as much as they could remember from the text and were encouraged to continue recalling for the entire 10-min period.

For scoring purposes, the text was divided into 30 idea units, and concept maps and free recall protocols were scored using the same criteria (see Blunt & Karpicke, 2014; Karpicke & Blunt, 2011). Specifically, students were given 1 point for correctly recalling an entire idea unit or 0.5 points for correctly recalling part of an idea unit. This made 30 the best possible score for each recall as well as the concept map. Two independent raters scored all of the concept maps and free recall protocols. The Pearson correlations (r) between the two sets of scores were .92 and .95 for concept maps and free recall protocols, respectively. Given the high levels of interrater reliability, the scores given by the two raters were averaged for the reported analyses.

The short answer test included six verbatim and six inference questions. All of the short answer questions along with their answers are provided in the online supplemental materials. Each question was shown one at a time on the computer screen, and students typed their answers on the computer. Verbatim questions assessed memory for information that was explicitly stated in the text and typically required a single idea unit (M = 1.3 idea units per verbatim question). For example, the verbatim question "What does the body do when it is dehydrated?" is answered with the idea unit "When the body becomes dehydrated, more ADH is released." Inference questions required students to remember and connect multiple idea units across the text (M = 2.3 idea units per inference question). For example, the inference question "If the glands of the endocrine system stopped working, homeostasis would no longer be possible. Why?" required students to remember (a) that the endocrine system is responsible for homeostasis, (b) that the endocrine system is a series of glands, and (c) that these glands produce hormones. Students then had to use this information to infer that without the endocrine system, the hormones necessary to maintain homeostasis would no longer be produced.

Verbatim and inference questions were scored by awarding 1 point for completely correct answers and 0.5 points for partially correct answers. This made 6 the best possible score for both the verbatim questions and inference questions. The Pearson correlation (r) between the two sets of scores was .98 for the short answer tests. Given the high level of interrater reliability, the scores given by the two raters were averaged for the reported analyses.

Procedure. Students completed two sessions spaced 1 week apart and were tested in small groups of up to six people. The top portion of Table 1 provides an overview of the conditions in Experiment 1. During the learning phase (Session 1), students in all conditions studied the text, which was shown on a computer screen, in an initial 5-min study period. After the study period, students made an initial judgment of learning. In the study only condition, students were dismissed after making this judgment of learning. In the retrieval practice condition, students practiced recalling the text for 10 min, restudied the text in another 5-min study period and recalled it again in a second 10-min recall period. In the CM + RP condition, after the initial 5-min study period, students constructed their concept maps using paper and a pen while the text was presented on the computer screen. After the 10-min concept mapping period, the concept maps were collected, and students then completed the retrieval practice procedure, which was identical to the one in the retrieval practice condition. Students freely recalled the text for 10 min, restudied for 5 min, and recalled again in a second 10-min recall period. At the end of the learning phase, students in the retrieval practice and CM + RP conditions made a second judgment of learning, using the same procedure as the one used to make the first judgment. Students were then dismissed and returned to the laboratory 1 week later for the final short answer and free recall tests.

The short answer test always occurred first, and students were required to spend at least 30 s attempting to answer each question before a button appeared that allowed them to advance to the next question, but they were encouraged to take as much time as they needed to answer each question. After students answered all of the short answer questions, they completed the final free recall test. The short answer test was intentionally chosen to come first because it is far more educationally authentic than free recall. Consequently, the final free recall test is an impure measure of memory performance because students could have obtained cues from the short answer questions to aid their free recall. At the end of Session 2, students were debriefed and thanked for their participation.

Results

The data and analysis scripts can be downloaded from the online supplemental materials.

Initial concept map and recall performance. Students in the CM + RP condition, on average, produced 46% of the ideas from the text on their maps. The left portion of Table 2 shows the proportion of idea units recalled on the initial free recall tests. A 2 (condition: retrieval practice vs. CM + RP) × 2 (recall test: 1 or 2) mixed factorial ANOVA was carried out. There was no main effect of condition, $F(1, 58) = 0.06$, $\eta_p^2 = 0.00$. There was a main effect of recall test $F(1, 58) = 131.44$, $\eta_p^2 = 0.69$, which reflects the finding that students recalled more idea units on their second recall attempt relative to their first. However, these effects were qualified by a significant Condition × Recall Test interaction, $F(1, 58) = 12.72$, $\eta_p^2 = 0.18$. Students who constructed a concept map performed slightly better on the first recall test relative to students who did not construct a map, $t(58) = 1.61$, $d = 0.42$ [−0.10, 0.93]. This pattern, however, was reversed on the second recall test: Students in the CM + RP condition performed slightly worse than those in the retrieval practice condition, $t(58) = -0.90$, $d = -0.23$ [−0.74, 0.28].

Final short answer performance. The left panel of Figure 1 shows performance on the verbatim questions on the final short answer test. Both the retrieval practice and CM + RP conditions outperformed the study only condition; for retrieval practice vs. study, $t(58) = 5.26$, $d = 1.36$ [0.79, 1.92]; for CM + RP vs. study, $t(58) = 3.20$, $d = 0.83$ [0.29, 1.35]. The most striking result of the experiment, however, was that the CM + RP condition led to worse performance relative to practicing retrieval without concept mapping, $t(58) = 2.12$, $d = 0.55$ [0.03, 1.06]. The pattern of results was similar for the final inference questions, shown in the middle panel of Figure 1. On the inference questions, the retrieval practice condition outperformed the study control condition, $t(58) = 6.13$, $d = 1.58$ [0.99, 2.16], but the CM + RP did not, $t(58) = 1.79$, $d = 0.46$ [−0.05, 0.97]. Once again, the CM + RP condition performed significantly worse than did the retrieval practice condition, $t(58) = 3.92$, $d = 1.01$ [0.47, 1.55]. Although students spent an additional 10 min actively studying the material in the CM + RP condition, that condition harmed learning relative to practicing retrieval without creating a concept map.

