www.sciencemag.org/cgi/content/full/science.1199327/DC1

Supporting Online Material for

Retrieval Practice Produces More Learning than Elaborative Studying with Concept Mapping

Jeffrey D. Karpicke* and Janell R. Blunt

*To whom correspondence should be addressed. E-mail: karpicke@purdue.edu

Published 20 January 2011 on *Science* Express
DOI: 10.1126/science.1199327

This PDF file includes:

Supporting Online Material

Materials and Methods

Experiment 1

Method

Subjects. Eighty Purdue University undergraduates participated in Experiment 1 in exchange for course credit.

Materials. A 276-word science text on "Sea Otters" was selected from (*S1*).

Design. A between-subjects design was used. There were four learning conditions, described below, and 20 subjects were assigned to each condition.

Procedure. Subjects were tested in small groups. At the beginning of the experiment the subjects were told they would read and study a brief science text. Students in all conditions studied the text in an initial 5-min study period. In the study condition, subjects only experienced the text in the 5-min study period. In the repeated study condition, subjects studied the text in three additional 5-min study periods (thus they studied the text in a total of four periods). There was a brief 1-min break between study periods. This condition was modeled after repeated study conditions used in prior research (*S1*).

In the elaborative concept mapping condition, after reading the text in the initial 5-min study period, the subjects were instructed about the nature of the concept mapping activity. They were told that a concept map is a diagram where the concepts in a set of material are represented as nodes and relationships among the concepts are represented as lines linking the nodes together. The subjects were shown an example of a concept map selected from (*S2*). They were given a sheet of paper and had 25 min to create a concept

map of the concepts in the text. They were allowed to refer to the example concept map and to the text throughout the entire 25-min period. The subjects were also told that if they finished before the end of the 25-min period, they should spend the remaining amount of time reviewing their map and making sure they had included all the details from the text in their map. The experimenter monitored compliance with these instructions.

In the retrieval practice condition, after studying the text for 5 min, a response box was shown on the computer screen and subjects were told to recall as much of the information from the text as they could, in any order they chose. Subjects made their responses by typing them into the computer. The recall test lasted 10 min. Pilot testing showed that this was enough time for subjects to express their knowledge and reach asymptotic levels of free recall. Subjects then reread the text in another 5-min study period and recalled it again in another 10-min recall period. Brief instructions, lasting a few seconds, were given prior to the restudy period and the second recall period. Overall, the total amount of learning time was identical in the elaborative concept mapping and retrieval practice conditions.

At the end of the learning phase, subjects were asked to predict how much of the material they would remember in one week (an aggregate judgment of learning). They made their predictions using an 11-point scale (0%, 10%, 20% ... 80%, 90%, 100%). In addition, at this point in the experiment the subjects in the elaborative concept mapping condition were asked whether they had any prior knowledge about or experience with making concept maps.

The subjects were dismissed and returned to the laboratory one week later for the final test. The paper-and-pencil test included 14 verbatim short-answer questions and 2 inference short-answer questions. Examples of questions are shown in Table S1. The time to complete the short answer test was not limited and was not precisely recorded, but all subjects were able to complete the final test within a 15-min final testing session. At the end of the session the subjects were debriefed and thanked for their participation.

Results

All results unless otherwise stated were significant at the .05 level.

Scoring. Both the free recall tests and concept maps were scored using the same criteria: Subjects were given 1 point for each idea unit recalled on the tests or produced on the concept maps (*S1, S3*). (Examples of concept maps created by students in Experiment 2 are shown in Fig. S1.) On the final short-answer test, there were a total of 21 possible points for the 14 verbatim questions and 4 possible points for the 2 inference questions, so the final test comprised a total of 25 points. Initially, all 20 concept maps, 20 initial recall tests, and 10 final test protocols were scored by two raters. The percentages of agreement between the two raters were 95%, 97%, and 97%, respectively. Given the high levels of interrater agreement, the remaining initial recall and final test protocols were scored by one rater.

