Journal of Experimental Psychology: Applied
2018, Vol. 24, No. 1, 43–56

1076-898X/18/$12.00 http://dx.doi.org/10.1037/xap0000133

Improving Self-Regulated Learning With a Retrieval Practice Intervention

Robert Ariel and Jeffrey D. Karpicke
Purdue University

Repeated retrieval practice is a powerful learning tool for promoting long-term retention, but students use this tool ineffectively when regulating their learning. The current experiments evaluated the efficacy of a minimal intervention aimed at improving students' self-regulated use of repeated retrieval practice. Across 2 experiments, students made decisions about when to study, engage in retrieval practice, or stop learning a set of foreign language word pairs. Some students received direct instruction about how to use repeated retrieval practice. These instructions emphasized the mnemonic benefits of retrieval practice over a less effective strategy (restudying) and told students how to use repeated retrieval practice to maximize their performance—specifically, that they should recall a translation correctly 3 times during learning. This minimal intervention promoted more effective self-regulated use of retrieval practice and better retention of the translations compared to a control group that received no instruction. Students who experienced this intervention also showed potential for long-term changes in self-regulated learning: They spontaneously used repeated retrieval practice 1 week later to learn new materials. These results provide a promising first step for developing guidelines for teaching students how to regulate their learning more effectively using repeated retrieval practice.

Keywords: retrieval practice, self-regulated learning, testing effect, strategy training

Applying empirically supported cognitive techniques to improve educational outcomes requires that educators can show students how to use these techniques and that students will spontaneously adopt them to learn course content. Although cognitive and educational psychologists have identified a number of promising techniques for improving academic achievement (for a recent review, see Dunlosky, Rawson, Marsh, Nathan, & Willingham, 2013), limited research exists concerning whether students can use and will adopt these techniques to regulate their learning. Retrieval practice, in particular, is one powerful learning tool for promoting long-term retention with overwhelming empirical support (for a comprehensive review, see Karpicke, 2017). Retrieval practice is effective across a range of age groups spanning all levels of education from elementary school (Goossens, Camp, Verkoeijen, & Tabbers, 2014; Karpicke, Blunt, Smith, & Karpicke, 2016) to middle school (Carpenter, Pashler, & Cepeda, 2009; McDaniel, Agarwal, Huelser, McDermott, & Roediger, 2011) to college aged students (Hopkins, Lyle, Hieb, & Ralston, 2015; Jones, 1923). Despite the overwhelming empirical support for this potent learning tool, only a few experiments have examined students' self-regulated use of retrieval practice strategies (Dunlosky & Rawson, 2015; Karpicke, 2009). This limited research suggests that students prefer not to use retrieval practice strategies—instead preferring to use other strategies like restudying—and when they do use a retrieval practice strategy, they do so ineffectively (Karpicke, 2009).

One way students use retrieval practice ineffectively is that they use a "one-and-done" strategy where they drop material from study after recalling it once during a single study session (Dunlosky & Rawson, 2015; Karpicke, 2009; Kornell & Son, 2009). Repeated successful retrieval fosters better retention than a single correct retrieval attempt (Grimaldi & Karpicke, 2014; Karpicke, 2012) and one empirically supported guideline for using retrieval practice to maximize long term retention is for students to recall information correctly a minimum of three times during each study session (Karpicke, 2009; Karpicke & Roediger, 2008; Rawson & Dunlosky, 2011; Vaughn & Rawson, 2011). Can students be trained to use retrieval practice more effectively to regulate their learning? The current experiments evaluated the efficacy of a minimal intervention aimed at improving learners' study habits by correcting their inaccurate metacognitive knowledge about the efficacy of retrieval practice as a learning technique.

Learners in general seem to believe that retrieval practice is useful for monitoring memory, but they fail to grasp the memorial benefits it provides (Hartwig & Dunlosky, 2012; Kornell & Son, 2009). For example, they incorrectly report that restudying is more effective for learning than self-testing (Agarwal, Karpicke, Kang, Roediger, & McDermott, 2008; Ariel, Hines, & Hertzog, 2014; Karpicke & Blunt, 2011; Roediger & Karpicke, 2006; Tullis, Finley, & Benjamin, 2013). Educators share this view and also use tests primarily as an assessment tool in their classrooms (More-

This article was published Online First June 12, 2017.

Robert Ariel and Jeffrey D. Karpicke, Department of Psychological Sciences, Purdue University.

This research was supported in part by Spencer Foundation Grant 201700034, National Science Foundation Grants DRL-1149363 and DUE-1245476, and the Institute of Education Sciences in the U.S. Department of Education Grant R305A150546. The opinions expressed are those of the authors and do not represent the views of the Spencer Foundation, National Science Foundation, the Institute of Education Sciences, or the U.S. Department of Education. We thank Ellis Farr, Kelsey Rogers, Katie Rauscher, and Samantha Wagner for their assistance with data collection.

Correspondence concerning this article should be addressed to Robert Ariel, Department of Psychological Sciences, Purdue University, West Lafayette, IN 47907. E-mail: rariel@purdue.edu

head, Rhodes, & DeLozier, 2016). These inaccurate metacognitive beliefs could undermine students' effective use of retrieval practice during study, causing them to choose not to engage in repeated self-testing of material. Consistent with this conclusion, the limited research examining self-regulated use of retrieval practice indicates that people do not repeatedly self-test. Instead they prefer to restudy material, and when they do self-test, they drop material after being recalled once (Karpicke, 2009; Kornell & Bjork, 2008).

Students' use of a one-and-done strategy, although suboptimal, is unsurprising given that they view retrieval practice as an assessment tool. Theories of self-regulated study such the discrepancy reduction theory (Dunlosky & Hertzog, 1998), the region of proximal learning theory (Metcalfe, 2002; Metcalfe & Kornell, 2005), the hieratical model (Thiede & Dunlosky, 1999), and agenda-based regulation framework (Ariel, Dunlosky, & Bailey, 2009; Dunlosky & Ariel, 2011) all propose that learners' decisions about what material to learn are influenced by their monitoring of their learning. Although, these theories differ in their predictions about what material learners will allocate the most time to learning, a key assumption of each theoretical approach is that learners will not study material that they believe they already know.

Monitoring learning is a heuristic process that involves inferring memory quality based on the cues available to the learner at the time of their monitoring judgment (Koriat, 1997). Retrieval practice provides the learner with cues that are typically more diagnostic of later memory performance than the cues available to them when studying information (Dunlosky & Nelson, 1994; Son & Metcalfe, 2005). This may lead students to overvalue retrieval-based cues such as accessibility and retrieval fluency when monitoring their learning, causing them to prematurely drop material from learning. Consistent with this conclusion, Karpicke (2009) observed that retrieval fluency simultaneously inflated judgments of learning and increased the probability that material would be dropped during learning.

Assuming that students use retrieval practice ineffectively because they lack appropriate metacognitive knowledge, correcting this knowledge could lead to improved self-regulatory behavior. Tullis et al. (2013) conducted a series of experiments examining whether people can learn from task experience that retrieval practice is better for learning than restudying. They found that people had difficulties attributing enhanced learning to the use of a retrieval practice strategy. Students only changed their beliefs about the effectiveness of each strategy when they were provided with extensive environmental support in the form of computer-based performance summaries for material either studied or tested. Even then, strategy knowledge only improved for 50% of participants. These results suggest that learning about the effectiveness of retrieval practice is unlikely to occur through experience and self-discovery alone. Instead, an instructional intervention may be warranted.

One potentially easy way to correct inaccurate metacognitive knowledge and promote spontaneous strategy use is to alert people to the effectiveness of a given strategy (Borkowski, Carr, & Pressley, 1987). This type of direct instruction is most effective when learners are also given detailed instructions about how to implement a strategy and why that strategy is effective (O'Sullivan & Pressley, 1984). Direct instruction may be more effective for initial strategy acquisition than other approaches that rely on experiential discovery because students do not need to initially generate the appropriate strategy on their own (Klahr & Nigam, 2004, but see Alfieri, Brooks, Aldrich, & Tenenbaum, 2011). Moreover, direct instruction is as effective as experiential discovery-based learning approaches for promoting transfer of strategy use to new learning contexts (McDaniel & Schlager, 1990).

In the current experiments, we attempted to correct students' inaccurate metacognitive knowledge about the memorial benefits of retrieval practice by instructing them that repeated retrieval practice is more beneficial for learning than repeated studying. Students were also given detailed instructions about how to use repeated retrieval practice to maximize retention of content. Specifically, they were told to learn items to a criterion of three correct recalls before dropping them from practice and were also provided an illustration that highlighted the effectiveness of repeated retrieval over repeated study strategies.

Across two experiments, we compared self-regulated learning strategies of people who experienced this minimal strategy intervention (retrieval practice instructions group) to a control group who experienced no strategy instructions. Both groups were allowed to regulate their learning of foreign language vocabulary words using a flashcard-like computer-based learning environment. Students regulated their learning by making decisions about whether to study, self-test, or drop translations from learning across multiple self-controlled practice blocks. Our goals were to evaluate whether a minimal instructional intervention would lead learners to alter their self-regulated learning strategies to practice information to a criterion of three recalls before dropping it from learning (Experiment 1) and whether learners would spontaneously use this strategy with new material following a one week delay (Experiment 2). Concerning the first goal, it is possible that learners would not implement this specific repeated retrieval strategy effectively. Accurate metacognitive knowledge about strategy effectiveness does not always translate into effective strategy implementation. Even under ideal circumstances, learners sometimes fail to implement appropriate strategies when they have accurate strategy knowledge (Borkowski et al., 1987).