Final free recall performance. The right panel of Figure 1 shows the proportion of idea units recalled on the final recall test. The pattern of results on the final free recall test was similar to the pattern of results on the short answer test. Both the retrieval practice and CM + RP conditions outperformed the study only condition; for retrieval practice vs. study: $t(58) = 5.88$, $d = 1.52$ [0.94, 2.09]; for CM + RP vs. study: $t(58) = 4.35$, $d = 1.12$ [0.57, 1.66]. The retrieval practice condition, without concept mapping, produced better final free recall relative to the CM + RP condition, $t(58) = 2.07$, $d = 0.53$ [0.02, 1.05].

Table 2
Proportion of Idea Units Recalled During the Learning Phase in Experiments 1 and 2

Variables	Experiment 1		Experiment 2	
	Recall 1	Recall 2	Recall 1	Recall 2
Retrieval practice	.36 (.03)	.60 (.04)	.56 (.04)	.78 (.03)
CM + RP	.43 (.03)	.55 (.03)	.61 (.03)	.71 (.03)

Note. CM + RP = concept mapping plus retrieval practice. Standard errors of the means are in parentheses. Recall 1 and Recall 2 refer to the first and second recall opportunities respectively, in the initial learning phase.

Judgments of learning. Table 3 shows students' judgments of learning (JOLs) solicited after the initial 5-min study period in all conditions (JOL 1) and again at the end of the learning phase in the retrieval practice and CM + RP conditions (JOL 2). After 5 min of studying, students' judgments of learning were similar in all three conditions, $F(2, 87) = 0.79$, $\eta_p^2 = 0.02$. In the retrieval practice condition, judgments of learning at the end of the learning phase had decreased significantly relative to the judgments after the first study period, $t(29) = -2.78$, $d = -0.51$ [−0.88, −0.12]. In contrast, students in the CM + RP condition did not show this decline in their judgments of learning, showing instead a small but negligible increase, $t(29) = 0.57$, $d = 0.10$ [−0.26, 0.46].

Input-output correspondence. One possible reason that concept mapping might be ineffective when combined with retrieval practice is that creating a concept map might emphasize a relational structure that differs from the serial order of ideas in the text, and this may harm initial free recall by interfering with students' default retrieval strategies. To explore this possibility, input-output correspondence was computed following the technique developed by Asch and Ebenholtz (1962). Specifically, the idea units that were recalled were compared to the order of the idea units in the text. For example, a student may have recalled idea units 1, 2, 5, and 4 in that order. Pairs of idea units can then be grouped (1–2, 2–5, 5–4), and the correspondence of their recall output relative to the original ordering of the idea units within the text can be evaluated by computing the proportion of forward transitions (e.g., 1–2) relative to the total number of transitions. In this example, three of the four pairs preserve the original order resulting in an input-output correspondence of .75. Two students in the retrieval practice condition failed to recall any of the idea units from the text for Recall 1, and one of these students failed to recall any idea units successfully during the final recall test (i.e., Recall 3). Consequently, an input-output correspondence index could not be computed for these subjects, and they were removed from the analysis. The results of this exploratory analysis of input-output correspondence are presented in Table 4. Overall, students tended to recall idea units in the order in which they were presented in the text ($M = 0.91$, $SD = 0.08$). Concept mapping before practicing retrieval made little difference in the order in which idea units were output during Recall 1, $t(56) = 1.44$, $d = 0.38$ [−0.14, 0.90], Recall 2, $t(56) = 0.67$, $d = 0.18$ [−0.34, 0.69], or the final recall test, $t(56) = 0.21$, $d = -0.05$ [−0.57, 0.46], relative to the retrieval practice condition.

Conditional recall analysis. Table 5 shows the results of an exploratory analysis of the idea units that were successfully recalled or not recalled on each of the free recalls during the learning phase of the experiment. We applied Tulving's (1964) method to analyze the fate of individual idea units across two free recall tests (see also Blunt & Karpicke, 2014; Karpicke & Zaromb, 2010). C_1 refers to idea units that were produced during Recall 1, and N_1 refers to idea units that were not produced during Recall 1.

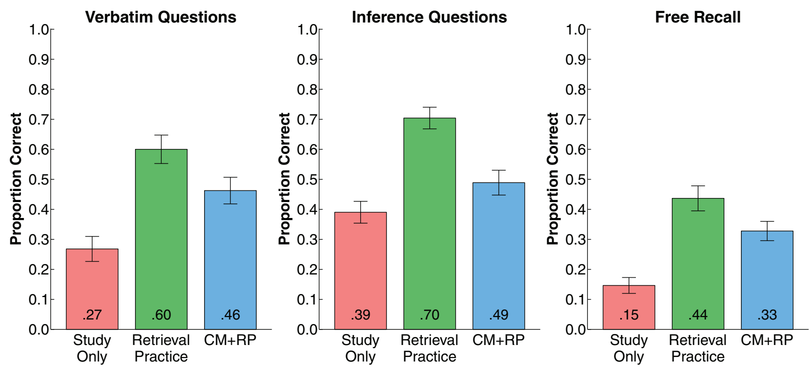

Figure 1. Final short answer performance on verbatim questions and inference questions (left and middle panels) and final recall performance (right panel) in Experiment 1. Error bars represent standard errors of the means. CM + RP = concept mapping plus retrieval practice. See the online article for the color version of this figure.

C_2 refers to idea units that were produced during Recall 2, and N_2 refers to idea units that were not produced during Recall 2. This analysis revealed one major difference between the two retrieval practice conditions: the probability of not recalling an idea on Recall 1 but recalling it on Recall 2 (N_1C_2) was higher in the retrieval practice condition (M = .28, SD = .11) than it was in the CM + RP condition (M = .19, SD = .11), $t(58) = 3.13$, $d = 0.81$ [0.28, 1.33]. The conditions did not differ in intertest retention (C_1C_2; d = 0.34 [−0.17, 0.85]), intertest forgetting (C_1N_2; d = 0.29 [−0.22, 0.80]), or failing to recall the item on both tests (N_1N_2; d = 0.04 [−0.46, 0.55]). This pattern of results suggests that concept mapping is selectively impairing the ability to gain items from Recall 1 to Recall 2, perhaps by changing the way in which students process the intervening restudy opportunity.