Learning Phase Performance. Subjects in the retrieval practice condition recalled .64 and .81 of the idea units on the first and second recall tests, respectively. Subjects in the elaborative concept mapping condition produced .78 of the idea units on their concept maps. The proportions correct in the mapping condition and on the second test in the retrieval practice condition were not statistically different ($F(1, 38) = 0.46$,

n.s.). Thus levels of performance in the initial learning phase did not differ across the elaborative concept mapping and retrieval practice conditions.

Final Test Performance. The final test data were entered into a 4 (Learning Condition) X 2 (Question Type) ANOVA, with learning condition as a between-subjects factor and question type as a within-subjects factor. There was a main effect of learning condition ($F(3, 76) = 18.25$, $\eta_p^2 = .42$) and a main effect of question type ($F(1, 76) = 7.51$, $\eta_p^2 = .09$), but no interaction ($F(3, 76) = 1.75$, n.s.). Because the pattern of results did not differ across question type, the results were combined and overall test scores were calculated for each subject. Pairwise comparisons among the overall test scores indicated that the repeated study, elaborative concept mapping, and retrieval practice conditions all performed better than the study-once condition ($M = .27$, all Fs$(1, 38) > 14.7$, all $\eta_p^2 > .28$). Performance was slightly better in the repeated study condition ($M = .49$) than in the elaborative concept mapping condition ($M = .45$), though this difference was not significant ($F(1, 38) = 0.31$, n.s.). Finally, performance in the retrieval practice condition ($M = .67$) was significantly better than performance in the repeated study condition ($F(1, 38) = 12.44$, $\eta_p^2 = .25$). Most importantly, retrieval practice produced significantly better performance than elaborative studying with concept mapping ($F(1, 38) = 21.63$, $\eta_p^2 = .36$). The effect size for the difference between the retrieval practice and mapping conditions was $d = 1.50$ (*S4*).

In the elaborative concept mapping condition, 14 subjects indicated that they had prior experience creating concept maps and 6 subjects indicated that they had no prior experience. This prior-experience factor did not interact with anything. The proportion correct on the initial map and the proportion correct on the final verbatim and inference

questions did not differ as a function of whether students had prior experience with concept mapping (all Fs < 1).

Judgments of Learning. The highest judgments of learning (JOLs) were given in the repeated study condition (M = .79), and these JOLs were significantly higher than those in the other three conditions (all Fs(1, 38) > 4.1, all η_p^2 > .10). JOLs in the study and mapping conditions were not reliably different (Ms = .680 and .675, F(1, 38) = 0.01, n.s.). Finally, JOLs were the lowest in the retrieval practice condition (M = .58). The difference between JOLs in the retrieval practice condition and those in (a) the study condition and (b) the elaborative concept mapping condition approached but did not reach significance (for retrieval practice vs. study: F(1, 38) = 2.67, η_p^2 = .07, p = .11; for retrieval practice vs. mapping: F(1, 38) = 2.08, η_p^2 = .05, p = .16).

Experiment 2

Method

Subjects. One hundred and twenty Purdue University undergraduates participated in Experiment 2 in exchange for course credit. None of the subjects had participated in Experiment 1.

Materials. Four science texts were selected from (*S5*). Two of the texts, "The Human Ear" and "The Digestive System", had sequential structures, which means the texts described a connected series of events and steps in a process (e.g., the sequence of events involved in the process of digestion). The sequence texts were 260 and 268 words in length, respectively. The other two texts, "Make-up of Human Blood" and "Kinds of Muscle Tissue", had enumeration structures, which means that the texts listed and

described a series of concepts (e.g., the properties of different muscle tissues). The enumeration texts were 235 and 248 words in length, respectively.

Design. A 2 (Learning Condition: Retrieval Practice vs. Elaborative Concept Mapping) X 2 (Text Structure: Enumeration vs. Sequence) X 2 (Final Test Format: Short Answer vs. Concept Mapping) mixed factorial design was used. Learning condition was manipulated within-subjects. Each subject studied two texts. The subjects created a concept map of the concepts in one of the texts and practiced retrieval of the concepts in the other text. The order in which subjects performed the two learning activities was counterbalanced across subjects.

Text structure was manipulated between-subjects. Half of the subjects studied the two enumeration texts and half studied the two sequence texts.