A repeated retrieval strategy requires learners to monitor the accuracy of their responses and keep track of the number of times they have recalled each item. Difficulties with either task could lead to failures to implement a repeated retrieval strategy. In the current experiments, we indirectly measured monitoring behavior by allowing people to make decisions about whether to receive correct answer feedback following self-testing trials. If learners are attempting to monitor the accuracy of their responses, they should choose to view the correct answer following every test trial. Recent evidence suggest that people are likely to seek feedback following retrieval practice (Dunlosky & Rawson, 2015). However, even if people check for feedback, they might still fail to use a repeated retrieval strategy due to difficulties in keeping track of the number of times they have successfully recalled items or due to overconfidence in their ability to remember translations after only a few correct attempts. People seem to be quite good at remembering whether a given item was recalled on a previous test (Ariel & Dunlosky, 2011; Finn & Metcalfe, 2008; Gardiner & Klee, 1976; Klee & Gardiner, 1976; King, Zechmeister, & Shaughnessy, 1980; Serra & Ariel, 2014), but no research exists examining whether people can keep track of the number of correct retrieval attempts for individual items across multiple tests.

Assuming that learners are capable of implementing a repeated retrieval practice strategy effectively, they might still be reluctant to do so because the experience of repeatedly retrieving information is subjectively more difficult than repeatedly restudying information (Karpicke, 2009; Kornell & Son, 2009; Roediger & Karpicke, 2006; Tullis et al., 2013). This could cause students to discount our strategy instructions because the instructions contradict students' typical metamnemonic experiences. Consistent with this hypothesis, direct instruction is often not effective for correcting inaccurate metacognitive beliefs when beliefs are derived from perceived fluency or disfluency during learning (Koriat & Bjork, 2006; Yan, Bjork, & Bjork, 2016). Thus, it is unclear whether repeated retrieval practice instructions will have any effect on learners' decisions to use retrieval practice. If learners are reluctant to use repeated retrieval practice or are incapable of implementing this strategy effectively, it would undermine the applied implications of this potent learning strategy.

Experiment 1

Method

Subjects. Sixty Purdue University undergraduates participated in exchange for course credit. They were randomly assigned to a neutral instruction control group ($n = 30$) or the retrieval practice instructions group ($n = 30$). We selected this sample size because an a priori power analysis conducted using G*Power software (Faul, Erdfelder, Lang, & Buchner, 2007) indicated that 60 subjects were required to achieve sufficient power (.95) to detect group differences in final recall performance based on the effect size ($d = .87$) observed from a previous experiment examining computer scheduled repeated retrieval practice (Experiment 1b from Grimaldi & Karpicke, 2014).

Materials. Twenty normatively difficult Lithuanian–English translations (e.g., *voras–spider*) were selected from Grimaldi, Pyc, and Rawson (2010).

Procedure. Subjects were tested in groups of 4 to 12 students. The experiment was administered on a computer using a custom program. Each subject was assigned to a single computer, and computers were separated by partitions so that subjects could not see other subjects or computers during the task. Subjects were told that they would learn 20 Lithuanian–English translations and would take a final test approximately 45 min after the start of the experiment. They were told that on the final test they would be asked to recall the English translations for each Lithuanian word. They were also told that their goal should be to learn all of the translations, so that they could recall as many as possible on the final test, and that they would be given control over how they practiced the translations in the initial learning phase.

Before beginning the learning phase, subjects in the control and retrieval practice instructions groups were given different instructions. Subjects in the control group were simply told to learn the translations so that they could remember as many as possible on the final test. Subjects in the retrieval practice instructions group were told that repeatedly self-testing was an effective strategy for improving learning. They were told that research indicates that people learn more from repeated testing than from repeatedly studying. They were also shown a bar graph (Figure 3a from Grimaldi & Karpicke, 2014) and told that the graph contained data from Purdue students who repeatedly studied information or repeatedly retrieved information with practice tests. Finally, they were told that the best strategy to ensure that they remembered all the translations on the final test in 45 min was to successfully retrieve each translation at least three times in the learning phase. The complete retrieval practice instructions are presented in the Appendix.

The learning phase consisted of alternating *choice blocks* and *practice blocks*. During choice blocks, subjects planned how they would practice each translation in the next practice block. Figure 1 shows the computer display during a choice block. Subjects saw a stack of 20 cards on the left side of the screen, and they made their choices by clicking on cards, dragging them, and dropping them into one of three piles on the right side of the screen. When subjects clicked down on a card, a translation appeared on it for the duration the mouse button was pressed down on it; otherwise, the cards were blank to minimize study during the choice blocks. Because we were most interested in retrieval practice decisions, both the Lithuanian and English translation were presented when subjects clicked down on a card to discourage covert retrieval during the choice phase. This procedure allowed us to ensure that all decisions to use retrieval practice were recorded.

The subjects were told that if they chose to study an item, they would be shown the translation (the Lithuanian and English words) in a study trial in the next practice block. If they chose to test themselves on an item, they would see the Lithuanian word and would be asked to recall the English translation in a test trial in the next practice block. If they chose to drop an item, they would no longer practice that translation in the learning phase. Subjects chose to study, test themselves, or drop items from further practice by dragging and dropping the cards into one of three piles labeled "Study Pile," "Practice Test Pile," and "Done Pile," respectively.

After making all of their choices during a choice block, subjects clicked a start button to begin a practice block. If subjects chose only to study items and did not choose to test on any items, the practice block contained only study trials. Likewise, if subjects chose only to test on items and did not choose to study any items, the practice block contained only test trials. If subjects chose to study some items and test other items, then before beginning the practice block, subjects indicated whether they wanted to complete the study trials or practice test trials first by clicking one of two buttons on the screen to indicate their choice.

A study block consisted of one or more study trials. On each study trial, subjects saw a Lithuanian word, its one-word English translation to the right of it, and a button labeled "Next" below the word pair. Study trials were self-paced, and subjects clicked the button to advance to the next trial. The order of trials was randomized within each study block. Each test block consisted of one or more test trials presented in a random order. On each test trial, subjects saw a Lithuanian word and a text entry box below it, and they were told to recall and type the English translation for that word. Test trials were self-paced, and subjects pressed the Enter key to enter each response. After entering a response, subjects indicated whether they wanted to see the correct answer for that translation by clicking a button labeled "Yes" or a button labeled "No." If subjects chose yes, the correct response was shown for 2

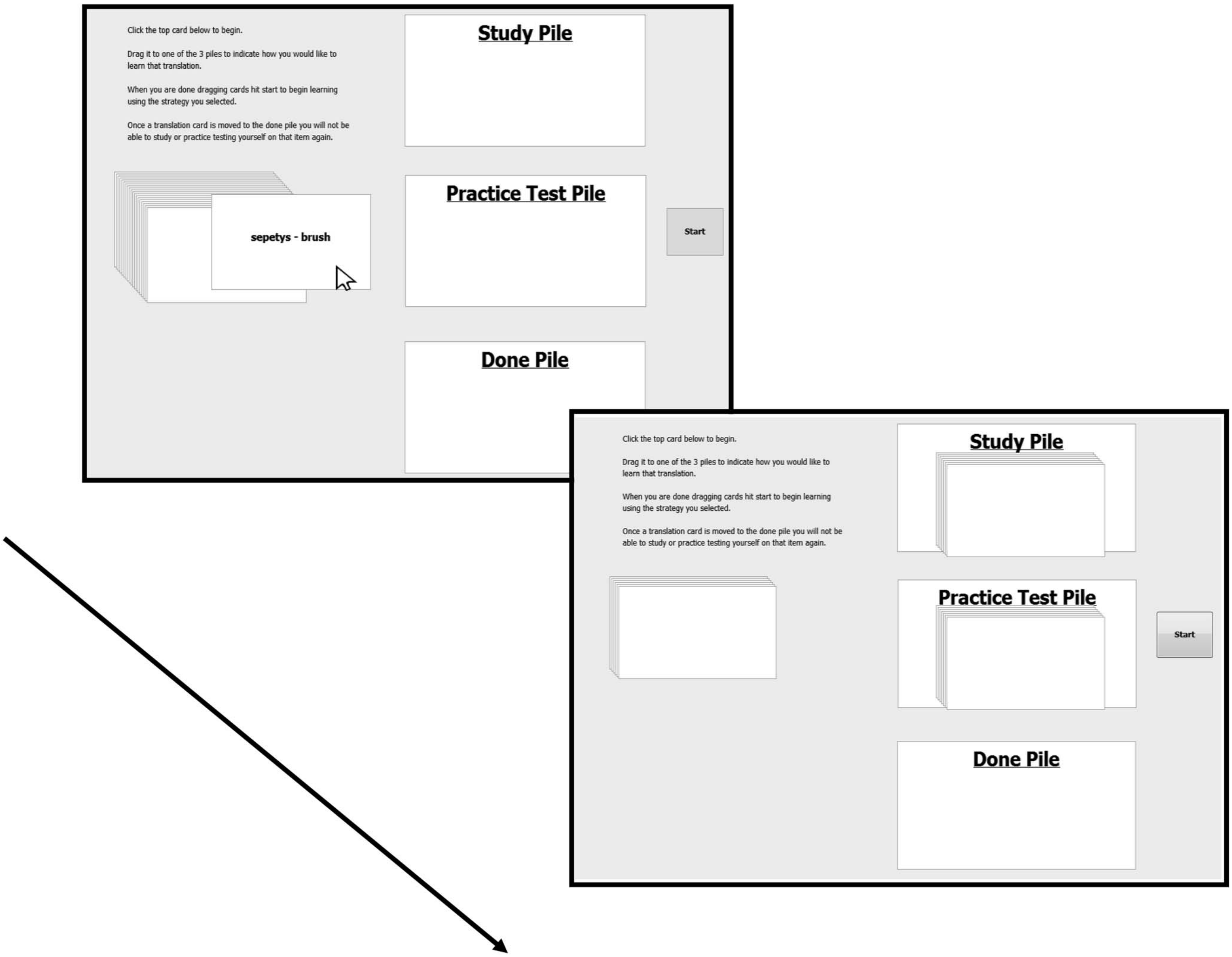

Figure 1. Display of a choice block during the learning phase in Experiments 1 and 2. Students clicked and dragged cards from the pile on the left side of the display to one of the piles on the right side of the display. Their choices determined how those items would be practiced during the subsequent practice block. The English–Lithuanian translation on a given card was only visible while the computer mouse was clicked down on that card.

s, after which the computer program advanced to the next test trial. If subjects chose no, the computer program advanced to the next test trial without displaying the correct response.

