Journal of Applied Research in Memory and Cognition 3 (2014) 198–206

Contents lists available at ScienceDirect

Journal of Applied Research in Memory and Cognition

journal homepage: www.elsevier.com/locate/jarmac

Retrieval-based learning: The need for guided retrieval in elementary school children

Jeffrey D. Karpicke*, Janell R. Blunt, Megan A. Smith, Stephanie S. Karpicke

Purdue University, United States

ARTICLE INFO

Article history:
Received 20 March 2014
Received in revised form 11 July 2014
Accepted 22 July 2014
Available online 30 July 2014

ABSTRACT

Three experiments were aimed at adapting retrieval practice techniques that are effective with college students to work with elementary school children. Children participated in their classrooms and completed activities with educational texts selected from the school curriculum. In Experiment 1, when children were asked to freely recall the texts, they recalled very little of the material (about 10%) and showed almost no improvement after rereading. In another condition that involved creating concept maps, the children produced only about 20% of the ideas on their maps, even though they viewed the texts during the entire activity. Experiments 2 and 3 explored ways to provide support during retrieval activities. In Experiment 2, children were very successful at retrieving knowledge on concept maps that were partially completed. In Experiment 3, a question map activity, where questions were displayed in a relational map format, was effective for guiding retrieval practice and improving learning relative to repeated studying. The results demonstrate the importance of examining strategies that work with college students with young children in educational settings using authentic materials. The results also highlight the need for guided retrieval practice in young children.

1. Introduction

Retrieval practice is a powerful way to improve learning. The general idea, which we refer to as *retrieval-based learning*, does not have to be complicated to implement. Essentially, retrieval practice simply involves having learners set aside the material they are learning and practice actively reconstructing it on their own. When students are capable of successfully retrieving knowledge, and when they practice doing it repeatedly, retrieval practice will promote learning that is robust, durable, and transferable to new contexts (Carpenter, 2012; Karpicke, 2012). Retrieval practice could be incorporated into a variety of existing educational activities. That is, many activities could be converted into retrieval-based learning activities simply by having students complete them in the absence of the to-be-learned material, so students are required to remember what they experienced in previous study episodes.

Because retrieval-based learning can be done in a variety of ways, a general challenge is to develop and test new ways to implement retrieval practice in educational settings. There are several criteria that are necessary to accomplish this translation from laboratory work to classrooms. Namely, research on learning strategies like retrieval practice needs to be conducted with actual educational materials selected directly from school curricula. In addition, learning activities need to be designed so that they could be implemented in classroom settings. Therefore, it follows that experimental work on such learning activities should be conducted in actual classrooms.

Perhaps most importantly, retrieval-based learning tasks need to be scaled to work with younger children. The vast majority of research on retrieval practice has been carried out with college students, and very little has been done to examine retrieval practice effects in children, specifically those in the late elementary grades (see Blunt & Karpicke, 2014b; Dunlosky, Rawson, Marsh, Nathan, & Willingham, 2013). Grades 3–5 represent a critical time in children's educational development, because by these grade levels children have learned to read, and they are now increasingly "reading to learn." That is, late elementary school children are beginning to read material and implement strategies on their own in order to learn from what they are reading. Thus it is essential to examine retrieval practice in elementary school children.

Some recent work has focused on applying retrieval practice to learning in young children. That research has confirmed that children do exhibit retrieval practice effects in simple word-pair experiments (Blunt & Karpicke, 2014b), can learn educationally relevant

* Corresponding author at: Department of Psychological Sciences, Purdue University, 703 Third Street, West Lafayette, IN 47907-2081, United States.
Tel.: +1 765 494 0273; fax: +1 765 496 1264.
E-mail address: karpicke@purdue.edu (J.D. Karpicke).

http://dx.doi.org/10.1016/j.jarmac.2014.07.008
2211-3681/

facts from multiple-choice tests (Marsh, Fazio, & Goswick, 2012), and benefit from retrieval practice of vocabulary words (Goossens, Camp, Verkoeijen, & Tabbers, 2014). In the present experiments, our goal was to examine retrieval-based strategies for learning from educational texts. The three experiments reported here were carried out with children in elementary school classrooms to identify retrieval activities with the potential to promote learning as well as those that may *not* work well with young children, even though the activities are quite effective with college students.

Regardless of the students' ages, the nature of the materials, or the setting of the learning tasks, there are a few critical features of retrieval practice that must be present in order for retrieval to promote learning (see Karpicke, Lehman, & Aue, 2014). First, retrieval practice tasks must afford successful retrieval. It would be foolish merely to give children "tests" and assume that testing will promote learning, because learning is enhanced by the act of retrieval, not testing, per se. Second, retrieval practice enhances learning when learners must *reinstate a prior context* during retrieval (Karpicke & Zaromb, 2010; Karpicke et al., 2014; Lehman, Smith, & Karpicke, in press). Massed retrieval immediately after experiencing items might guarantee retrieval success, but it obviates the context reinstatement and updating processes that promote learning, so massed retrieval is ineffective (Carpenter & DeLosh, 2005; Karpicke & Bauernschmidt, 2011; Karpicke & Roediger, 2007). An effective retrieval practice task will afford both retrieval success and context reinstatement. Third, free recall tasks have features that make them especially effective for learning. Free recall provides little support in the immediate environment, so learners must engage in a great degree of context reinstatement. Free recall also requires learners to establish an organizational strategy to guide recall output, sometimes called a retrieval structure (Raaijmakers & Shiffrin, 1981), and learners must recover the individual items or elements within that structure. However, free recall may be problematic if it results in very little retrieval success, which is a risk because the task affords little external support. We had these elements of retrieval practice in mind when we conducted the present experiments.