Discussion

In Experiment 1, creating a concept map produced a small gain on an immediate recall test during the initial learning phase, but this benefit vanished on a subsequent recall test and completely reversed on the final short answer and free recall tests 1 week later. Students spent an additional 10 min studying the material in the CM + RP condition, but this activity ultimately failed to produce any benefit beyond practicing retrieval. Concept mapping improved performance on an immediate free recall test, but this advantage disappeared when students reread the text and recalled a second time. Exploratory analyses revealed that concept mapping before practicing retrieval did not disrupt the order in which idea units were recalled during practice but did disrupt students' ability to gain new idea units from the first and second recall periods relative to when they only read the text before practicing retrieval. Experiment 2 further investigated the efficacy of combining concept mapping and retrieval practice.

Table 3
Judgments of Learning in Experiments 1 and 2

Variables	Experiment 1 JOL 1	Experiment 1 JOL 2	Experiment 2 JOL
Study only	.58 (.04)	—	.67 (.04)
Concept mapping	—	—	.71 (.04)
Retrieval practice	.64 (.03)	.54 (.03)	.61 (.04)
CM + RP	.61 (.03)	.62 (.04)	.64 (.03)

Note. CM + RP = concept mapping plus retrieval practice; JOL = judgment of learning. Standard errors of the means are in parentheses. In Experiment 2, judgments of learning occurred only once at the end of the learning phase.

Experiment 2

Experiment 2 had two primary purposes. The first purpose was to replicate and extend the results of Experiment 1 to a new set of materials. Experiment 2 also included a condition in which students created a concept map without practicing retrieval, resulting in four conditions: study only, concept mapping, retrieval practice, and CM + RP. This design allowed us to examine the effects of concept mapping and retrieval practice both in isolation and in combination. The second purpose was to increase performance on the initial concept map activity relative to where it was in Experiment 1. In Experiment 2, students were given 20 min to create their concept maps, twice the amount of time that students were given in Experiment 1. While constructing their concept maps, students received encouragement at 5-min intervals to continue drawing relations among all of the key ideas presented in the text. Experiment 2 included a few additional changes relative to Experiment 1. Specifically, students made a single judgment of learning at the end of the learning phase, rather than two judgments of learning. Students also created their concept maps on the computer, rather than on paper, using a computer program created in our laboratory.

Method

Subjects and design. One hundred twenty undergraduate students (51 females, M_{age} = 20.0 years, SD = 2.1, range = 18–33) participated for partial fulfillment of course requirements. They were recruited from the introductory psychology pool at Purdue University. A between-subjects design was used with students randomly assigned to a study only condition (n = 30), a concept mapping condition (n = 30), a retrieval practice condition (n =

Table 4
Input-Output Correspondence for Recall 1, Recall 2, and Recall 3 in Experiments 1 and 2

	Experiment 1			Experiment 2		
Variables	Recall 1	Recall 2	Recall 3	Recall 1	Recall 2	Recall 3
Retrieval practice	.92 (.01)	.92 (.01)	.90 (.02)	.90 (.02)	.93 (.02)	.92 (.02)
CM + RP	.90 (.01)	.91 (.01)	.90 (.03)	.87 (.02)	.91 (.01)	.91 (.01)

Note. CM + RP = concept mapping plus retrieval practice. Standard errors of the means are in parentheses. Recall 1 and Recall 2 refer to the first and second recall opportunities respectively, in the initial learning phase. Recall 3 refers to the final free recall test in Session 2.

30), or a CM + RP condition (n = 30). The conditions did not significantly differ with respect to age, $F(3, 116) = 0.27$, $\eta_p^2 = 0.01$, or in the proportions of men and women, $\chi^2(3) = 2.83$, $p = .42$. All subjects were fluent English speakers and none had participated in Experiment 1. This research was approved by the Purdue University Institutional Review Board.

Materials and measures. All of the materials and measures are provided in the online supplemental materials. In Experiment 2, half of the students in each condition studied a text titled "Make-up of Human Blood" (four paragraphs, 18 sentences, 236 words, Flesh-Kincaid ease of reading score of 63.7, and a grade level of 7.6), and the other half studied a text titled "Kinds of Muscle Tissue" (three paragraphs, 18 sentences, 248 words, Flesh-Kincaid ease of reading score of 56.7, and a grade level of 8.7). These texts were selected from Cook and Mayer (1988). The randomly assigned text was presented on the computer screen during the study (5 min) and concept mapping (20 min) periods (see Table 1).

Students were asked to make judgments of learning in an effort to assess the metacognitive impact of the learning conditions. Specifically, they were asked to indicate how much of the material from the text they would remember in 1 week, and they made their ratings on a scale from 0% to 100% in increments of 10 (0, 10, 20, . . . , 80, 90, 100) by clicking a radio button that was displayed on the screen.

During the concept mapping period, students were told that a concept map is a diagram in which concepts are represented as nodes and relationships among the concepts are represented as lines linking the nodes together. Following previous research (i.e., Blunt & Karpicke, 2014; Karpicke & Blunt, 2011), students were also provided an example of a concept map on paper that was adapted from Blunt and Karpicke (2014) that they could view while they created their concept map (see online supplemental materials for the example concept map students were provided). During the concept map period for the concept mapping and CM + RP conditions, the text was shown on the left side of the computer screen, and students created their concept maps on a canvas displayed on the right side of the screen. Students were given 20 min to construct their concept maps and were encouraged to continue improving their concept maps for the entire 20-min period. At 5-min intervals, the experimenter reminded the students that they should keep working for the entire 20-min period and that it was important to build connections among all of the key ideas contained in the text when possible.