When subjects finished a practice block, a new choice block began. During this and subsequent choice blocks, only items that were not dropped in the previous choice block were presented in the choice pile. These items were presented in the same order as the previous choice block. The experiment continued with alternating choice and practice blocks until the subject moved all cards to the *drop* pile during a choice phase.

When subjects completed the learning phase by moving all items to the drop pile, they then completed a distracter task, which was unrelated to the learning task, for 15 min. After completing the distracter task, subjects took a final cued-recall test over all translations. The final test procedure was the same as the test procedure used in the learning phase with the exception that participants were not given the option to view feedback for their responses. At the end of the experiment, the subjects were debriefed and thanked for their participation.

Results and Discussion

Final recall. The retrieval practice instructions group recalled more translations (M = .87, SE = .03) than did the control group (M = .64, SE = .06) on the delayed criterial test, $t(58) = 3.36$, $p <$ .001, d = 0.88, confidence interval (CI) [.33, 1.39]. Thus, retrieval practice instructions enhanced performance suggesting that students altered their self-regulated learning choices based on the instructions. Subsequent analyses below focus on the effects of retrieval practice instructions on self-regulated learning choices.

Self-regulated learning behavior.

Study, test, and feedback choices. The next analyses focused on three self-regulated learning measures: (a) the frequency of study decisions per item, (b) the frequency of self-test decisions per item, and (c) the proportion of feedback choices following self-testing trials. The means for study choices, test choices, and feedback decisions are presented in Table 1 as a function of instruction group.

Both groups chose to study translations at least once, on average, and did not differ in the average number of study choices for translations, $t(58) = 0.70$, $p = .49$, $d = 0.18$ [−0.34, 0.69]. Consistent with the goals of the strategy intervention, the retrieval practice instructions group chose to self-test more than did the control group, $t(58) = 3.23$, $p < .01$, $d = 0.83$ [0.30, 1.36]. Both groups also chose to receive correct answer feedback following the majority of self-testing trials (~96% of trials), presumably to monitor the accuracy of their responses. The proportion of feedback choices did not differ much across groups, $t(58) = 1.67$, $p = .10$, $d = 0.43$ [−0.08, 0.94], but this outcome should be interpreted cautiously because feedback choices for both groups were at ceiling.

Cumulative learning. Table 2 shows both cumulative recall and cumulative attempted recall for each group during the learning phase as measured in each test practice block. Cumulative recall is the proportion of items recalled at least once and attempted recall is the proportion of unique items self-tested (see Karpicke, 2009; Karpicke & Roediger, 2007). Table 2 reveals that retrieval practice instructions increased both cumulative recall and attempted recall during learning. The retrieval practice instructions group recalled 90% of items at least once during learning compared to 70% in the control group, $t(58) = 2.73$, $p < .01$, $d = 0.71$ [0.18, 1.22]. Moreover, the retrieval practice instruction group used a retrieval practice strategy at least once for nearly all items (99%) compared to the control group who chose not to use retrieval practice for 12% of items, $t(58) = 2.51$, $p < .05$, $d = 0.65$ [0.13, 1.17].

Number of successful retrieval attempts during learning. The number of correct retrieval practice trials for each item during the learning phase is presented in the left panel of Figure 2. This figure illustrates that the retrieval practice instruction group recalled items to a higher accuracy criterion before dropping them from learning than did the control group, $t(58) = 3.52$, $p < .01$, $d = 0.92$ [0.37, 1.44]. The retrieval practice instruction group recalled translations approximately three times before dropping them from the learning task, whereas the control group dropped translations on average after one correct retrieval practice trial. Consistent with this observation, the mean number of successful recall attempts per item differed significantly from 3 for the control group, $t(29) = 8.09$, $p < .001$, $d = 1.48$, but did not differ from 3 for the retrieval practice instructions group, $t(29) = 0.55$, $p = .58$, $d = 0.10$.

Given that cumulative learning was less than 100% for both groups, learners obviously did not use the same criterion for all items before dropping them from learning. The right panel of Figure 2 shows the mean proportion of items reaching four possible criteria before they were dropped from study (0, 1, 2, or 3 or more correct retrieval practice trials). Figure 2 reveals that the minimal retrieval practice instructions used in our intervention decreased the number of items that were never recalled, $t(58) = 2.07$, $p < .05$, $d = 0.52$ [0.02, 1.05], or recalled only once, $t(58) = 2.15$, $p < .05$, $d = 0.56$ [0.04, 1.07], while increasing the number of items that were recalled 3 or more times during practice, $t(58) = 3.49$, $p < .01$, $d = 0.91$ [0.37, 1.43]. Groups did not differ in the proportion of items that they recalled 2 times during practice, $t(58) = 0.25$, $p = .81$, $d = 0.06$ [−0.44, 0.57]. These results indicate that our intervention was effective in promoting use of a repeated retrieval practice strategy during learning but also suggest

Table 1
Mean Number of Study Choices, Number of Test, and Proportion of Feedback Choices in Experiment 1 and During the Initial Intervention Session and Transfer Session in Experiment 2

Variable	Control group	Retrieval practice instructions
Experiment 1		
Study choice frequency	1.49 (0.16)	1.67 (0.20)
Test choice frequency	2.49 (0.28)	4.07 (0.40)
Proportion feedback choices	0.96 (0.02)	0.99 (0.01)
Experiment 2—Initial session		
Study choice frequency	2.81 (0.28)	1.77 (0.20)
Test choice frequency	3.32 (0.35)	4.32 (0.33)
Proportion feedback Cchoices	0.96 (0.02)	0.96 (0.02)
Experiment 2—Transfer session		
Study choice frequency	2.31 (0.41)	1.83 (0.40)
Test choice frequency	2.19 (0.27)	3.67 (0.38)
Proportion feedback choices	0.90 (0.04)	0.94 (0.14)

Note. Standard error of the means are in parentheses.

Table 2
Mean Cumulative Proportion of Items Recalled and Attempted for Recall During Learning as a Function of Chosen Test Block for Each Group in Experiment 1 and During the Initial Intervention Session and Transfer Session in Experiment 2

Variable	Test block 1	2	3	4	5+
Experiment 1					
Control					
Recalled	.17 (.03)	.38 (.05)	.52 (.06)	.60 (.06)	.70 (.06)
Attempted	.54 (.07)	.72 (.05)	.84 (.04)	.86 (.05)	.88 (.05)
Retrieval practice instructions					
Recalled	.21 (.03)	.45 (.04)	.66 (.04)	.77 (.04)	.90 (.03)
Attempted	.63 (06)	.83 (.04)	.91 (.03)	.95 (.02)	.99 (.01)
Experiment 2—Initial session					
Control					
Recalled	.15 (.03)	.42 (.06)	.56 (.06)	.64 (.06)	.81 (.06)
Attempted	.53 (.07)	.78 (.05)	.90 (.03)	.92 (.03)	.98 (.01)
Retrieval practice instructions					
Recalled	.22 (.04)	.42 (.06)	.56 (.06)	.64 (.07)	.73 (.06)
Attempted	.62 (.07)	.78 (.05)	.85 (.05)	.88 (.05)	.92 (.04)
Experiment 2—Transfer session					
Control					
Recalled	.17 (.03)	.43 (.04)	.62 (.05)	.74 (.04)	.94 (.03)
Attempted	.53 (06)	.81 (.04)	.91 (.03)	.94 (.03)	.98 (.01)
Retrieval practice instructions					
Recalled	.19 (.03)	.42 (.05)	.61 (.05)	.69 (.06)	.92 (.04)
Attempted	.57 (07)	.73 (.06)	.80 (.05)	.82 (.05)	.97 (.03)

Note. Standard error of the means are in parentheses.