Free recall retrieval practice has been shown to be effective in college students, and findings with college students were the impetus for the present project with elementary school children. To provide a reference point for interpreting the results reported in this article, we describe an experiment by Karpicke and Blunt (2011). In that study, college students (ages 18–22) read brief educational texts and practiced retrieval by freely recalling them. The students read the texts for 5 min, freely recalled the texts in a 10 min recall period, then reread the texts and recalled again. Performance on the recall tasks was scored as the proportion of idea units recalled. Students in Karpicke and Blunt's first experiment recalled 64% of the ideas on the first recall and improved to 81% recall after rereading the text. Practicing free recall greatly enhanced long-term retention, relative to repeatedly studying the material, on a delayed short answer test. Similar results have been obtained in other work. For instance, Karpicke and Roediger (2010) also examined free recall of brief texts and obtained similar results, with students improving from about 50% to 80% recall across multiple rereading and recall cycles. Repeated free recall produced a large enhancement on a delayed final test, relative to reading the material one time.

Karpicke and Blunt (2011) also examined another technique that is popular in educational settings known as concept mapping (Novak & Gowin, 1984). Students create concept maps by making node-and-link diagrams, where the concepts within a set of materials are represented as nodes, the relations among concepts are represented as links connecting the nodes, and labels describing the relations are written next to the links. Karpicke and Blunt had college students create concept maps while they read texts, in order to promote elaborative studying and ensure high levels of success on the task. Indeed, students were able to create concept maps easily, and on average they included about 80% of the ideas from the texts on their concept maps. Nevertheless, Karpicke and Blunt showed that practicing retrieval produced better long-term performance than elaborative studying with concept mapping, although concept mapping did produce gains relative to studying the material once.

In subsequent work, Blunt and Karpicke (2014a) showed that concept mapping could work as an effective technique when it was done as a retrieval-based learning activity. That is, when students practiced retrieval by creating concept maps without viewing the material, this retrieval-based concept mapping activity improved long-term retention to the same extent as recalling material in paragraph format (which was the retrieval practice method used by Karpicke & Blunt, 2011; Karpicke & Roediger, 2010; and Roediger & Karpicke, 2006, among others). Concept mapping is effective when it is implemented as a retrieval-based learning activity, and it may work well as a retrieval practice activity for young children. Retrieval-based concept mapping tasks could be designed to provide retrieval support, for example, by providing students with the overall relational structure of the node-and-link diagram, or by having some portion of the map already filled in for students. At the same time, the activity would still require students to practice retrieval of knowledge from a prior study context. Retrieval-based concept mapping was one of the retrieval practice techniques explored in the present experiments.

Three experiments are reported in the present article. The first experiment was an attempt to implement methods that are effective with college students, as shown by Karpicke and Blunt (2011) and other researchers, in a classroom setting with children. Fourth-grade students read brief educational texts, which were chosen in collaboration with the students' teachers from the fourth grade curriculum, and created concept maps or practiced retrieval by freely recalling the texts. The original intent of the experiment was to examine the effects of these learning activities on a delayed test, but it was immediately evident that the key results pertained to children's performance on the initial learning activities themselves. To preview, the levels of performance were surprising in light of the results typically obtained on these activities with college students. The outcome of Experiment 1 led us to explore new ways of guiding retrieval practice with children by using novel concept mapping tasks in Experiments 2 and 3.

2. Experiment 1

In Experiment 1, elementary school-age children read brief educational texts and engaged in four different learning activities. In the concept mapping condition, the children read the text and created a concept map with the text in front of them. In the free recall condition, the children read the text and then attempted to recall as much of the information from the text as they could, without the text in front of them. The children then reread the text and attempted to recall it a second time. In the cued recall condition, the children read the text and were given a sheet with sentence stems as prompts to guide them as they practiced retrieval. The children then reread the text and attempted free recall, without the sentence stems, in the second recall period. This condition allowed us to examine whether initial sentence-cued recall would aid performance during the second free recall period. Finally, for a control condition, children read the text and did not engage in any additional activity. The children then made a series of metacognitive judgments (judgments of learning and ratings of how interesting, difficult, and enjoyable they thought the activities were). The effects of the initial learning conditions were assessed on a final test four days after the original learning

phase that included verbatim, inference, and application short answer questions (Karpicke & Blunt, 2011).

2.1. Method

2.1.1. Subjects

Ninety-four children, ages 9–11, participated in Experiment 1. They were recruited from four general education classrooms in a public elementary school in Indianapolis, Indiana. The school's total student population was 52% African American, 27% Hispanic, 14% Caucasian, and 7% other race/ethnicities. In exchange for participating in the experiment, the children received a gift card to use at their school's semiannual book fair. Parental consent and student assent was received for each child. Not all children participated in all eight experimental sessions, due to absences across the four-week experiment period. Out of the total group, 54 completed all four conditions, 33 completed three conditions, 6 completed two conditions, and one did not complete any condition.

2.1.2. Materials

Four brief texts were adapted from elementary science textbooks (Basca, Burke, Campbell, & Sherman, 2012; Holdren, 2011; MacMillan/McGraw-Hill, 2008a, 2008b). Each text covered a single topic (*Learning More about Rocks, Surviving in the Wild, Classifying Animals*, and *The Earth's Changing Surface*). The texts were 273, 251, 226, and 238 words in length and had Flesch–Kincaid reading levels of 5.3, 6.5, 3.6, and 3.4, respectively. The texts were modified to increase coherence, for example, by replacing pronouns with nouns, replacing difficult or unfamiliar words, and adding explicit connections among concepts in the text (see Graesser, McNamara, Louwerse, & Cai, 2004). The texts were chosen in collaboration with the teachers of the four fourth grade classrooms, based on topics that would be covered in the fourth grade curriculum later in the school year.