During the two free recall periods during Session 1 for the retrieval practice and CM + RP conditions and the final free recall during Session 2 for all conditions, students were asked to recall as much of the material as they could remember from the text. They were shown the title of the text ("Make-up of Human Blood" or "Kinds of Muscle Tissue") and a response box below it, and they made their responses by typing on the computer. Students were given 10 min to recall as much as they could remember from the text and were encouraged to continue recalling for the entire 10-min period.

For scoring purposes, the two texts were divided into 27 idea units each, and concept maps and free recall protocols were scored using the same criteria as Experiment 1 (see Blunt & Karpicke, 2014; Karpicke & Blunt, 2011). Specifically, students were given 1 point for correctly recalling an entire idea unit or 0.5 points for correctly recalling part of an idea unit. This made 27 the best possible score for each recall as well as the concept map. Two independent raters scored all of the concept maps and free recall protocols. The Pearson correlations (r) between the two sets of scores were .91 and .94 for concept maps and free recall protocols, respectively. Given the high levels of interrater reliability, the scores given by the two raters were averaged for the reported analyses.

The short answer test included 10 verbatim and four inference questions for each text. Students were tested only on the questions that corresponded to the text that they had studied during Session

Table 5
Conditional Recall Analysis for Recall 1 and Recall 2 in Experiments 1 and 2

	Experiment 1				Experiment 2			
Variables	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	.36 (.04)	.05 (.01)	.28 (.02)	.31 (.04)	.56 (.04)	.03 (.01)	.24 (.02)	.17 (.03)
CM + RP	.42 (.03)	.07 (.01)	.19 (.02)	.32 (.03)	.60 (.03)	.05 (.01)	.14 (.02)	.21 (.03)

Note. CM + RP = concept mapping plus retrieval practice; C_1 = items recalled on the first recall test; N_1 = items that were not recalled on the first recall test; C_2 = items recalled on the second recall test; N_2 = items that were not recalled on the second recall test. Standard errors of the means are in parentheses.

1. All of the short answer questions along with their answers are provided in the online supplemental materials. Verbatim questions assessed memory for information that was explicitly stated in the text and typically required a single idea unit (collapsed across texts, $M = 1.2$ idea units per verbatim question). For example, the verbatim question "How many kinds of muscle tissue do vertebrates have in their bodies?" is answered with the idea unit "Vertebrates have three kinds of muscle tissue within their body systems." Inference questions required students to remember and connect multiple idea units across the text (collapsed across texts, $M = 2.6$ idea units per inference question). For example, the inference question "What are the two kinds of muscle tissue that are under involuntary control?" required students to remember (a) that there are three types of muscle tissue (smooth, cardiac, and striated) and (b) that striated muscle tissue is controlled by the thinking part of the brain. Students then had to use this information to infer that that the other two types of muscle tissue are under involuntary control, which was not explicitly stated in the text.

Verbatim and inference questions were scored by awarding 1 point for completely correct answers and 0.5 points for partially correct answers. This made 10 the best possible score for the verbatim questions and 4 the best possible score for inference questions. The Pearson correlation (r) between the two sets of scores was .97. Given the high level of interrater reliability, the scores given by the two raters were averaged for the reported analyses.

Procedure. The bottom portion of Table 1 provides an overview of the experimental conditions. The procedure was identical to the one used in Experiment 1 with three exceptions. First, the concept mapping tasks were completed on the computer using custom software created in our laboratory. Students were given instruction about how to create concept maps with the software and practiced using the software for 2 min. The experimenter answered any questions about the software during the practice period. Second, students were given 20 min to create their concept maps in the concept mapping and CM + RP conditions. At 5-min intervals, the experimenter reminded the students that they should keep working for the entire 20-min period and that it was important to build connections among all of the key ideas contained in the text when possible. Third, students in all conditions made a single judgment of learning at the end of the learning phase, following the procedure used in Experiment 1. The final short answer test included 10 verbatim questions and four inference questions. As in Experiment 1, students took the final tests 1 week after the initial learning session, and they completed the short answer test prior to the final free recall test. In all other respects, the procedure in Experiment 2 followed the one used in Experiment 1.

Results

An initial analysis indicted that there were no differences between the two texts and text did not interact with any of the conditions. Consequently, the results have been collapsed across the two texts. The data and analysis scripts can be downloaded from the online supplemental materials.

Initial concept map and recall performance. Students produced, on average, 73% of the ideas from the texts on their maps, and performance was similar in the concept mapping and CM + RP conditions (.72 vs. .75), $t(58) = 0.98$, $d = 0.25$ [−0.26, 0.76]. The right portion of Table 2 shows the proportion of idea units recalled on the initial free recall tests in Experiment 2. A 2 (condition: retrieval practice vs. CM + RP) × 2 (recall test: 1 or 2) mixed factorial ANOVA was carried out. There was no main effect of condition, $F(1, 58) = 0.01$, $\eta_p^2 = 0.00$. There was a main effect of recall test, $F(1, 58) = 88.68$, $\eta_p^2 = 0.61$, which reflects the finding that students recalled more idea units on their second recall attempt relative to their first. However, these effects were qualified by a significant Condition × Recall Test interaction, $F(1, 58) = 13.62$, $\eta_p^2 = 0.19$. As in Experiment 1, students who constructed a concept map performed slightly better on the first recall test relative to students who did not construct a map, $t(58) = 1.17$, $d = 0.30$ [−0.21, 0.81]. This pattern, however, was reversed on the second recall test: Students in the CM + RP condition performed slightly worse than those in the retrieval practice condition, $t(58) = -1.46$, $d = -0.38$ [−0.89, 0.14].