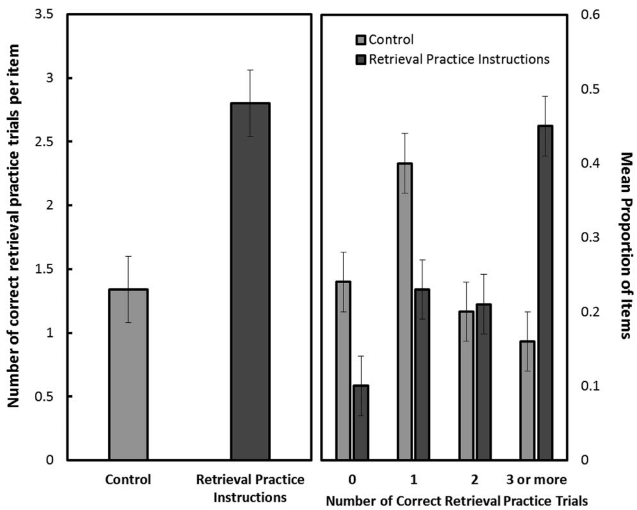

Figure 2. Mean number of correct retrieval practice trials per item during the learning phase (left panel) and mean proportion of items correctly recalled 0, 1, 2, or 3+ times (right panel) for each group in Experiment 1. Errors bars represent standard error of the means.

that students may need more formal training or environmental support to recall all items to the same accuracy criterion.

Decisions before successful recall. The next analysis focused on learners' decisions to study, test, or drop items *before* they had successfully recalled the items. Because the goal of our intervention was to promote *repeated* retrieval practice after each item had been recalled once, one might expect students' self-regulated learning decisions to be similar in the two instruction conditions before items were recalled and to differ only after items were recalled. The mean proportion of study, test, and drop decisions for items before their initial successful retrieval are presented in the left panel of Figure 3 collapsed across the first 5 choice blocks. Groups did not differ in the proportion of study choices, $t(58) = 0.36$, $p = .73$, $d = 0.09$ [−0.41, 0.60], or test choices, $t(58) = 1.34$, $p = .19$, $d = 0.35$ [−0.17, 0.85], for items before initial recall. However, students in the control group were more likely to drop items before successful retrieval than were students in the retrieval practice instructions group, $t(58) = 2.41$, $p < .05$, $d = 0.62$ [0.10, 1.14]. This finding is consistent with the cumulative learning data above which indicated that instructions to repeatedly self-test increased the number of items that were learned during practice.

Decisions after successful recall. The right panel of Figure 3 presents study, test and drop decisions for items after their initial successful recall. After an initial successful recall for an item, subjects in the retrieval practice instructions group were more likely to continue testing themselves than were subjects in the control group, $t(55) = 3.52$, $p < .001$, $d = 0.93$ [0.38, 1.48]. In contrast, the control group had a greater preference for either restudying items, $t(55) = 2.02$, $p < .05$, $d = .53$ [0.01, 1.06], or dropping them from study compared to the retrieval practice instructions group, $t(55) = 2.24$, $p < .05$, $d = .59$ [0.06, 1.12]. Overall, the strategy preferred most by the retrieval practice instructions group was to engage in repeated retrieval practice. In contrast, the control group was equally likely to drop items from learning as they were to engage in repeated retrieval practice.

Total time allocated to learning. Next, we examined the total time allocated to learning which was computed as the total duration of the learning phase of the experiment beginning at the start of the first choice block and ending after a student chose to drop the last translation from learning. The retrieval practice instructions group ($M = 21.89$, $SE = 1.61$) allocated more time (in minutes) to learning translations than did the control group ($M = 15.55$, $SE = 1.31$), $t(58) = 3.06$, $p < .01$, $d = 0.79$, CI [.26, 1.31]. These group differences in time allocation are not surprising because the goal of strategy intervention was to encourage students to continue using retrieval practice for items that they would otherwise drop from learning. Thus, increasing the number of retrieval practice trials necessitates that learner's also increase learning time. The total time allocated to learning was highly correlated with the number of successful recall trials per item, $r = .80$, $p < .001$ which suggests that students who allocated more time to learning used their time to implement a highly effective learning strategy.

Effects of repeated successful retrieval on final performance. Finally, Table 3 shows final recall performance as a function of the number of correct retrieval trials during learning, as well as the proportion of items that fall into each category. Table 3 shows that successful repeated retrieval during study was associated with improved learning performance in both groups. To estimate the effects of repeated retrieval success on final recall performance we entered the data into a multilevel fixed effects logistical regression model (for rationale, see Jaeger, 2008). This allowed us to account for group and individual differences in our analysis of the frequency of successful retrieval practice for items evident in Table 3. Fixed effects for the number of successful recall attempts for an item were entered into the model (group mean centered) with both random-subject level and random-group level (retrieval practice instructions) intercepts (i.e., subjects nested in groups). Fixed effects for the model are summarized in Table 4. The regression coefficients for the fixed effects in Table 4 reflect the log odds of successful final recall with every 1 unit increase in each variable and can be interpreted in terms of probabil-

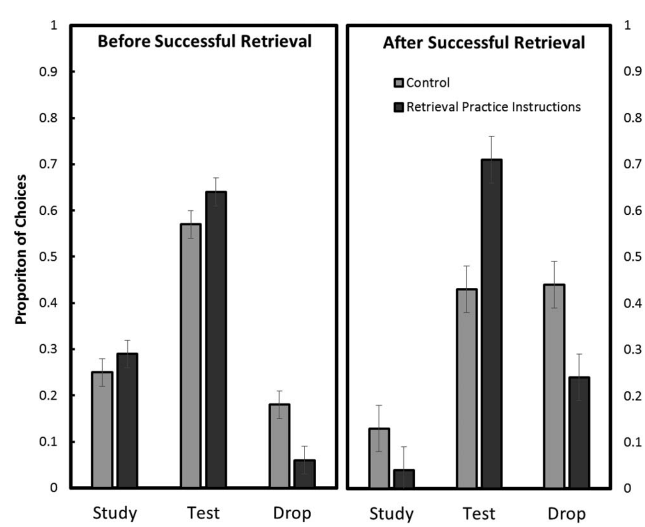

Figure 3. Mean proportion of study, test, and drop decisions for items before (left panel) and after (right panel) the first successful retrieval practice trial for them in Experiment 1. Errors bars represent standard error of the means.

Table 3
Proportion of Items Recalled 0, 1, 2, or 3+ Times During Learning and Final Recall Performance in Experiment 1 and During the Initial Intervention Session and the Transfer Session in Experiment 2

	Control group		Retrieval practice instructions	
Number of initial recalls	Proportion of items	Final recall	Proportion of items	Final recall
Experiment 1				
0	.24 (.06)	.31 (.08)	.10 (.04)	.60 (.10)
1	.40 (.06)	.68 (.06)	.24 (.05)	.86 (.05)
2	.20 (.04)	.84 (.06)	.21 (.04)	.93 (.03)
3+	.16 (.04)	.98 (.02)	.45 (.07)	.86 (.05)
Experiment 2—Initial session				
0	.19 (.06)	.35 (.10)	.06 (.03)	.48 (.14)
1	.26 (.05)	.71 (.07)	.17 (.04)	.80 (.06)
2	.17 (.03)	.86 (.06)	.23 (.03)	.91 (.03)
3+	.38 (.07)	.93 (.04)	.53 (.07)	.95 (.02)
Experiment 2—Transfer session				
0	.31 (.06)	.50 (.08)	.12 (.05)	.68 (.11)
1	.28 (.05)	.76 (.07)	.16 (.04)	.90 (.05)
2	.18 (.03)	.96 (.02)	.23 (.03)	.90 (.04)
3+	.23 (.06)	.91 (.06)	.50 (.07)	.93 (.03)

Note. Standard errors of the means are in parentheses.

ity. Thus, Table 4 shows that the probability of correctly remembering a translation on the final test increased by .67 with each successful retrieval attempt ($\beta = .72$). In sum, repeated successful retrieval practice was highly effective for improving subsequent retention when students chose to engage in it.

Experiment 2

Experiment 1 indicated that learners were capable of implementing a retrieval practice strategy when given minimal strategy instructions. One implication of this finding is that it may be possible for educators to help students improve their self-regulated learning by drawing attention to the memorial benefits of repeated retrieval practice and providing them with a few guidelines about how to implement this strategy to promote durable learning. Although these results are promising, it is unclear whether this simple intervention would have long-term effects on students' self-regulated learning strategies. A more powerful demonstration of the efficacy of the current intervention would show preserved strategy use after a delay with no additional instructions. Will students internalize the knowledge that retrieval practice is an effective learning strategy and generalize it to a new learning context after a delay?

Einstein, Mullet, and Harrison (2012) provided promising but indirect evidence that students may be capable of generalizing use of a retrieval practice strategy after experiencing its benefits. They had students in a psychology lab course read text passages and then either reread some passages or practice retrieval of them. One week later, students were administered a surprise quiz. Students were later given summaries of class performance for each passage and instructed to graph the data and speculate about the goals of the experiment. Students performed better on the quiz when they used retrieval practice compared to restudy. Students also correctly speculated that this improved performance was caused by using the retrieval practice strategy. At the end of the semester, students were asked to report how often they used retrieval practice to learn course content relative to the beginning of the semester. Eighty-two percent of students reported that they were much more likely to use retrieval practice during study than at the beginning of the semester.