2.1.3. Design

The four learning conditions (concept mapping, free recall, cued recall, and the study only control condition) were manipulated within-subjects. The experiment took place across 4 weeks: children completed one of the four learning activities each Monday and took a short answer test each Friday. Thus, children studied four different texts across the four weeks, one in each condition. The order of the texts was held constant, while the order in which children performed the learning activities was counterbalanced across classrooms.

2.1.4. Procedure

The experimental sessions were conducted in classrooms, and the children were tested as a group, with about 25 children per classroom, but were instructed to work independently. The initial learning session was conducted each week on Monday and the final short answer test was conducted each week on Friday (4 days later). During the learning phase, in all conditions the children began by studying a text for 5 min. In the control condition, the children then moved on to other classroom activities, while in the other conditions they completed a learning activity for approximately 20 min. In the concept mapping condition, the experimenter gave a brief introduction to the concept mapping task, using an example about dogs. Children were told that concepts were represented in bubbles and that those concepts were connected to other concepts using words and phrases that described how they related to each other. They were then given a blank sheet of paper and told to create a concept map while viewing the text they had read in the 5-min study phase. The concept mapping task lasted 20 min. In the free recall condition, children were given a blank sheet of lined paper and had 7.5 min to write as much of the text as they could remember, without looking back at the text. They then restudied the text in a 5-min rereading period and recalled it again in a second recall period, lasting 7.5 min. In the cued recall condition, the children were given a sheet with sentences from the text that were missing words or phrases (see Appendix A for an example). They were asked to think back to the text they had just read and write the word or phrase that best completed the sentence. The cued recall task lasted 7.5 min. After completing the task, the children were given 5 min to restudy the text and 7.5 min to freely recall it, using the same instructions used in the free recall condition.

At the end of each learning activity, the children were asked to rate several aspects of the activity, including how well they thought they had learned the material (a judgment of learning), how interesting they found the activity, how difficult they found the activity, and how much fun they thought the activity was. The children made their ratings on a 0–10 scale where 0 was the lowest rating and 10 was the highest rating.

The final short answer tests occurred 4 days after each learning activity and contained three types of questions: verbatim questions, which referred to concepts and ideas that were stated directly in the texts; inference questions, which required children to make inferences about ideas that were not explicitly stated in the text; and application questions, which asked the children to apply what they had learned to new situations (see Anderson et al., 2001). There were 8 verbatim, 3 inference, and 2 application questions. Examples of the short answer questions are shown in Appendix A. The children were encouraged to take as much time as they needed to answer the questions.

2.2. Results

2.2.1. Scoring

The texts were divided into 30 idea units for scoring purposes. Free recall protocols were scored by giving one point for each correctly recalled idea unit (Karpicke & Blunt, 2011; Karpicke & Roediger, 2010). Concept maps were scored by giving one point for each correct idea unit represented in their concept map. Cued recall protocols were scored by giving the students one point for each correct word or phrase completed on the test. The final short answer assessments were scored by giving students 1 point for each correct response and .5 points for partially correct responses.

2.2.2. Initial performance

Table 1 shows the performance results on the initial learning activities. Overall, students did not perform well on any of the activities, and there are several striking results depicted in the table. First, in the free recall condition, the children recalled only 7% and 9% of the idea units on the first and second recalls, respectively. It is worth reiterating that the materials were selected directly from school textbooks and were modified to improve coherence. Free recall of brief texts is feasible for college students; recall

Table 1
Proportion of ideas correctly produced or recalled during initial learning and proportion correct on the final short answer tests in Experiment 1.

	Initial performance		Final short answer	
	Period 1	Period 2	Verbatim	Higher order
Control	–	–	.40 (.27)	.27 (.25)
Concept mapping	.20 (.16)	–	.35 (.28)	.29 (.26)
Free recall	.07 (.07)	.09 (.08)	.34 (.21)	.30 (.25)
Cued recall	.26 (.20)	.10 (.09)	.35 (.24)	.26 (.26)

Note: Standard deviations are reported in parentheses. The concept mapping condition lasted about the same duration as the free and cued recall conditions. On the final short answer test, the inference and application questions have been combined into a single measure labeled "higher order" questions.

that college students in Karpicke and Blunt's (2011) Experiment 1 recalled 64% and 81% of the ideas in a text, in a procedure nearly identical to the one used here with elementary school children. The second striking observation from the data in Table 1 is that children showed virtually no improvement in recall after 5 min of rereading the text. Out of 85 students who completed the free recall condition, about half (42/85 = 49%) showed no gain in recall on the second attempt. Of those students who did improve, the average improvement was only 7% (SD = 5%), or about 2 out of 30 ideas gained by rereading. Third, students performed better on the cued recall task relative to free recall, but the cued recall task also produced relatively low levels of retrieval success. In addition, initial cued recall produced a miniscule benefit on free recall in the second period (10% vs. 9%). Finally, students exhibited relatively poor performance in the concept map condition. Even though they had the material available during the entire activity, the children were able to produce only about 20% of the idea units. In experiments with nearly identical methods, college students produced much more material on their concept maps (e.g., about 60% of the idea units in Blunt & Karpicke, 2014a, and about 80% in Karpicke & Blunt, 2011). Taken together, these results provided preliminary evidence that elementary school children would need much more guidance and support to accomplish successful retrieval practice than what is needed by college students.

2.2.3. Final assessment performance

The right portion of Table 1 shows the proportion correct on the final short answer tests, with data from the inference and application questions combined. Given that students had significant difficulties with the initial learning activities, it is not surprising that there were essentially no effects of the learning activities on the final short answer tests. This observation was confirmed with separate one-way ANOVAs conducted on the verbatim and higher-order question data, which both yielded Fs < 1, and by multilevel analyses that are capable of handling missing data, which occurred due to student absences. The multilevel analyses are provided as Supplementary Material.