Final test format was also manipulated between-subjects. Half of the subjects took a final test that involved answering short-answer questions (a combination of verbatim and inference questions). The other half of the subjects created concept maps on the final test. Thus subjects took either a final short-answer test or a final concept-mapping test over both of the texts they experienced in the initial learning phase (the text for which they created a concept a map and the text for which they practiced retrieval).

Procedure. The procedure was identical to the procedure used in Experiment 1 with a few exceptions. In the elaborative concept mapping task, subjects first studied one of the science texts, which was shown on a computer screen for 5 min. After studying the text the subjects were given the same instructions about concept mapping used in Experiment 1. Subjects were given a sheet of paper and had 20 min to create their concept map (a shorter amount of time was used in Experiment 2 because all subjects in

Experiment 1 had completed their concept maps well within a 25-min time period). After subjects created the concept map, they made a judgment of learning, again using the same procedure used in Experiment 1.

In the retrieval practice task, subjects first studied one of the texts on the computer screen for 5 min. They were then shown a response box on the computer screen and were told to recall as much of the information from the text as they could. Subjects made their responses by typing them into the computer. The recall period lasted 7 min. Preliminary pilot testing showed that a 7 min recall period provided enough time for subjects to express their knowledge and reach asymptotic levels of recall of these texts. Subjects then reread the text in another 5-min reading period and recalled it again in another 7-min recall period. The subjects made a judgment of learning at the end of the retrieval practice task. Again, note that the order of the two learning tasks (elaborative concept mapping and retrieval practice) was counterbalanced across subjects.

The subjects returned to the laboratory one week later and took either a final short-answer test or a final concept-mapping test. The final short-answer test included 10 verbatim questions and 4 inference questions for each text (thus each subject answered a total of 28 questions). Examples of questions are shown in Table S1. The final short answer test was administered on the computer. Subjects saw title of the text, a question, and a response box on the screen and were told to type their answer to the question in the response box. After 20 s had elapsed, a button labeled "Next" appeared at the bottom of the screen, and subjects clicked the button to advance to the next question. This insured that subjects spent at least 20 s attempting to answer each question. In all other respects the time to answer each question was unlimited.

Subjects in the final concept mapping test condition were told they would be creating concept maps of both of the texts they studied in the initial learning phase, but that they would not be allowed to review the texts while they created their maps. They were given a brief reminder about the nature of the concept mapping activity (all subjects had created a concept map of one of the texts in the initial learning phase). Subjects were given the title of one of the texts, and they were given a sheet of paper and had 25 min to create a concept map of the concepts in the text. After creating a concept map of the first text, subjects were given the title of the second text and created a concept map of the concepts in that text. At the end of the experiment, all subjects were debriefed and thanked for their participation.

Results

Scoring. The scoring procedure was same as the one used in Experiment 1. Initially, 20 concept maps, 20 initial recall tests, and 20 final short answer tests, were scored by two raters. The percentage of agreement between the two raters was 95%, 96%, and 95%, respectively. Given the high levels of interrater agreement, the remaining protocols were scored by one rater.

Learning Phase Performance. The proportion of ideas produced on the initial concept maps was .74, while the proportions of ideas recalled on the initial tests in the retrieval practice condition were .46 and .65 on the first and second test, respectively. There was a significant difference between the proportion of ideas produced on the concept maps and the proportion recalled on the second test in the retrieval practice condition ($F(1, 117) = 23.13$, $\eta_p^2 = .17$).