The results of Einstein et al. (2012) indicate that guided exposure to retrieval practice benefits may alter future self-regulated study behavior. However, this conclusion is based on uncorroborated self-reported strategy use, and hence, it is unclear whether students actually used retrieval practice to learn course content and how much they used it. To examine the efficacy of the current strategy intervention, which is based on direct instruction about strategy effectiveness, students in Experiment 2 completed a second learning Session 1 week after the initial learning session, which was identical to the one used in Experiment 1. During this second transfer session, students studied a new set of translations and were not given any strategy instructions or reminded about the effectiveness of retrieval practice as a learning strategy. This allowed us to examine whether students who received the initial intervention would spontaneously adopt a repeated retrieval strategy in a new learning context following a 1-week delay. Put in terms of Barnett and Ceci's (2002) taxonomy of transfer, our focus concerned whether learners would transfer strategy use to a different knowledge domain (a new set of materials) and a new temporal context. If students do not continue to use a repeated retrieval practice strategy without explicit instructions and after a delay, then a more extensive strategy training intervention may be required than the current approach to improve students' long-term strategy use.

Method.

Subjects. Seventy-one undergraduate students from Purdue University completed the initial session. Sample size was determined using our effect size for group differences in final recall performance from Experiment 1 ($d = .88$). To achieve at least .95 power, at least 58 subjects (29 in each group) were required. We oversampled slightly to account for expected attrition and to ensure that our final sample would afford sufficient power. Sixty-four subjects returned to

Table 4
Fixed Effects for the Multilevel Logistic Regression Models Predicting Final Recall in Experiment 1 and During the Initial Intervention Session and Transfer Session in Experiment 2

Fixed effects	β	*SE*	Wald *Z*	*p*	95% CI Lower	95% CI Upper
Experiment 1						
Intercept	2.01	.70	2.89	<.01	0.65	3.38
Number of initial recalls	0.72	.10	6.99	<.001	0.52	0.92
Experiment 2—Initials						
Intercept	2.79	.36	7.83	<.001	2.09	3.48
Number of initial recalls	0.37	.11	3.55	<.001	0.17	0.60
Experiment 2—Transfer session						
Intercept	3.04	.44	6.80	<.05	2.16	3.90
Number of initial recalls	0.18	.08	2.28	<.05	0.03	0.34

Note. CI = confidence interval.

complete the transfer session of the experiment (n = 33 in the neutral instruction group, n = 31 in the retrieval practice instructions group). Only subjects who completed both sessions were included in the analyses.

Materials and procedure. The experiment consisted of two sessions that occurred one week apart. The materials and procedures in Session 1 were identical to those used in Experiment 1: Subjects completed a self-regulated learning phase and a final test over a list of 20 Lithuanian–English translations. For the transfer session (Session 2), 20 normatively difficult Swahili–English translations were selected from the norms of Nelson and Dunlosky (1994). The procedure in Session 2 was the same as the procedure in Session 1 with two exceptions: All subjects learned and were tested on Swahili–English materials in the learning phase and final test, and all subjects received neutral instructions. Thus, subjects who received retrieval practice instructions in Session 1 received neutral instructions in Session 2.

Results and Discussion

Final recall. The retrieval practice instructions group recalled a higher proportion of items than the control group during both the initial session—retrieval practice instructions: M = .89, SE = .02; control: M = .75, SE = .05), $t(62) = 2.34$, $p < .05$, $d = 0.59$ [0.08, 1.08]—and the transfer session—retrieval practice instructions: M = .88, SE = .04; control: M = .73, SE = .06), $t(62) = 2.23$, $p <$ .05, $d = 0.56$ [0.06, 1.06]. Thus, the retrieval practice intervention administered during the initial session had a lasting effect on how students learned a new list with neutral instructions 1 week later. Next, we unpack the contribution of students' self-regulated learning strategies to this improved learning.

Self-regulated learning behavior.

Study, test, and feedback choices. Mean study, test, and proportion of feedback choices for the initial session and transfer session are presented in Table 1 as a function of group. Table 1 shows that the retrieval practice instructions influenced decisions to self-test but not decisions to study items or receive feedback on testing trials. Consistent with this observation, participants chose to self-test more in the retrieval practice instructions group than in the control group for the initial session, $t(62) = 2.08$, $p <$ .05, $d = 0.52$ [.02,1.02], and transfer session, $t(62) = 3.20$, $p <$ .01, $d = 0.80$ [0.29, 1.31]. However, the frequency of study choices and feedback choices did not differ between groups for either Session 1, $t(62) = 0.90$, $p = .37$, $d = 0.23$ [−0.27, 0.72], or Session 2, $t(62) = 0.42$, $p = .68$, $d = 0.11$ [−0.39, 0.60].

Cumulative learning. Cumulative recall and cumulative attempted recall during the learning phases of both sessions are presented in Table 2. In contrast to Experiment 1, the number of unique items attempted for recall during learning was not significantly different between groups in either Session 1, $t(62) = 1.18$, $p = .24$, $d = 0.30$ [−0.20, 0.79], or Session 2, $t(62) = .86$, $p =$ .39, $d = 0.22$ [−0.28, 0.71]. Both groups attempted to recall nearly all the items at least once during learning in Session 1 (98%) and more than 90% of items in Session 2. Table 2 shows that the retrieval practice instructions group successfully recalled more unique items during learning than the control group. These group differences were only statistically significant during the transfer session, $t(62) = 2.43$, $p < .05$, $d = 0.61$, [0.10,1.11]. They were not statistically significant for Session 1, $t(62) = 1.91$, $p = .06$, $d = 0.48$ [−0.02, 0.97]. To summarize, students in both groups used a retrieval practice strategy at least once for nearly all items. Of importance, however, is whether students continued to use retrieval practice for learning translations after an initial successful recall for them. We examine this key question in the next analysis.

Number of successful retrieval attempts during learning. Figure 4 shows the mean number of correct retrieval practice trials for each item during both sessions. Consistent with the goals of the strategy intervention, the retrieval practice instructions group recalled each translation successfully more times than did the control group during both the initial session, $t(62) = 2.04$, $p < .05$, $d =$ 0.51 [0.01, 1.01], and the transfer session of the experiment, $t(62) = 3.79$, $p < .001$, $d = 0.95$ [0.43, 1.46]. Most important, the retrieval practice instructions group recalled items on average 3 times immediately after the intervention (initial session) and also one week later (transfer session), whereas the control group recalled items significantly less than 3 times in both the initial session, $t(32) = 2.93$ $p < .01$, $d = 0.51$, and the transfer session, $t(32) = 7.18$, $p < .001$, $d = 1.25$. Put differently, students who experienced the initial intervention spontaneously used a repeated retrieval strategy when studying new materials one week after the intervention occurred. Consistent with this conclusion, there was a high correlation between the number of successful retrieval practice trials during the initial session and transfer session for both groups (control: $r = .85$, $p < .001$; retrieval practice instructions group: $r = .54$, $p < .01$). A Fischer r to z test indicated that this correlation was higher for the control group than for the retrieval practice instruction group, $z = 2.48$, $p < .05$, which suggests that the amount of retrieval practice used by the neutral instructions group was more consistent across sessions.

Figure 5 shows the mean proportion of translations never recalled, recalled only one time, recalled only 2 times, and recalled 3 or more times during the learning phases of the initial session (left panel) and transfer session (right panel). Figure 5 shows that the retrieval practice instructions group was more likely to successfully recall translations 3 or more times than was the control group during the initial session, $t(62) = 2.02$, $p < .05$, $d = 0.51$ [0.01, 1.01], and the transfer session, $t(62) = 3.17$, $p < .01$, $d =$ 0.79 [0.28, 1.30], whereas the control group was more likely to never recall a translation before dropping it from the learning task during the initial session, $t(62) = 2.36$, $p < .05$, $d = 0.59$, CI [0.09, 1.09], and transfer session, $t(62) = 2.37$, $p < .05$ $d = 0.59$ [0.09, 1.09]. There were no significant differences during the initial session, $t(62) = 1.33$, $p = .19$, $d = 0.33$ [−0.16, 0.83], or transfer session, $t(62) = 1.83$, $p = .07$, $d = 0.46$ [−0.04, 0.95], for the number of items recalled only once. There was also no significant differences for items recalled twice during the initial session, $t(62) = 1.15$, $p = .73$, $d = 0.29$ [−0.21, 0.78], and transfer session, $t(62) = 0.93$, $p = .36$, $d = 0.23$ [−0.26, 0.72]. To summarize, the strategy intervention had persistent effects on decisions to use a repeated retrieval strategy one week after the intervention occurred. However, consistent with the outcomes from Experiment 1, students failed to recall all items to the criterion specified by the strategy instructions (i.e., recall every translation correctly 3 times) in both sessions of the experiment. Why students failed to use this strategy consistently for all items is unclear. We discuss this outcome in more detail in the General Discussion.