2.2.4. Activity ratings

Table 2 shows the students' ratings of the learning activities. The pattern of results in these subjective rating data was similar to the pattern often seen in college students (Blunt & Karpicke, 2014a; Karpicke & Blunt, 2011). The table shows that children gave the highest JOLs in the concept mapping condition, and JOLs were lower in the two retrieval practice conditions relative to the control condition (see Karpicke & Blunt, 2011). Even though the children were not especially successful at completing the concept mapping task, they rated concept mapping as most interesting and enjoyable of the four learning activities.

Table 2
Mean subjective ratings of the learning activities in Experiments 1 and 2.

	Judgment of learning	Interest	Difficulty	Enjoyment
Experiment 1				
Control	5.2 (2.9)	6.5 (3.4)	2.3 (3.2)	5.7 (3.3)
Concept mapping	6.1 (2.9)	7.1 (3.3)	3.3 (3.4)	7.1 (3.4)
Free recall	4.8 (2.8)	6.4 (3.5)	4.3 (3.3)	5.9 (3.6)
Cued recall	4.3 (3.0)	5.7 (3.4)	5.8 (3.7)	5.1 (3.6)
Experiment 2				
More-more	3.7 (1.3)	3.8 (1.4)	1.9 (1.3)	3.8 (1.4)
Less-more	3.6 (1.3)	3.9 (1.3)	2.0 (1.3)	3.9 (1.3)
More-less	3.6 (1.2)	3.8 (1.3)	2.5 (1.4)	3.9 (1.4)
Less-less	3.6 (1.3)	3.7 (1.5)	2.3 (1.4)	3.7 (1.5)

Note: Standard deviations are reported in parentheses. Ratings were made on a scale from 0 (lowest) to 10 (highest).

2.2.5. Discussion

The results of Experiment 1 indicated that some activities known to be feasible and effective with college students, including free recall and concept mapping with minimal support, were not feasible for promoting learning of educational texts with elementary school children. Both concept mapping and free recall can be important learning tools, because both activities require learners to focus on relational and item-specific processing, and because retrieval practice specifically requires learners to reconstruct what occurred in a prior episodic context. The clear challenge, based on the data from Experiment 1, is to adapt activities so that younger learners can benefit from practicing retrieval with educational materials. Experiments 2 and 3 were aimed at exploring possible ways of accomplishing that outcome.

3. Experiment 2

Experiment 2 explored the use of concept mapping as a retrieval-based learning activity (see Blunt & Karpicke, 2014a). Concept mapping can be implemented as a retrieval activity when students create concept maps without viewing the material. The advantage of using a concept mapping task is that it allows us to manipulate the level of support provided during an activity. In Experiment 2, students read educational texts, completed a concept

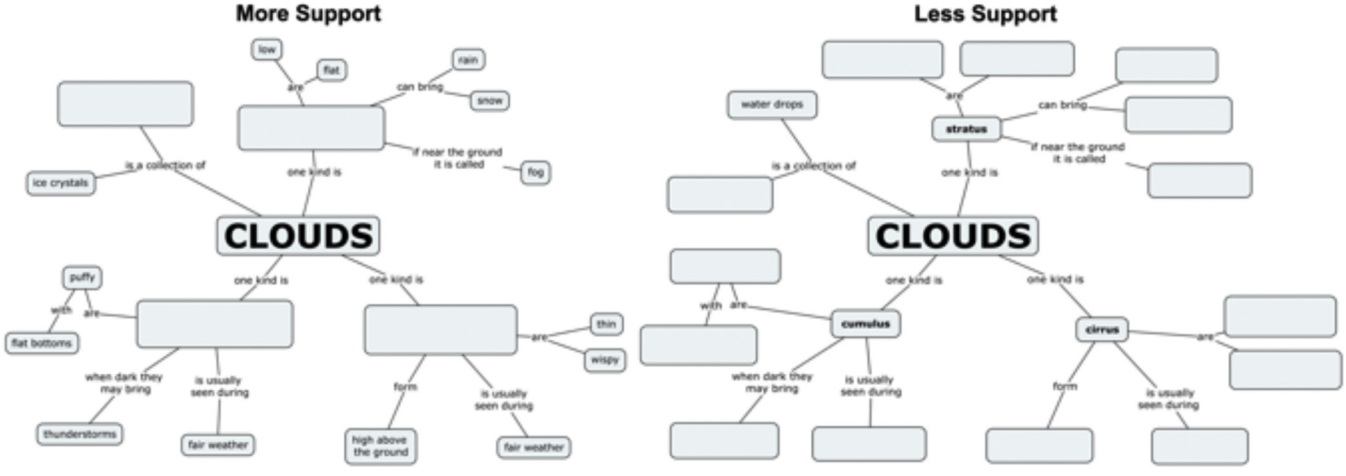

Fig. 1. Examples of the more support (left) and less support (right) concept maps used in Experiment 2. Children were also given a word bank that consisted of the missing concepts.

Table 3
Proportion of ideas correct during initial learning activities, proportion correct on the final fill-in-the-blank test, and effect sizes in Experiment 2.