Final short answer performance. The left panel of Figure 2 shows performance on the verbatim questions on the final short answer test. The concept mapping, retrieval practice, and CM + RP conditions all outperformed the study only control condition; for concept mapping vs. study, $t(58) = 2.27$, $d = 0.59$ [0.07, 1.10]; for retrieval practice vs. study, $t(58) = 7.93$, $d = 2.05$ [1.41, 2.67]; for CM + RP vs. study, $t(58) = 6.95$ $d = 1.79$ [1.19, 2.39]. Retrieval practice enhanced performance relative to concept mapping, $t(58) = 5.88$, $d = 1.52$ [0.94, 2.09]. Adding retrieval practice to concept mapping in the CM + RP condition enhanced learning relative to concept mapping, $t(58) = 4.79$, $d = 1.24$ [0.68, 1.79]. However, the CM + RP condition was slightly worse than retrieval practice alone, $t(58) = -1.44$, $d = -0.37$ [−0.88, 0.14].

The pattern was similar for inference questions, shown in the middle panel of Figure 2. There was a 10% difference favoring the concept mapping condition over the study condition, but the difference did not reach significance, $t(58) = 1.49$, $d = 0.38$ [−0.13, 0.89]. Both the retrieval practice and CM + RP conditions outperformed the study condition; for retrieval practice vs. study, $t(58) = 4.83$, $d = 1.25$ [0.69, 1.80]; for CM + RP vs. study, $t(58) = 3.75$, $d = 0.97$ [0.43, 1.50]. Retrieval practice enhanced performance relative to concept mapping, $t(58) = 3.30$, $d = 0.85$ [0.32, 1.38], and the CM + RP condition outperformed the concept mapping condition, $t(58) = 2.23$, $d = 0.57$ [0.06, 1.09]. The CM + RP condition, once again, performed slightly worse than the retrieval practice condition, $t(58) = -1.13$, $d = -0.29$, [−0.80, 0.22].

Final free recall performance. The right panel of Figure 2 shows performance on the final recall test. The concept mapping, retrieval practice, and CM + RP conditions all outperformed the study only control condition; for concept mapping vs. study, $t(58) = 3.02$, $d = 0.78$ [0.25, 1.30]; for retrieval practice vs. study, $t(58) = 9.00$, $d = 2.32$ [1.66. 2.98]; for CM + RP versus study, $t(58) = 7.06$, $d = 1.82$ [1.21, 2.42]. Retrieval practice enhanced performance relative to concept mapping, $t(58) = 5.96$, $d = 1.54$ [0.96, 2.11], and adding retrieval practice to concept mapping in the CM + RP condition enhanced learning relative to concept mapping, $t(58) = 4.19$, $d = 1.08$ [0.54, 1.62]. Following the results for verbatim and inference short answer questions, the CM + RP condition produced slightly worse final free recall relative to retrieval practice alone, $t(58) = -1.46$, $d = -0.38$ [−0.89, 0.14].

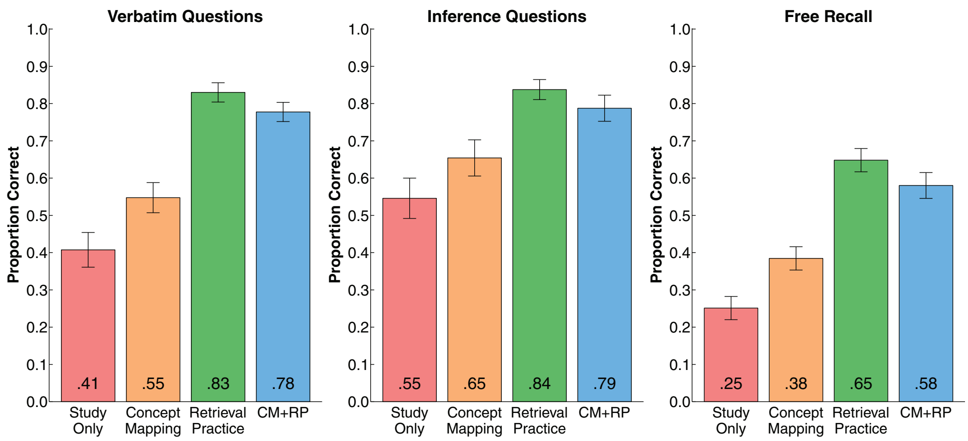

Figure 2. Final short answer performance on verbatim questions and inference questions (left and middle panels) and final recall performance (right panel) in Experiment 2. Error bars represent standard errors of the means. CM + RP = concept mapping plus retrieval practice. See the online article for the color version of this figure.

Judgments of learning. The right portion of Table 3 shows students' judgments of learning in Experiment 2. A one-way ANOVA did not detect any significant differences in judgments of learning among the condition, $F(3, 116) = 1.14$, $\eta_p^2 = 0.03$. Numerically, students predicted better learning in the conditions that did not involve retrieval (study only and concept mapping), which directly opposes the actual pattern of final test performance and replicates the pattern observed in Karpicke and Blunt (2011). Passive studying and concept mapping led learners to believe that they had far better mastery over the material, whereas retrieval-based learning led learners to be better calibrated in their metacognitive predictions.

Input-output correspondence. Table 4 shows the results of the analysis of input-output correspondence. Overall, students tended to recall idea units in the order in which they were presented in the text ($M = 0.91$, $SD = 0.09$). Replicating the results of Experiment 1, concept mapping before practicing retrieval made little difference in the order in which idea units were output during Recall 1, $t(58) = -1.20$, $d = -0.31$ [−0.82, 0.20], Recall 2, $t(58) = -1.06$, $d = -0.27$ [−0.78, 0.23], or the final recall test $t(58) = -0.38$, $d = -0.10$ [−0.60, 0.41], relative to the retrieval practice condition.