Decisions before successful recall. The mean proportions of study, test, and drop decisions for items before their initial successful

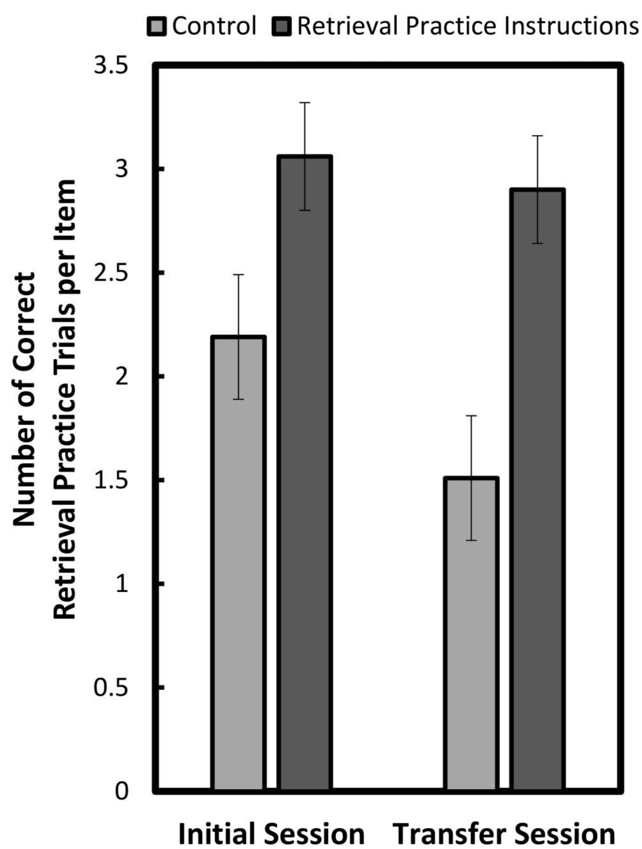

Figure 4. Mean number of correct retrieval practice trials per item in the learning phase during the initial intervention session and transfer Session 1 week later in Experiment 2. Errors bars represent standard error of the means.

retrieval are presented in Figure 6 collapsed across the first 5 choice blocks. The left panel presents data from the initial training session and the right panel shows data from the transfer session that occurred 1 week later. Figure 6 shows that the retrieval practice instructions group was more likely to choose to engage in retrieval practice than the control group during the initial session, $t(62) = 2.43$, $p < .05$, $d = 0.61$ [0.10, 1.11], and the transfer session, $t(62) = 2.10$, $p < .05$, $d = 0.53$ [0.02, 1.02]. This outcome should be interpreted cautiously because it was not obtained in Experiment 1, although the same trend was present in the Experiment 1 data (see the left panel of Figure 3). If true, this finding would suggest that an additional benefit of the current intervention is that it may promote early use of retrieval practice before material is sufficiently encoded to be recalled successfully. Because failed retrieval attempts are more effective than restudying material (Grimaldi & Karpicke, 2012; Kornell, Hays, & Bjork, 2009; Kornell, Klein, & Rawson, 2015), attempting retrieval early in the learning process may be more effective for promoting durable learning relative to students' default strategies.

Consistent with findings from Experiment 1, the control group chose to drop more items from learning before successfully retrieving them than the retrieval practice instructions group did during the initial session, $t(62) = 2.47$, $p < .05$, $d = 0.62$ [0.11, 1.12], and transfer session, $t(62) = 2.47$, $p < .05$, $d = 0.62$ [0.11, 1.12]. Groups also did not differ in their proportion of study decisions prior to a successful retrieval attempt during the initial session, $t(62) = 1.19$, $p = .24$, $d = 0.30$ [−0.20, 0.79], or transfer session, $t(62) = 0.79$, $p = .43$, $d = 0.20$ [−0.29, 0.69].

Decisions after successful recall. Figure 7 shows the mean proportion of study, test, and drop decisions for items after their first successful retrieval practice trial for the initial training session (left panel) and the transfer session (right panel). Consistent with Experiment 1, students were unlikely to choose to continue studying items after initially recalling them, and there were no differences between groups in the proportion of study choices for items in either the initial session, $t(61) = 0.15$, $p = .89$, $d = .04$ [−0.46, 0.53], or transfer session, $t(54) = 0.74$, $p = .46$, $d = 0.20$ [−0.33, 0.72]. Most important, the retrieval practice instructions group was more likely to choose to engage in repeated retrieval practice for items than the control group was during the transfer session of the experiment, $t(54) = 2.93$, $p < .01$, $d = 0.78$ [0.24, 1.33]. They were also more likely to continue using retrieval practice during the initial session, although these group differences did not reach significance, $t(61) = 1.93$, $p = .06$, $d = 0.49$ [−0.02, 0.99]. The control group was more likely to drop items from learning after successfully recalling them than the retrieval practice instructions group was during both the initial session, $t(61) = 2.14$, $p < .05$, $d = 0.54$ [0.03, 1.04], and transfer session, $t(54) = 2.42$, $p < .05$, $d = 0.65$ [0.11, 1.18]. In summary, the retrieval practice instructions altered self-regulated learning decisions for items that were successfully recalled one week later. Students who received this minimal strategy intervention were more likely to spontaneously use repeated retrieval practice and less likely to a use a one-and-done recall strategy.

Total time allocated to learning. The total time allocated to learning items did not differ for the retrieval practice instructions group ($M = 21.65$, $SE = 1.24$) and the control group ($M = 19.43$, $SE = 1.71$) during the initial session, $t(62) = 1.04$, $p = .30$, $d = 0.30$ [−.23, .75]. However, the retrieval practice instructions group ($M = 17.89$, $SE = 1.12$) allocated more time to learning items in the transfer session than did the control group ($M = 13.44$, $SE = 1.38$), $t(62) =$

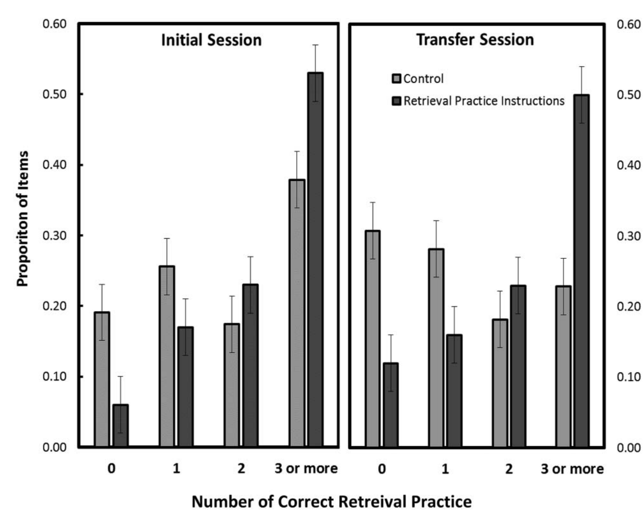

Figure 5. Mean proportion of items reaching learning accuracy criterions of 0, 1, 2, or 3+ correct retrieval practice trials for each group during the initial intervention session (left panel) and transfer Session 1 week later (right panel) in Experiment 2. Errors bars represent standard error of the means.

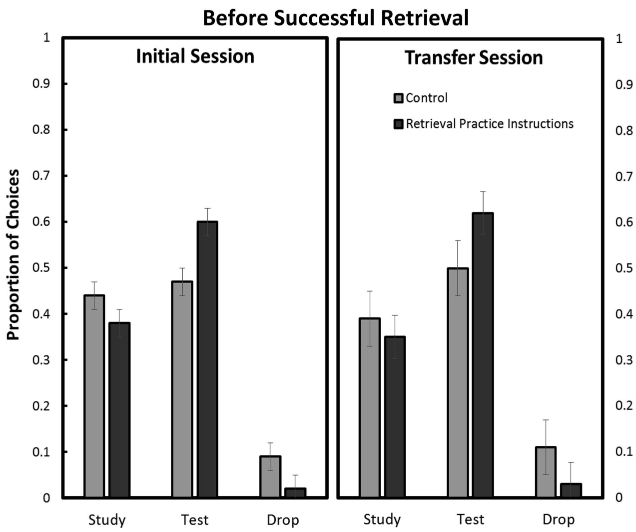

Figure 6. Mean proportion of study, test, and drop decisions for items before the first successful retrieval practice trial for them during the initial intervention session (left panel) and the transfer Session 1 week later (right panel) in Experiment 2. Errors bars represent standard error of the means.

2.91, $p < .01$, $d = 0.73$ [0.22, 1.23]. The number of successful retrieval attempts per item was highly correlated with learning time during the initial session, $r = .72$, $p < .001$ and the transfer session, $r = .68$, $p < .001$ which indicates that students who allocated more time to learning items used their time to engage in repeated retrieval practice.

Effects of repeated successful retrieval on final performance. Table 3 presents final recall performance as a function of the number of correct retrieval trials during learning and the proportion of items that fall into each category. The data in Table 3 show that items that were recalled more often during the learning phase were also more likely to be recalled on the final test. This pattern is evident in both sessions of the experiment for both groups. The effect of repeated successful retrieval on final recall performance was again examined using a multilevel fixed effects logistic regression model. Fixed effects for the number of successful recall attempts for an item were entered into the model (group mean centered) with both random-subject level and random-group level (retrieval practice instructions) intercepts (i.e., subjects nested in groups). For simplicity, we computed separate models for each session. Table 4 shows that the probability of recalling translations on the final test increased by .59 ($\beta = .37$) during the initial session and by .54 ($\beta = .18$) during the transfer session for each successful recall attempt during the learning phase. Repeated retrieval practice was a highly effective learning strategy in both sessions of the experiment.

General Discussion

The current experiments examined the efficacy of training students to use repeated retrieval practice to improve their self-regulated learning outcomes via direct instruction. Across two experiments, some participants received a minimal strategy intervention that highlighted the effectiveness of using repeated retrieval practice over the less effective strategy of repeated studying. The intervention involved telling students that repeated retrieval practice is most effective when individual to-be-learned items are recalled three or more times during learning. Students who experienced this intervention were less likely to terminate learning for items after their initial successful retrieval than were students who did not receive any strategy instructions. Perhaps most important from an applied standpoint, students who were given retrieval practice instructions in an initial session continued to use a repeated retrieval practice strategy when learning new material in a second session one week later, without receiving any strategy instructions or reminders in the second session (Experiment 2). These results suggest that students can be trained to regulate their study more effectively and also provide a few guidelines for how to do so.