	Period 1	Period 2	Final performance	*N*	*d*	95% CI
More-more	.91 (.24)	.90 (.24)	.42 (.26)	–	–	–
Less-more	.92 (.17)	.94 (.21)	.48 (.30)	92	0.17	[−0.04, 0.38]
More-less	.90 (.25)	.83 (.29)	.47 (.27)	87	0.17	[−0.04, 0.38]
Less-less	.85 (.22)	.70 (.32)	.46 (.30)	90	0.12	[−0.09, 0.33]

Note: Standard deviations are reported in parentheses. Effect sizes and 95% confidence intervals around the effect sizes were calculated for each condition relative to the more-more condition.

map with the text in front of them, and then practiced retrieval by completing a concept map without the text. There were two levels of support in the concept map task: *more* support, where much of the map was filled in and the children completed 4 concepts, and *less* support, where less of the map was completed and children were required to complete 13–14 concepts. Fig. 1 shows examples of concept maps from the more and less support conditions. The students completed two concept map activities, the first with the text and the second without it, and we manipulated level of support across the two periods. Thus, there were four learning conditions: more-more, more-less, less-more, and less-less. Immediately after the learning phase, the children then took a final fill-in-the-blank test, similar to the sentence cued recall task used in Experiment 1. Note that we did not examine the effects of the retrieval conditions relative to a study-only control condition. The purpose of the experiment was to examine possible ways to bolster initial retrieval success in elementary school children.

3.1. Method

3.1.1. Subjects

One hundred and three children, ages 9–11, participated in Experiment 2. They were recruited from the same classrooms as Experiment 1, and the experiment was carried out in the spring semester, following Experiment 1. Eighty-seven of the children in Experiment 2 had participated in Experiment 1. As in Experiment 1, not all children participated in all four experimental sessions, due to absences across the four-week experiment period. Out of the total group, 79 completed all four conditions, 19 completed three conditions, 3 completed two conditions, one completed one condition, and one did not complete any condition.

3.1.2. Materials

Four brief texts were adapted from two of the elementary textbooks used in Experiment 1 (MacMillan/McGraw-Hill, 2008a, 2008b). Texts were modified to increase text coherence and reduce the Flesch–Kincaid reading level to a 4th grade reading level. Each text covered a single topic (*Clouds, Oceans, Storms*, and *Deserts*). The texts were 151, 124, 123, and 152 words in length, and had Flesch–Kincaid reading levels of 4.4, 4.5, 4.5, and 4.2, respectively.

3.1.3. Design

There were four learning conditions (more-more, more-less, less-more, and less-less), which were manipulated within-subjects. Thus, children studied four texts and engaged in one learning activity for each text. The order of the texts was held constant while the order in which the children performed the learning activities was counterbalanced across classroom.

3.1.4. Procedure

The experiment took place once a week (either Monday or Tuesday) for four sessions during children's normal class time in the spring semester. As in Experiment 1, the experimental sessions were conducted in classrooms, and the children were tested as a group, with about 25 children per classroom, but were instructed to work independently. During the learning phase, the experimenter read the text out loud to children while they read silently on paper, and the experimenter then answered any questions children had about unfamiliar words or pronunciations. Children completed the first concept map with the text in front of them and then completed the second concept map without the text, as a retrieval activity. The link and node structure of the concept maps was intact, but certain nodes were blank. The missing concepts that belonged in the nodes were represented in a word bank that was included to the right of the map. In the *more support* conditions, children viewed maps that were missing 4 concepts, while in the *less support* conditions, children viewed maps that were missing 13 or 14 concepts (see Fig. 1). Children worked at their own pace to complete the maps, but were given no more than 11 min per map. On average, children spent approximately 3–5 min completing the *more support* maps and approximately 6–10 min completing the *less support* maps.

After completing the second map, the children made a series of ratings about the learning task, using the same procedure used in Experiment 1. The children then completed a final fill-in-the-blank test, which was similar to the cued retrieval activity in Experiment 1 (see Appendix A for an example). They were given 5 min to complete sentences with a word or phrase from the text they had studied.

3.2. Results

3.2.1. Scoring

Initial maps were scored by giving one point for each word or concept correctly filled in. The fill-in-the-blank final assessments were scored by giving the students one point for each correct word filled in the blank space on the test.

3.2.2. Initial performance

Table 3 shows performance during the initial learning activities. It was clear that students were much more successful on these activities, regardless of the level of support, than they were on the recall and concept map activities in Experiment 1, which provided minimal support. A one-way ANOVA on the first map data yielded $F(3, 234) = 1.82$, $\eta_p^2 = .02$, and an ANOVA on the second map data yielded $F(3, 234) = 18.45$, $\eta_p^2 = .19$, which reflects the fact that students tended to perform better on the retrieval task with more support than they did with less support, Ms = .93 vs. .77, $t(97) = 6.20$, $d = 0.63$ [0.41, 0.84]. Looking across the two experiments, it is noteworthy that even in the less support condition, students were much more successful at recalling ideas (77%) than they were at producing a concept map that contained the ideas from scratch (where they produced 20% of the ideas in Experiment 1), even with the text present during the entire activity.

3.2.3. Final assessment performance

Table 3 also shows results on the final fill-in-the-blank assessment. There were small differences among the four conditions, favoring the conditions that experienced less support during the learning activity relative to the condition with the most support (the more-more condition). The table shows the effect sizes for the differences between the three less-support conditions and the more-more condition, respectively, which ranged from

$d = 0.12–0.17$. The effect sizes are small, but they hint at a general advantage of concept map activities that provided less support relative to the condition that provided the most support and thus required the least amount of retrieval effort.

3.2.4. Activity ratings

The bottom of Table 2 shows students' subjective ratings of the four learning activities. There were virtually no differences in the students' ratings of the four conditions in terms of judgments of learning, interest, or enjoyment; one-way ANOVAs on each rating yielded Fs < 1. Not surprisingly, students rated the conditions with less support during the retrieval activity as more difficult than the conditions with more support, $F(3, 219) = 6.05$, $\eta_p^2 = .08$.