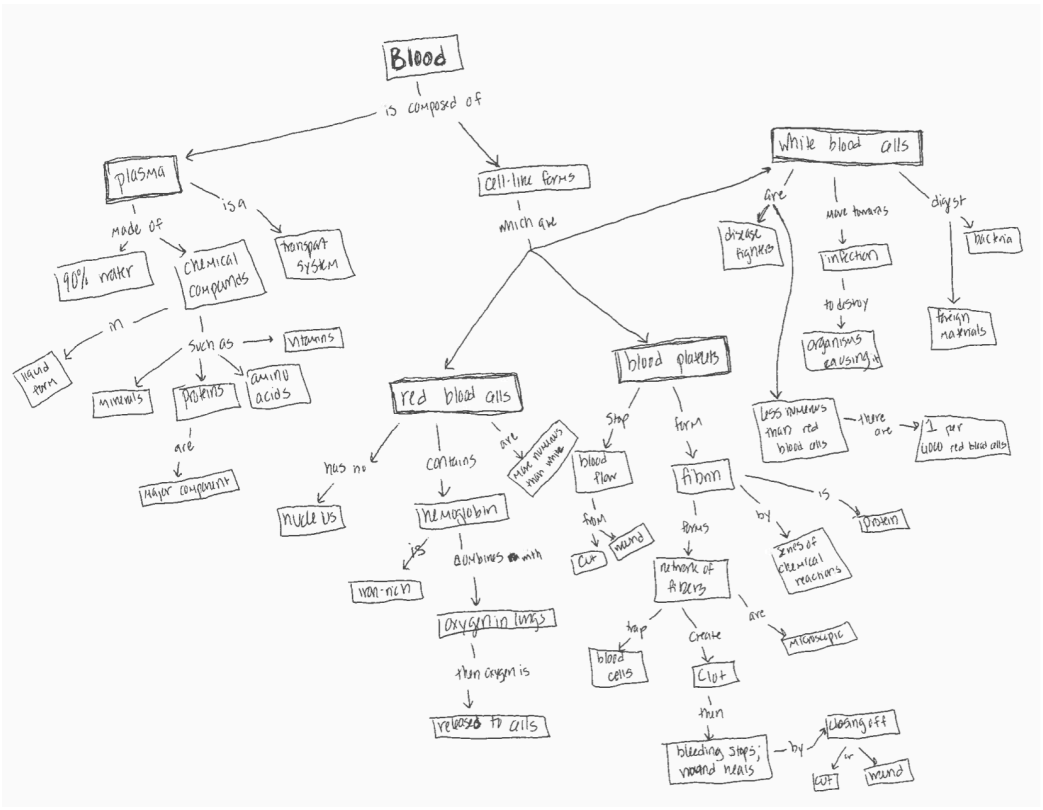
Final Test Performance. An initial analysis indicated that the pattern of results on the final short-answer test was the same for verbatim and inference questions. Therefore, as was done in Experiment 1, the short-answer results were collapsed across question type.

On the final short answer test, retrieval practice produced better performance than elaborative studying with concept mapping for both the enumeration texts (*M*s = .77 vs. .59, $F(1, 29) = 26.72$, $\eta_p^2 = .48$) and the sequence texts (*M*s = .69 vs. .48, $F(1, 29) = 42.99$, $\eta_p^2 = .60$). Collapsed across the two text formats, on the final short-answer test the effect size for the difference between the retrieval practice and elaborative concept mapping conditions was $d = 1.07$ (*M*s = .73 and .54, respectively, $F(1, 59) = 68.54$, $\eta_p^2 = .54$).

When the final test involved creating concept maps, retrieval practice still produced better final test performance than initial elaborative studying with concept mapping. This result occurred for both the enumeration texts (*M*s = .48 vs. .28, $F(1, 29) = 51.75$, $\eta_p^2 = .64$) and the sequence texts (*M*s = .40 vs. .28, $F(1, 29) = 16.03$, $\eta_p^2 = .36$). Collapsed across the two text formats, on the final concept-mapping test the effect size for the difference between the retrieval practice and elaborative concept mapping conditions was $d = 1.01$ (*M*s = .44 and .28, respectively, $F(1, 59) = 58.42$, $\eta_p^2 = .50$).

Judgments of Learning. Judgments of learning were solicited after students had experienced each text in the initial learning phase. The JOL data were collapsed across final test type (short-answer vs. concept map) and analyzed separately for each text type. For enumeration texts, JOLs were higher after students created concept maps than after they practiced retrieval (*M*s = .54 and .47, respectively, $F(1, 59) = 4.73$, $\eta_p^2 = .07$).

Likewise, for sequence texts, JOLs were higher following elaborative studying with concept mapping than following retrieval practice (*M*s = .55 and .48, respectively, $F(1, 59) = 5.64$, $\eta_p^2 = .09$).