Conditional recall analysis. Table 5 shows the results of an exploratory analysis of the idea units that were successfully recalled on each of the free recall tests during the learning phase of the experiment. The pattern replicates Experiment 1 with the main difference between the retrieval practice and CM + RP conditions occurring for intertest gains (N_1C_2). Specifically, students in the retrieval practice condition ($M = .24$, $SD = .12$) gained more items from Recall 1 to Recall 2 than did students in the CM + RP condition ($M = .14$, $SD = .09$), $t(58) = 3.49$, $d = 0.90$ [0.37, 1.43]. Surprisingly, there was a significant difference in ntertest forgetting (C_1N_2) with students in the retrieval practice condition ($M = .03$, $SD = .03$) forgetting items on Recall 2 less frequently than students in the CM + RP condition ($M = .05$, $SD = .06$), $t(58) = 2.26$, $d = 0.58$ [0.07, 1.10]). In both conditions, few items were initially recalled but then forgotten between Recall 1 and Recall 2. The situation in which an item was not recalled during Recall 1 but was recalled on Recall 2 (N_1C_2) was a more frequent occurrence. The 10% difference in intertest gains is what produced the interaction in recall performance during the learning phase. The groups did not differ in the probability of retaining items across recalls (C_1C_2; $d = 0.17$ [−0.34, 0.68]) and the probability of not recalling items on either Recall 1 or Recall 2 (N_1N_2; $d = 0.21$ [−0.30, 0.72]). As in Experiment 1, concept mapping before retrieval practice disrupted intertest gains across the two initial recall attempts.

Discussion

Experiment 2 replicated the results of Experiment 1 with a new set of materials. Students were given more time (20 min) to create their concept maps, received verbal reminders every 5 min to keep working on their maps, and were encouraged to build connections among all the key ideas in the texts. These changes improved students' performance on the initial concept map activity to levels observed in previous research (Blunt & Karpicke, 2014; Karpicke & Blunt, 2011). Concept mapping produced a small benefit on an initial recall test immediately after the concept map activity. That benefit, however, did not persist when students reread the text and recalled a second time, a finding also observed in Experiment 1. Creating a concept map enhanced long-term retention relative to the study only control condition, but both retrieval practice conditions outperformed concept mapping alone. Most importantly, combining concept mapping and retrieval practice again failed to improve learning relative to retrieval practice alone, replicating the key finding from Experiment 1. The CM + RP condition required a large increase in total time, with students spending 20 min creating a concept map before practicing retrieval, yet the concept mapping activity produced no additional learning. Indeed, the trend was negative, with the CM + RP condition performing about 5% worse than the retrieval practice condition on the final test.

General Discussion

The purpose of the present experiments was to investigate whether combining retrieval practice with an elaborative study activity would improve learning. In two experiments, students studied educationally relevant texts about biology topics. Some students read the text in a study only control condition, whereas other students practiced retrieval by freely recalling the text. In elaborative study conditions, students created a concept map after they studied the text. The critical condition involved combining the concept mapping and retrieval practice activities: Students studied the text, created a concept map, and then practiced retrieval by freely recalling the text. Learning was assessed 1 week later with a short answer test that included both verbatim and inference questions and a final free recall test. The pattern of results in both experiments was unequivocal. Practicing retrieval enhanced performance on all final assessments relative to the study only condition in both experiments, and in Experiment 2 the retrieval practice conditions outperformed the concept mapping condition. Most importantly, combining concept mapping with retrieval practice failed to produce any benefit beyond retrieval practice alone.

The failure to find any benefit of combining concept mapping and retrieval practice was surprising. Concept mapping is often touted as an effective learning activity (Fiorella & Mayer, 2015; Nesbit & Adesope, 2006; Novak & Cañas, 2008; Schroeder et al., 2018), and concept mapping would appear to encourage learners to elaborate on the material they are learning (Blunt & Karpicke, 2014; Karpicke & Blunt, 2011). Constructing a concept map is thought to be effective because it allows learners to create a graphical representation of material that emphasizes key ideas, specifies relations among ideas, and contains linking phrases that provide additional information about those relations (Novak, 1990). Another view is that concept mapping might promote learning because it requires students to select and isolate key pieces of information, organize that key information in a graphical form, and integrate those pieces of information together with relationship links. The end product is a student-generated visual representation that has removed unimportant details while emphasizing the relationships among the most important ideas (Fiorella & Mayer, 2015). Despite these theoretical explanations for the potential benefits of concept mapping, in the present experiments, benefits of concept mapping were seen only on an immediate free recall test and on a final test when compared to the study only control condition. Retrieval practice outperformed concept mapping on the final test, and there were no benefits of combining concept mapping with retrieval practice.

It is important that benefits of concept mapping were observed in some facets of the present experiments because this confirms that concept mapping was successfully implemented and eliminates manipulation failure as an explanation for the present results. In both experiments, creating a concept map improved initial free recall. Had the experiments stopped at that point, conclusions about the effectiveness of concept mapping would have been much more positive. Concept mapping also improved performance on the final test in Experiment 2 relative to the study only control condition. Both findings, however, represent rather unimpressive benefits of concept mapping. Immediate tests of free recall are not as prized as assessments of long-term meaningful learning (Anderson et al., 2001), and moreover, those immediate positive effects were reversed when students reread and recalled the text a second time. In Experiment 2, the concept mapping condition spent substantially more time with the material than did the study only control condition (25 min vs. 5 min) so that advantage was likely driven by the large discrepancy in overall study time.

In an effort to explain why concept mapping was ineffective at improving retrieval practice, two exploratory analyses were conducted. First, input-output correspondence was examined because concept mapping might have afforded an alternative organizational strategy that was incompatible with subjects' natural or default organizational strategies. To explore this possibility, input-output correspondence was computed following the technique by Asch and Ebenholtz (1962). However, there was no difference between the two retrieval practice conditions in the order in which idea units were recalled. Students tended to recall the idea units in the order in which they were presented within the text, and concept mapping did not disrupt this organization.