3.2.5. Discussion

Experiment 2 explored a possible approach to guiding retrieval practice with children by using concept maps as retrieval activities and by manipulating the amount of support provided by the concept maps. The results were promising and showed that students were much more successful at retrieving concepts in these activities than they were in the activities in Experiment 1, which offered minimal support during the tasks. There was a trend suggesting that experiencing a condition with less support, in which students completed 13–14 concepts, was more effective than experiencing only the most support, in which students completed only 4 concepts. The results of Experiment 2 suggested that concept mapping could serve as an effective retrieval activity with children because the activity affords the opportunity to provide retrieval support. Experiment 3 leveraged this idea and added an additional element of support to the concept mapping task.

4. Experiment 3

One of our observations in the previous experiments was that children sometimes found the concept mapping task confusing. When the children were asked informally, they could often recall more of the material than what they included on their maps or recall protocols (see too Blunt & Karpicke, 2014a). Experiment 3 explored a new technique that we refer to as *question mapping*. In the question map task, students completed a set of questions that were arranged in a map format, as shown in Fig. 2. The purpose of the task was to provide students with more effective cues for retrieving the material while also emphasizing how the questions related to the central topic of the material, which is an inherent advantage of a concept mapping task. Experiment 3 included two conditions. In the guided retrieval condition, the children first completed a question map while viewing a text and then completed a question map without the text as a retrieval practice activity. In the restudy condition, instead of completing the question maps, the students spent additional time rereading and studying the text. The effects of the two activities were assessed on a final free recall test, which occurred immediately after completing each activity.

4.1. Method

4.1.1. Subjects

Eighty-nine children, ages 9–11, participated in Experiment 3. They were recruited from the same school and classrooms as Experiments 1 and 2 in the year after those experiments were completed. Thus, none of the children had participated in the previous experiments. Three students did not complete the learning activities and one student was not able to complete the free recall tests. Thus, the results are based on data from 85 students.

4.1.2. Materials

Two brief texts were developed: a modified version of the *Clouds* text from Experiment 2, and another text, *Apache Indian Homes*, was adapted from a children's website (Technological Solutions, Inc., 2013). The texts were 68 and 75 words in length and had 15 and 18 idea units, respectively. Both texts had a Flesch–Kincaid reading level of 4.4.

4.1.3. Design

The two learning conditions (guided retrieval vs. restudy) were manipulated within-subjects. Children studied two texts and engaged in a guided retrieval activity with one text and restudied the other. The order of the two texts and the order

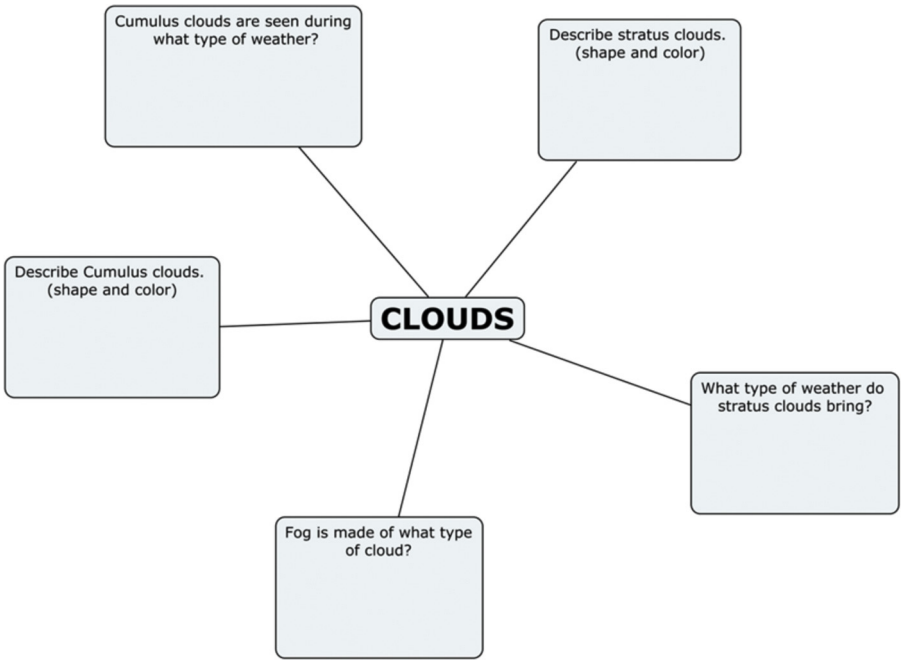

Fig. 2. Example question map used in Experiment 3.

in which children performed the two learning activities were counterbalanced across classrooms.

4.1.4. Procedure

The experiment took place in a single session during the children's normal class period. As in the previous experiments, the session was conducted in classrooms, with about 25 children per classroom, and the children were instructed to work independently. In both conditions, the experimenter read the text out loud while the children read silently on paper, and the experimenter then answered any questions the children had about unfamiliar words or pronunciations. In the guided retrieval condition, the children were given a question map, which included 5 questions in boxes arranged around the title of the text (see Fig. 2). The children completed the map by writing their answers to the questions in the boxes, and they viewed the text the entire time during the first period. The first question map task lasted approximately 5 min, though children were given extra time to finish their answers if needed. The children were then given a second identical question map to complete without viewing the text, as a retrieval practice activity. The second question map task lasted 2–3 min. In the reread condition, after the experimenter had read the text the children were given approximately 8 min of additional time to reread and study the text. Thus, total time was matched as closely as possible across the two conditions, within the constraints of testing large groups of children in classroom settings.

Experiment 3 did not include subjective ratings of the activities. Instead, after completing each learning activity, the children proceeded to the criterial free recall test. The procedure for the free recall test was the same as the one used in Experiment 1: the children were given a blank sheet of paper and told to write down as much of the material as they could remember from the text. To help encourage the children to recall more material, the experimenter told them to think about what they would tell a friend if they were explaining what they had learned about in the text. The children were given as much time as they needed to complete the free recall task, and most children had finished recalling within 5–7 min.