Our intervention involved relatively minimal instructions that educators could adapt to a variety of learning contexts to promote more effective self-regulated learning for their students. If teachers follow our approach, they may not need to devote much lecture time to training students to use repeated retrieval strategies. Instead, they might embed retrieval practice strategy instructions into homework or study guides. Based on the present results, effective instructions should emphasize the mnemonic benefits of retrieval practice to counter students' inaccurate metacognitive beliefs and should also provide students with direct instruction about how to use a repeated retrieval practice strategy to regulate their learning (e.g., telling students they should correctly recall items at least three times before dropping them from study).

The current guidelines emphasize the role of direct instruction in teaching students to use repeated retrieval practice, but other experience-based interventions might also be effective (see Einstein et al., 2012). However, because students largely view retrieval practice as a self-assessment tool (Karpicke, Butler, & Roediger, 2009; Kornell & Son, 2009), strategy interventions that rely on experience alone may be unlikely to teach students about the effectiveness of *repeated* retrieval practice for learning. Learning about strategy effectiveness from experience requires a person

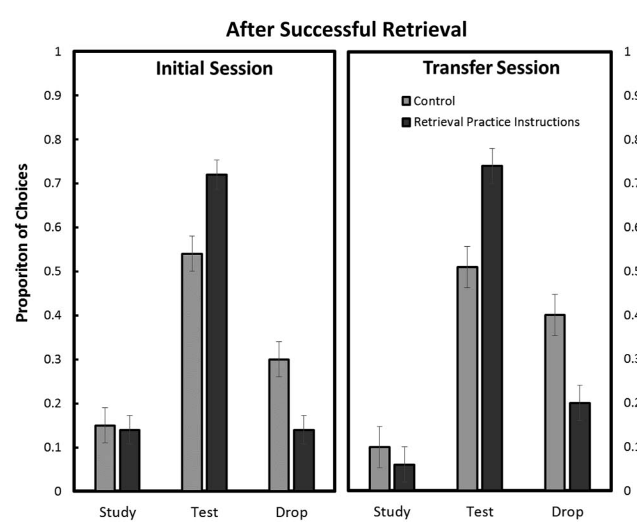

Figure 7. Mean proportion of study, test, and drop decisions for items after the first successful retrieval practice trial for them during the initial intervention session (left panel) and the transfer Session 1 week later (right panel) in Experiment 2. Errors bars represent standard error of the means.

to monitor learning gains, attribute those gains to a specific strategy, and update one's knowledge about the effectiveness of the strategy (for a formal model of strategy learning from experience, see Dunlosky & Hertzog, 2000). Therefore, because students view self-testing as an assessment tool for monitoring learning, rather than as an opportunity to engage in retrieval-based learning, they may be unlikely to attribute learning gains to the act of retrieval itself. The present results show that direct instruction can be effective for teaching students that retrieval is more than just a self-assessment tool and that repeated retrieval practice is an effective learning strategy.

Of course, there may be ways that the present intervention could be adapted and expanded to promote even better use of a repeated retrieval strategy. Our focus was on training students to recall information to a specific accuracy criterion, but other factors exists that also improve the effectiveness of retrieval practice, including the spacing of retrieval attempts for items within study sessions (Karpicke & Bauernschmidt, 2011; Karpicke & Roediger, 2007; Pyc & Rawson, 2009) and relearning with repeated retrieval to an accuracy criterion across multiple study sessions (Rawson & Dunlosky, 2013). Whether students can be trained and will adopt either strategy is an open question worth pursuing. Regardless, the current experiments provide a promising first step in developing guidelines for training students to regulate their learning more effectively using repeated retrieval practice.

Another open question is what the appropriate age is to train students to use a repeated retrieval strategy. Children as young as first graders can benefit from guided retrieval practice (Lipowski, Pyc, Dunlosky, & Rawson, 2014), but they may have difficulties implementing a repeated retrieval strategy on their own, because the strategy requires monitoring response accuracy and keeping track of the number of correct responses for each item across multiple tests. Elementary schoolchildren may lack the ability to use feedback from monitoring (e.g., their judgments of learning) when they make study decisions (Lockl & Schneider, 2004; Masur, McIntyre, & Flavell, 1973; Schneider & Lockl, 2008; Schneider & Loffler, 2016). Therefore, additional training or scaffolding may be necessary to help children use repeated retrieval practice strategies more effectively.

In some cases, even college students may require additional training to maximize the benefits of repeated retrieval practice. In the current experiments, many students failed to recall all translations to the accuracy criterion specified by the retrieval practice instructions even though on average items were recalled to this criterion (see the distributions of choices displayed in Figures 2 and 5). Failure to recall a specific item to a goal criterion could be due to errors in remembering the number of correct retrieval practice trials that have occurred for items. This type of error could be overcome by removing the memory component required when students must mentally track the frequency of correct responses for items, perhaps by instructing students to physically mark items after each correct response.

Errors in tallying could also arise from overconfidence in retrieved responses. For conceptual materials such as key-term definitions from an introductory psychology course (e.g., *What is the definition of confirmation bias?*), many students believe that their retrieved responses are completely accurate even when the responses are only partially correct and are missing important details (Dunlosky, Hartwig, Rawson, & Lipko, 2011; Dunlosky & Rawson, 2012; Grimaldi & Karpicke, 2014). This type of overconfidence could reduce the effectiveness of a repeated retrieval practice strategy because students may believe they have recalled all the important information accurately multiple times when they have not done so. Thus, additional training or interventions targeting student's monitoring accuracy might be necessary to fully reap the benefits of repeated retrieval practice with more complex material.

A final open question is whether students chose to practice repeated retrieval to different degrees with different items. The present experiments were not designed to address this question, but other research may point to possible answers. For example, students may have chosen to engage in more retrieval practice for subjectively difficult material than for subjectively easier material. Such a strategy would be relatively effective and metacognitively savvy because normatively easy items may require fewer successful retrieval practice trials than normatively difficult items to achieve equivalent benefits to retention (Vaughn, Rawson, & Pyc, 2013).

Although limited research has examined the processes influencing retrieval practice decisions, these decisions, like other self-regulatory decisions, are likely influenced by metacognitive monitoring processes. Specifically, students appear to set different decision criteria when they choose to study, retrieve, or drop information. Many students prefer to study items that they give low judgments of learning, test themselves on items they give medium level judgments of learning, and drop items they give high judgments of learning (Karpicke, 2009). A single correct retrieval attempt will typically increase one's judgment of learning for an item to levels exceeding their drop decision criterion. Our strategy intervention may have caused learners to increase their drop decision criterion to ensure that items were recalled multiple times during learning. Easy items may still be more likely to exceed this decision criterion following fewer retrieval practice trials relative to difficult items, which might explain why students in the retrieval practice instructions group did not recall all items correctly three times (see Figures 2 and 5). Because our experiments were not designed to evaluate the effects of item difficulty on retrieval practice decisions, these questions require additional future research (but see Karpicke, 2009).

In summary, the current experiments indicate that direct instructions about the effectiveness of repeated retrieval practice and how to use this strategy can alter and improve learners' self-regulated learning behavior. These results are especially encouraging because people hold inaccurate metacognitive beliefs about the efficacy of retrieval practice for improving learning outcomes (Ariel et al., 2014; Hartwig & Dunlosky, 2012; Karpicke et al., 2009; Kornell & Son, 2009) and these types of metacognitive illusions are often difficult to counteract (for recent discussion, see Yan et al., 2016). A simple instructional intervention can improve learning outcomes and promote effective self-regulated learning strategies that persist over the long term.

References

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

Alfieri, L., Brooks, P. J., Aldrich, N. J., & Tenenbaum, H. R. (2011). Does discovery-based instruction enhance learning. *Journal of Educational Psychology, 103,* 1–18. http://dx.doi.org/10.1037/a0021017

Ariel, R., & Dunlosky, J. (2011). The sensitivity of judgment-of-learning resolution to past test performance, new learning, and forgetting. *Memory & Cognition, 39,* 171–184. http://dx.doi.org/10.3758/s13421-010-0002-y

Ariel, R., Dunlosky, J., & Bailey, H. (2009). Agenda-based regulation of study-time allocation: When agendas override item-based monitoring. *Journal of Experimental Psychology: General, 138,* 432–447. http://dx.doi.org/10.1037/a0015928

Ariel, R., Hines, J. C., & Hertzog, C. (2014). Test framing generates a stability bias for predictions of learning by causing people to discount their learning beliefs. *Journal of Memory and Language, 75,* 181–198. http://dx.doi.org/10.1016/j.jml.2014.06.003

Barnett, S. M., & Ceci, S. J. (2002). When and where do we apply what we learn? A taxonomy for far transfer. *Psychological Bulletin, 128,* 612–637. http://dx.doi.org/10.1037/0033-2909.128.4.612

Borkowski, J., Carr, M., & Pressley, M. (1987). "Spontaneous" strategy use: Perspectives from metacognitive theory. *Intelligence, 11,* 61–75. http://dx.doi.org/10.1016/0160-2896(87)90027-4

Carpenter, S. K., Pashler, H., & Cepeda, N. J. (2009). Using tests to enhance 8th grade students' retention of US history facts. *Applied Cognitive Psychology, 23,* 760–771. http://dx.doi.org/10.1002/acp.1507

Dunlosky, J., & Ariel, R. (2011). Self-regulated learning and the allocation of study time. In the B. Ross (Ed.), *Psychology of learning and motivation* (Vol. 54, pp. 103–140). http://dx.doi.org/10.1016/B978-0-12-385527-5.00004-8

Dunlosky, J., Hartwig, M. K., Rawson, K. A., & Lipko, A. R. (2011). Improving college students' evaluation of text learning using idea-unit standards. *Quarterly Journal of Experimental Psychology, 64,* 467–484. http://dx.doi.org/10.1080/17470218.2010.502239

Dunlosky, J., & Hertzog, C. (1998). Training programs to improve learning in later adulthood: Helping older adults educate themselves. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Metacognition in Educational Theory and Practice* (pp. 249–275). Mahwah, NJ: Erlbaum.