The second exploratory analysis revealed that concept mapping before practicing retrieval impaired item gains from the first recall to the second. Students could use the restudy opportunity as a way to locate the information they had failed to initially retrieve—an item-specific approach to text comprehension. On other hand, students who created concept maps may have been primed to process the text relationally and may have approached the restudy trial from the standpoint of reading the text and looking for the important relationships they constructed earlier. This relational processing would be redundant because students had already learned the text in this manner when they had constructed their concept maps. In other words, concept mapping would lead to ineffective use of the restudy trial, whereas students who did not create a concept map were not primed to process the text relationally and would look for specific items that they did not recall during the first retrieval event. By focusing their mental effort on finding items they had not successfully retrieved or retrieved incorrectly, they can update this information in memory, allowing for better performance on the second retrieval opportunity. Practically, these differences on the first and second free recall tests were small. If concept mapping were a more effective strategy, then a large advantage would be expected on the first recall attempt, making a lack of gains in the restudy trial trivial. Because concept mapping did not produce a large advantage on the immediate recall test, the detrimental effect of concept mapping on item gains was evident.

The present results fit with prior research that also failed to find benefits of combining elaborative study strategies with retrieval practice. Fritz et al. (2007) and Karpicke and Smith (2012) did not observe benefits of combining retrieval practice with the keyword mnemonic, an imagery-based elaborative encoding strategy. Rummer et al. (2017) did not find benefits of combining retrieval practice with note taking. Identifying the most effective ways to enhance student learning by combining learning strategies is an important educational problem that has received relatively little attention. The present research examined the effectiveness of combining retrieval practice with a popular educational activity, used authentic educational materials, and assessed performance on delayed tests that measured knowledge in multiple ways with verbatim questions, inference questions, and final free recall. The research base on the effects of combining multiple learning strategies remains quite small, and there is considerable room for further exploration of ways to combine retrieval practice with other effective educational activities.

The present project has important implications for research on the effectiveness of concept mapping. Although claims about the effectiveness of concept mapping are often based on observations, opinions, or correlational studies (Novak & Cañas, 2008), two meta-analyses of experimental studies (Nesbit & Adesope, 2006; Schroeder et al., 2018) concluded that there were generally positive effects of concept mapping on learning. Those analyses, however, point out three critical issues in the research base on concept mapping: The meta-analyses included studies that lacked random assignment of students to concept mapping or control conditions (e.g., Nicoll, Francisco, & Nakhleh, 2001; Okebukola, 1992), included studies in which total time was confounded, with concept mapping conditions given substantially more study time than control conditions (e.g., Chang, Sung, & Chen, 2002; Willerman & Mac Harg, 1991) and included studies in which concept mapping was combined with a second activity that was not present in the control condition (e.g., Chularut & DeBacker, 2004; Lehman, Carter, & Kahle, 1985; Okebukola & Jegede, 1988; Schmid & Telaro, 1990). At the same time, the meta-analyses excluded some studies that did not suffer from these problems—studies in which students were randomly assigned to conditions, in which total time was controlled across conditions, and in which concept map conditions were not confounded with other activities (e.g., Blunt & Karpicke, 2014; Grimaldi, Poston, & Karpicke, 2015; Karpicke & Blunt, 2011).

The present experiments illustrate why these issues, especially the latter two, present problems for evaluating the true benefits of concept mapping as a learning activity. As previously noted, concept mapping enhanced long-term retention in Experiment 2 relative to the study only control condition, and concept mapping enhanced performance on immediate free recall tests in both experiments. But total time was confounded in those comparisons, with the concept map conditions spending considerably more time with the materials relative to the control condition. In both experiments, the CM + RP condition outperformed the study only control condition. Had the present experiments included only those two conditions, one might want to conclude that concept mapping produced an enormous effect on learning. That conclusion is obviously incorrect. Because both experiments included a condition that only practiced retrieval, it is clear that the entire benefit in the CM + RP condition was driven by retrieval practice. Indeed, the concept mapping activity produced a slightly negative effect on performance. A comparison of the CM + RP condition with the study only control condition may seem obviously problematic, but conceptually it is the same comparison made in other studies in which concept mapping was combined with a second activity (e.g., Chularut & DeBacker, 2004; Lehman et al., 1985; Okebukola & Jegede, 1988; Schmid & Telaro, 1990), studies that were included in meta-analyses used as evidence for the overall effectiveness of concept mapping (Nesbit & Adesope, 2006; Schroeder et al., 2018). There is a continuing need for rigorous examinations of learning strategies like concept mapping so that students, teachers, and educators can draw appropriate, evidence-based conclusions about their effectiveness.

References

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

Asch, S. E., & Ebenholtz, S. M. (1962). The principle of associative symmetry. *Proceedings of the American Philosophical Society, 106,* 135–163.

Blunt, J. R., & Karpicke, J. D. (2014). Learning with retrieval-based concept mapping. *Journal of Educational Psychology, 106,* 849–858. http://dx.doi.org/10.1037/a0035934

Carpenter, S. K. (2017). Spacing effects in learning and memory. In J. T. Wixted & J. H. Byrne (Eds.), *Learning and memory: A comprehensive reference*. Oxford, England: Academic Press. http://dx.doi.org/10.1016/B978-0-12-809324-5.21054-7

Chang, K., Sung, Y., & Chen, I. (2002). The effect of concept mapping to enhance text comprehension and summarization. *Journal of Experimental Education, 71,* 5–23. http://dx.doi.org/10.1080/00220970209602054

Chularut, P., & DeBacker, T. K. (2004). The influence of concept mapping on achievement, self-regulation, and self-efficacy in students of English as a second language. *Contemporary Educational Psychology, 29,* 248–263. http://dx.doi.org/10.1016/j.cedpsych.2003.09.001

Cook, L. K., & Mayer, R. E. (1988). Teaching readers about the structure of scientific text. *Journal of Educational Psychology, 80,* 448–456. http://dx.doi.org/10.1037/0022-0663.80.4.448

Craik, F. I. M., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General, 104,* 268–294. http://dx.doi.org/10.1037/0096-3445.104.3.268

Fiorella, L., & Mayer, R. E. (2015). *Learning as a generative activity: Eight learning strategies that promote understanding*. New York, NY: Cambridge University Press. http://dx.doi.org/10.1017/CBO9781107707085