4.2. Results

4.2.1. Scoring

The question maps were scored by giving 1 point for correct responses and .5 points for partially correct responses. Free recall protocols were scored by giving one point for each correctly recalled idea unit.

4.2.2. Initial performance

Overall, the children were rather successful in completing the question map activity both with and without the text. The students answered 82% (SEM = 2%) of the questions with the text and answered 74% (SEM = 2%) without the text in the second period. The high percentage of correct answers on the guided retrieval activity suggests that completing the activity with the text may have helped improve performance during the retrieval task when the text was absent. There were small differences across the two texts. The children answered more questions correctly about the *Clouds* text than about the *Apache Indians* text both with the text (.92 vs. .72), $t(84) = 5.56$, $d = 0.60$ [0.37, 0.83] and without the text (.79 vs. .69), $t(84) = 2.30$, $d = 0.50$ [0.07, 0.93].

4.2.3. Free recall performance

The other key result of Experiment 3 is shown in Fig. 3, which shows the proportion of ideas recalled on the criterial free recall test. There was an advantage of guided retrieval over restudying, *Ms* = .40 vs. .28, $t(84) = 3.89$, $d = 0.42$ [0.20, 0.65], indicating that the guided retrieval method used here was effective for promoting subsequent learning performance. The effect of guided retrieval practice was large for the *Clouds* text (.35 vs. .19), $t(84) = 3.89$, $d = 0.82$ [0.40, 1.28] and was less pronounced but still evident for the *Apache Indians* text (.43 vs. .36), $t(84) = 1.66$, $d = 0.36$ [−0.07, 0.79]. Although this is a cross-experiment comparison, and there are a number of differences across the experiments, it is noteworthy that the guided retrieval practice condition produced much better free recall performance (40%) than the free recall conditions in Experiment 1, where fewer than 10% of ideas were recalled even on a second recall attempt.

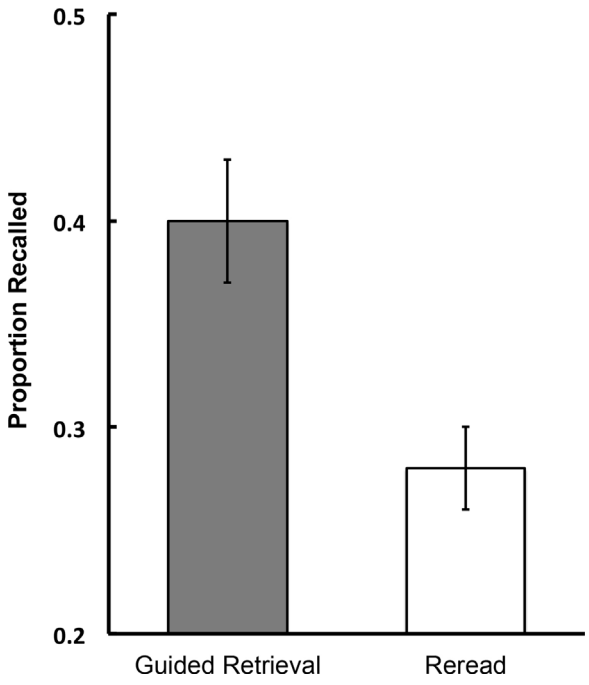

Fig. 3. Proportion correctly recalled on the final free recall test in Experiment 3. Error bars represent standard error of the mean. Children recalled more ideas when they practiced retrieval with question maps (guided retrieval) than they did when they reread.

4.2.4. Discussion

Experiment 3 demonstrated the efficacy of a guided retrieval procedure using a question mapping task. Children performed well on the question mapping task under conditions with and without the text, and the task produced a positive enhancement in free recall performance. The results provide preliminary evidence indicating that question mapping may serve as an effective retrieval-based learning activity. The results also suggest that, under certain conditions, free recall of educational texts is indeed a feasible task for young children when appropriate scaffolds are in place to guide and support children's successful recall. Whereas conditions with minimal support produced poor free recall in Experiment 1, the guided retrieval tasks used in Experiment 3 were effective for boosting free recall of educational texts in children.

5. General discussion

The purpose of the present experiments was to examine retrieval-based learning in elementary school children and to identify effective and feasible retrieval activities for children. A current challenge is to scale retrieval practice activities that work with college students so that they also work with younger children. Learning activities also need to be designed so that they may be implemented in classroom settings. Thus, the present work was carried out in elementary school classrooms with educational

content selected directly from curriculum materials. The experiments were also designed based on the elements of effective retrieval practice tasks. Retrieval-based learning activities must afford high levels of successful retrieval and must also require learners to reinstate a prior study context (Karpicke et al., 2014).

The experiments produced some surprising results. Experiment 1 showed that activities that are easy for college students (free recall and concept mapping of educational texts) were extremely difficult for elementary school children, while Experiments 2 and 3 showed that adaptations of those tasks were much more feasible for children. In the first experiment, children were only able to produce 20% of the ideas when they were asked to create concept maps from scratch, even though the text was available the entire time. Contrast that with the results of Experiment 2, where in all of the conditions with support during the activities, the children were able to produce much higher percentages of the ideas on their maps, with and without the text in front of them. Likewise, in Experiment 3, children were able to answer 82% of the questions correctly on their initial question maps, which they completed while viewing the texts. Thus, even when the to-be-learned materials were provided throughout an entire activity, children benefitted greatly from additional support within the activity, in the form of partially complete maps or specific questions.