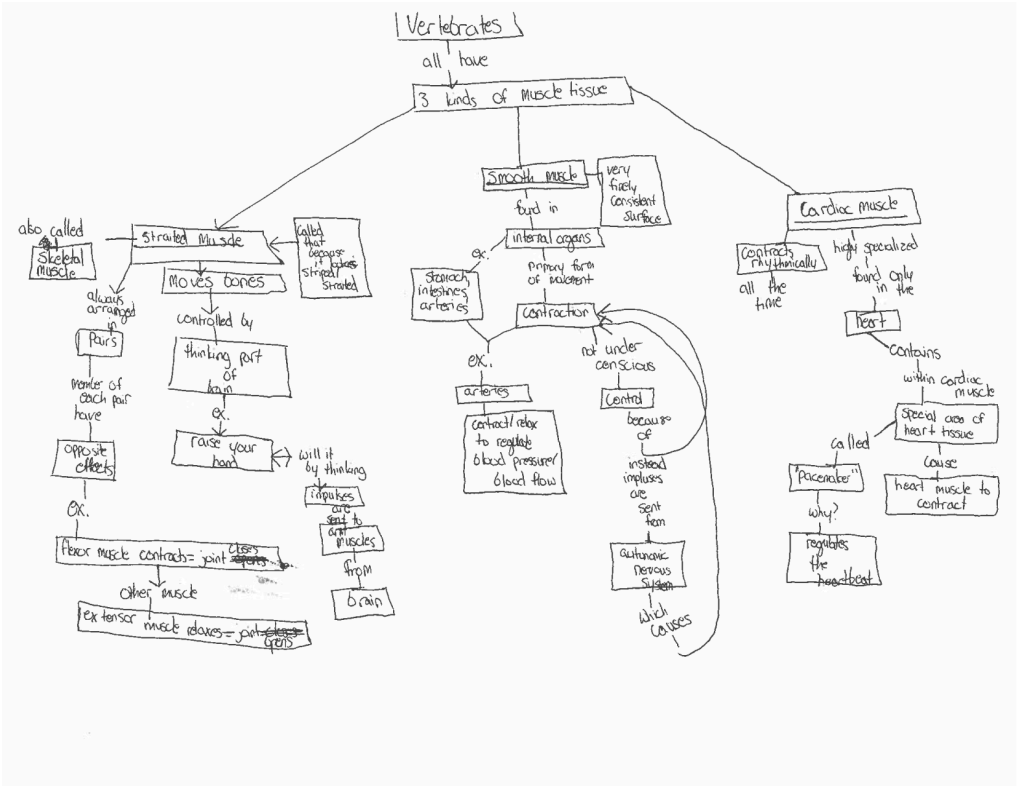


Fig. S1. Examples of concept maps created by students in Experiment 2.

Table S1

Examples of verbatim and inference questions used in Experiments 1 and 2

Experiment 1: Sample questions from text on "Sea Otters"
Verbatim Question:
"What does sea otter fur consist of?"
(Answers: Sea otters have a double-layered fur, with a coarse outer layer and a finer inner layer)
Inference Question:
"What would be the consequences of removing sea otters from their environment?" (Answers: There would be a lack of protection of kelp and seaweed, because fewer otters would eat the invertebrates that destroy kelp and seaweed. The presence of more invertebrates would change the ecosystem.)
Experiment 2: Sample questions from text on "Make-Up of Human Blood"
Verbatim Question:
"What happens when hemoglobin combines with oxygen?"
(Answers: 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?"
(Answers: Blood would not clot, because fibrin is needed to form a meshwork of fibers that trap blood cells and aid in clotting.)

References

S1. H. L. Roediger, J. D. Karpicke, *Psychol. Sci.,* **17***,* 249 (2006).

S2. J. D. Novak, *Res. Sci. Ed.* **35***,* 23-40 (2005).

S3. J. D. Karpicke, H. L. Roediger, *Mem. Cog.,* **38**, 116 (2010)

S4. J. Cohen, *Statistical Power Analysis for the Behavioral Sciences* (2nd Ed., Erlbaum, Hillsdale NJ, 1988)

S5. L. K. Cook, R. E. Mayer, *J. Ed. Psy.,* **80***,* 448-456 (1988).