Fritz, C. O., Morris, P. E., Acton, M., Voelkel, A. R., & Etkind, R. (2007). Comparing and combining retrieval practice and the keyword mnemonic for foreign vocabulary learning. *Applied Cognitive Psychology, 21,* 499–526. http://dx.doi.org/10.1002/acp.1287

Grimaldi, P. J., Poston, L., & Karpicke, J. D. (2015). How does creating a concept map affect item-specific encoding? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 41,* 1049–1061. http://dx.doi.org/10.1037/xlm0000076

Hunt, R. R. (2012). Distinctive processing: The co-action of similarity and difference in memory. In B. H. Ross (Ed.), *The psychology of learning of motivation* (Vol. 56, pp. 1–46). Oxford, England: Elsevier. http://dx.doi.org/10.1016/B978-0-12-394393-4.00001-7

Hunt, R. R., & McDaniel, M. A. (1993). The enigma of organization and distinctiveness. *Journal of Memory and Language, 32,* 421–445. http://dx.doi.org/10.1006/jmla.1993.1023

Kahana, M. J. (2017). Memory search. In J. H. Byrne (Ed.), *Learning and memory: A comprehensive reference* (2nd ed., Vol. 2, pp. 181–200). Oxford, England: Academic Press. http://dx.doi.org/10.1016/B978-0-12-809324-5.21038-9

Kang, S. H. K. (2017). The benefits of interleaved practice for learning. In J. C. Horvath, J. Lodge, & J. A. C. Hattie (Eds.), *From the laboratory to the classroom: Translating science of learning for teachers* (pp. 79–93). New York, NY: Routledge.

Karpicke, J. D. (2017). Retrieval-based learning: A decade of progress. In J. T. Wixted (Ed.), *Cognitive psychology of memory, Vol. 2 of learning and memory: A comprehensive reference* (J. H. Byrne, Series Ed.) (pp. 487–514). Oxford, England: Academic Press. http://dx.doi.org/10.1016/B978-0-12-809324-5.21055-9

Karpicke, J. D. (2018). Concept mapping. In B. Frey (Ed.), *The SAGE encyclopedia of educational research, measurement, and evaluation* (pp. 351–354). Thousand Oaks, CA: SAGE Publishing, Inc.

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

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

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

Lechuga, M. T., Ortega-Tudela, J. M., & Gómez-Ariza, C. J. (2015). Further evidence that concept mapping is not better than repeated retrieval as a tool for learning from texts. *Learning and Instruction, 40,* 61–68. http://dx.doi.org/10.1016/j.learninstruc.2015.08.002

Lehman, J. D., Carter, C., & Kahle, J. B. (1985). Concept mapping, vee mapping, and achievement: Results of a field study with black high school students. *Journal of Research in Science Teaching, 22,* 663–673. http://dx.doi.org/10.1002/tea.3660220706

McDaniel, M. A., & Pressley, M. (1984). Putting the keyword method incontext. *Journal of Educational Psychology, 76,* 598–609. http://dx.doi.org/10.1037/0022-0663.76.4.598

Nairne, J. S. (2006). Modeling distinctiveness: Implications for general memory theory. In R. R. Hunt & J. Worthen (Eds.), *Distinctiveness and memory* (pp. 27–46). New York, NY: Oxford University Press. http://dx.doi.org/10.1093/acprof:oso/9780195169669.003.0002

Nesbit, J. C., & Adesope, O. O. (2006). Learning with concept and knowledge maps: A meta-analysis. *Review of Educational Research, 76,* 413–448. http://dx.doi.org/10.3102/00346543076003413

Nicoll, G., Francisco, J., & Nakhleh, M. (2001). An investigation of the value of using concept maps in general chemistry. *Journal of Chemical Education, 78,* 1111–1117. http://dx.doi.org/10.1021/ed078p1111

Novak, J. D. (1990). Concept mapping: A useful tool for science education. *Journal of Research in Science Teaching, 27,* 937–949. http://dx.doi.org/10.1002/tea.3660271003

Novak, J. D., & Cañas, A. J. (2008). *The theory underlying concept maps and how to construct and use them*. Technical report IHMC CmapTools 2006 -01 Rev 01 -2008. Pensacola, FL: Florida Institute for Human and Machine Cognition.

Okebukola, P. A. (1992). Concept mapping with a cooperative learning flavor. *American Biology Teacher, 54,* 218–221. http://dx.doi.org/10.2307/4449458

Okebukola, P. A., & Jegede, O. J. (1988). Cognitive preference and learning-mode as determinants of meaningful learning through concept mapping. *Science Education, 72,* 489–500. http://dx.doi.org/10.1002/sce.3730720408

Raaijmakers, J. G. W., & Shiffrin, R. M. (1981). Search of associative memory. *Psychological Review, 88,* 93–134. http://dx.doi.org/10.1037/0033-295X.88.2.93

Rummer, R., Schweppe, J., Gerst, K., & Wagner, S. (2017). Is testing a more effective learning strategy than note-taking? *Journal of Experimental Psychology: Applied, 23,* 293–300. http://dx.doi.org/10.1037/xap0000134

Schmid, R. F., & Telaro, G. (1990). Concept mapping as an instructional strategy for high school biology. *Journal of Educational Research, 84,* 78–85. http://dx.doi.org/10.1080/00220671.1990.10885996

Schroeder, N. L., Nesbit, J. C., Anguiano, C. J., & Adesope, O. O. (2018). Studying and constructing concept maps: A meta-analysis. *Educational Psychology Review, 30,* 431–455. http://dx.doi.org/10.1007/s10648-017-9403-9

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

Willerman, M., & Mac Harg, R. A. (1991). The concept map as an advance organizer. *Journal of Research in Science Teaching, 28,* 705–712. http://dx.doi.org/10.1002/tea.3660280807

Received September 27, 2019
Revision received March 23, 2020
Accepted March 23, 2020 ■