The experiments also shed some light on what kinds of activities will and will not support successful immediate free recall in children. Experiment 1 showed that children produced very few ideas when they were asked to freely recall brief texts, even though the texts were selected from the curriculum materials and were modified to improve cohesion. Moreover, children gained very little after rereading the text and recalling again. Providing sentence cues improved initial recall to some extent, but that task did not help promote high levels of free recall after rereading. Experiment 3 showed that completing a question map transferred and enhanced performance on an immediate free recall task. Free recall remains an important activity because it requires learners to create a relational retrieval structure and then recover the individual ideas within that structure, two sets of processes that are critical for promoting long-term learning. Ultimately, effective retrieval-based learning activities will need to incorporate scaffolding, so that children are required to retrieve greater amounts of material across iterations and can do so successfully.

5.1. Practical applications

Emerging work has confirmed that young elementary school-aged children do indeed benefit from the mnemonic effects of retrieval (Blunt & Karpicke, 2014b; Marsh et al., 2012; Goossens et al., 2014). To identify and develop retrieval-based learning activates that work in educational settings, there is a continuing need for research carried out with children in classrooms using materials from the curriculum. The present results point up the need for guided retrieval practice with elementary school children and provide initial insights into how to structure and scaffold retrieval practice activities when children are learning from educational texts.

Conflict of interest

The authors declare that they have no conflict of interest.

Acknowledgements

We are grateful to Fox Hill Elementary School, principals Kathy Levine and Sean Taylor, fourth grade teachers Rhonda Jenkins, Linda Perez, Elizabeth Christy, Erin Cecil, Desiree Brooks, James Valdez, and Lindsey Cash, and all of the 2011-2013 students and parents. This research was supported by a grant from the Institute of Education Sciences in the U.S. Department of Education (R305A110903), and additional support during the writing of this article was provided by Grants from the National Science Foundation (DRL-1149363 and DUE-1245476).J.R.B. is supported by a National Science Foundation Graduate Research Fellowship (2012124747). The opinions expressed in this article are those of the authors and do not represent the views of the National Science Foundation, the Institute of Education Sciences, or of the U.S. Department of Education.

Appendix A.

Examples of short answer questions and cued recall prompts used in Experiments 1 and final test questions used in Experiment 2.

Experiment 1: Sample questions from text on "Layers of the Earth"
Verbatim Question:
"What is the thinnest layer of the earth?"
(Sample Answer: the crust)

Inference Question:
"Name one way the mantle and core are the same."
(Sample Answer: Both have a combination of solid and molten rock)

Application Question:
"When a volcano erupts, semi-molten rock is brought to the surface of the earth. Which layer of the earth does this come from?"
(Sample Answer: Mantle because the lower portion of the mantle is soft and semi-molten)

Cued Recall:
"The layer below the crust is the _____."
(Sample Answer: mantle)

Experiment 2: Sample final test question from text on "Clouds"
"A cloud is a collection of tiny _____ crystals in the air."
(Sample Answer: ice)

Appendix B. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jarmac.2014.07.008.

References

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

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

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

Blunt, J. R., & Karpicke, J. D. (2014b). *Retrieval-Based Learning: Positive Effects of Retrieval Practice in Elementary Children*. Purdue University. Unpublished manuscript

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

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

Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving students' learning with effective learning techniques: Promising directions for cognitive and educational psychology. *Psychological Science in the Public Interest*, *14*, 4–58.

Graesser, A. C., McNamara, D. S., Louwerse, M. M., & Cai, Z. (2004). Coh-Metrix: Analysis of text on cohesion and language. *Behavior Research Methods, Instruments, and Computers*, *36*, 193–202.

Goossens, N. A. M. C., Camp, G., Verkoeijen, P. P. J. L., & Tabbers, H. K. (2014). The effect of retrieval practice in primary school vocabulary learning. *Applied Cognitive Psychology*, *28*(1), 135–142.

Holdren, A. (2011). Habitats. In D. Sherman (Ed.), *Student reference book*. Chicago, IL: Chicago Educational Publishing Company, LLC.

Karpicke, J. D., & Roediger, H. L. (2007). Expanding retrieval practice promotes short-term retention, but equally spaced retrieval enhances long-term retention. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*, 704–719.

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

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.
Karpicke, J. D. (2012). Retrieval-based learning: Active retrieval promotes meaningful learning. *Current Directions in Psychological Science*, *21*, 157–163.
Karpicke, J. D., & Bauernschmidt, A. (2011). Spaced retrieval: Absolute spacing enhances learning regardless of spacing. *Journal of Experimental Psychology: Learning Memory, and Cognition*, *37*, 1250–1257.
Karpicke, J. D., & Blunt, J. R. (2011). Retrieval practice produces more learning than elaborative study with concept mapping. *Science*, *331*, 772–775.
Karpicke, J. D., 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.
Lehman, M., Smith, M.A., Karpicke, J.D. (in press). Toward an episodic context account of retrieval-based learning: Dissociating retrieval practice and elaboration. *Journal of Experimental Psychology: Learning, Memory, and Cognition*.
MacMillan/McGraw-Hill. (2008a). *Science (grade 3): A closer look*. Columbus, OH: MacMillan/McGraw-Hill.
MacMillan/McGraw-Hill. (2008b). *Science (grade 4): A closer look*. Columbus, OH: MacMillan/McGraw-Hill.
Marsh, E. J., Fazio, L. K., & Goswick, A. E. (2012). Memorial consequences of testing school-aged children. *Memory*, *20*, 899–906.
Novak, J. D., & Gowin, D. B. (1984). *Learning how to learn*. New York: Cambridge University Press.
Raaijmakers, J. G. W., & Shiffrin, R. M. (1981). Search of associative memory. *Psychological Review*, *88*, 93–134.
Roediger, H. L., & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science*, *17*, 249–255.
Technological Solutions, Inc. (2013). *Native American history for kids: Apache tribal peoples*. Available from: http://www.ducksters.com/history/native_american_apache.php