Dunlosky, J., & Hertzog, C. (2000). Updating knowledge about encoding strategies: A componential analysis of learning about strategy effectiveness from task experience. *Psychology and Aging, 15,* 462–474. http://dx.doi.org/10.1037/0882-7974.15.3.462

Dunlosky, J., & Nelson, T. O. (1994). Does the sensitivity of judgments of learning (JOLs) to the effects of various study activities depend on when the JOLs occur? *Journal of Memory and Language, 33,* 545–565. http://dx.doi.org/10.1006/jmla.1994.1026

Dunlosky, J., & Rawson, K. A. (2012). Overconfidence produces underachievement: Inaccurate self evaluations undermine students' learning and retention. *Learning and Instruction, 22,* 271–280. http://dx.doi.org/10.1016/j.learninstruc.2011.08.003

Dunlosky, J., & Rawson, K. A. (2015). Do students use testing and feedback while learning? A focus on key concept definitions and learning to criterion. *Learning and Instruction, 39,* 32–44. http://dx.doi.org/10.1016/j.learninstruc.2015.05.003

Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving students' learning with effective learning techniques: Promising directions from cognitive and educational psychology. *Psychological Science in the Public Interest, 14,* 4–58. http://dx.doi.org/10.1177/1529100612453266

Einstein, G. O., Mullet, H. G., & Harrison, T. L. (2012). The testing effect: Illustrating a fundamental concept and changing study strategies. *Teaching of Psychology, 39,* 190–193.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39,* 175–191. http://dx.doi.org/10.3758/BF03193146

Finn, B., & Metcalfe, J. (2008). Judgments of learning are influenced by memory for past test. *Journal of Memory and Language, 58,* 19–34. http://dx.doi.org/10.1016/j.jml.2007.03.006

Gardiner, J. M., & Klee, H. (1976). Memory for remembered events: An assessment of output monitoring in free recall. *Journal of Verbal Learning and Verbal Behavior, 15,* 227–233. http://dx.doi.org/10.1016/0022-5371(76)90021-9

Goossens, N. A., Camp, G., Verkoeijen, P. P., & Tabbers, H. K. (2014). The effect of retrieval practice in primary school vocabulary learning. *Applied Cognitive Psychology, 28,* 135–142. http://dx.doi.org/10.1002/acp.2956

Grimaldi, P. J., & Karpicke, J. D. (2012). When and why do retrieval attempts enhance subsequent encoding? *Memory & Cognition, 40,* 505–513. http://dx.doi.org/10.3758/s13421-011-0174-0

Grimaldi, P. J., & Karpicke, J. D. (2014). Guided retrieval practice of educational materials using automated scoring. *Journal of Educational Psychology, 106,* 58–68. http://dx.doi.org/10.1037/a0033208

Grimaldi, P. J., Pyc, M. A., & Rawson, K. A. (2010). Normative multitrial recall performance, metacognitive judgments, and retrieval latencies for Lithuanian–English paired associates. *Behavior Research Methods, 42,* 634–642. http://dx.doi.org/10.3758/BRM.42.3.634

Hartwig, M. K., & Dunlosky, J. (2012). Study strategies of college students: Are self-testing and scheduling related to achievement? *Psychonomic Bulletin & Review, 19,* 126–134. http://dx.doi.org/10.3758/s13423-011-0181-y

Hopkins, R. F., Lyle, K. B., Hieb, J. L., & Ralston, P. A. (2016). Spaced retrieval practice increases college students' short- and long-term retention of mathematics knowledge. *Educational Psychology Review, 28,* 853–873.

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language, 59,* 434–446. http://dx.doi.org/10.1016/j.jml.2007.11.007

Jones, H. E. (1923). The effects of examination on the performance of learning. *Archives de Psychologie, 10,* 1–70.

Karpicke, J. D. (2009). Metacognitive control and strategy selection: Deciding to practice retrieval during learning. *Journal of Experimental Psychology: General, 138,* 469–486. http://dx.doi.org/10.1037/a0017341

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

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

Karpicke, J. D., & Bauernschmidt, A. (2011). Spaced retrieval: Absolute spacing enhances learning regardless of relative spacing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37,* 1250–1257. http://dx.doi.org/10.1037/a0023436

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

Karpicke, J. D., Blunt, J. R., Smith, M. A., & Karpicke, S. S. (2014). Retrieval-based learning: The need for guided retrieval in elementary school children. *Journal of Applied Research in Memory & Cognition, 3,* 198–206. http://dx.doi.org/10.1016/j.jarmac.2014.07.008

Karpicke, J. D., Butler, A. C., & Roediger, H. L. (2009). Metacognitive strategies in student learning: Do students practice retrieval when they study on their own? *Memory, 17,* 471–479.

Karpicke, J. D., & Roediger, H. L. (2007). Repeated retrieval during learning is the key to long-term retention. *Journal of Memory and Language, 57,* 151–162.

Karpicke, J. D., & Roediger, H. L., III. (2008). The critical importance of retrieval for learning. *Science, 319,* 966–968. http://dx.doi.org/10.1126/science.1152408

King, J. F., Zechmeister, E. B., & Shaughnessy, J. J. (1980). Judgments of knowing: The influence of retrieval practice. *The American Journal of Psychology, 93,* 329–343. http://dx.doi.org/10.2307/1422236

Klahr, D., & Nigam, M. (2004). The equivalence of learning paths in early science instruction: Effect of direct instruction and discovery learning. *Psychological Science, 15,* 661–667. http://dx.doi.org/10.1111/j.0956-7976.2004.00737.x

Klee, H., & Gardiner, J. M. (1976). Memory for remembered events: Contrasting recall and recognition. *Journal of Verbal Learning and Verbal Behavior, 15,* 471–478. http://dx.doi.org/10.1016/S0022-5371(76)90042-6

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

Koriat, A., & Bjork, R. A. (2006). Mending metacognitive illusions: A comparison of mnemonic-based and theory-based procedures. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32,* 1133–1145. http://dx.doi.org/10.1037/0278-7393.32.5.1133

Kornell, N., & Bjork, R. A. (2008). Optimizing self-regulated study: The benefits and costs of dropping flashcards. *Memory, 16,* 125–136.

Kornell, N., Hays, M. J., & Bjork, R. A. (2009). Unsuccessful retrieval attempts enhance subsequent learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35,* 989–998. http://dx.doi.org/10.1037/a0015729

Kornell, N., Klein, P. J., & Rawson, K. A. (2015). Retrieval attempts enhance learning, but retrieval success (versus failure) does not matter. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 41,* 283–294. http://dx.doi.org/10.1037/a0037850

Kornell, N., & Son, L. K. (2009). Learners' choices and beliefs about self-testing. *Memory, 17,* 493–501. http://dx.doi.org/10.1080/09658210902832915

Lipowski, S. L., Pyc, M. A., Dunlosky, J., & Rawson, K. A. (2014). Establishing and explaining the testing effect in free recall for young children. *Developmental Psychology, 50,* 994–1000. http://dx.doi.org/10.1037/a0035202

Lockl, K., & Schneider, W. (2004). The effects of incentives and instructions on children's allocation of study time. *European Journal of Developmental Psychology, 1,* 153–169. http://dx.doi.org/10.1080/17405620444000085

Masur, E. F., McIntyre, C. W., & Flavell, J. H. (1973). Developmental changes in apportionment of study time among items in a multitrial free recall task. *Journal of Experimental Child Psychology, 15,* 237–246. http://dx.doi.org/10.1016/0022-0965(73)90145-8

McDaniel, M. A., Agarwal, P. K., Huelser, B. J., McDermott, K. B., & Roediger, H. L., III. (2011). Test-enhanced learning in a middle school science classroom: The effects of quiz frequency and placement. *Journal of Educational Psychology, 103,* 399–414. http://dx.doi.org/10.1037/a0021782

McDaniel, M. A., & Schlager, M. S. (1990). Discovery learning and transfer of problem-solving skills. *Cognition and Instruction, 7,* 129–159. http://dx.doi.org/10.1207/s1532690xci0702_3

Metcalfe, J. (2002). Is study time allocated selectively to a region of proximal learning? *Journal of Experimental Psychology: General, 131,* 349–363. http://dx.doi.org/10.1037/0096-3445.131.3.349

Metcalfe, J., & Kornell, N. (2005). A region of proximal learning model of study time allocation. *Journal of Memory and Language, 52,* 463–477. http://dx.doi.org/10.1016/j.jml.2004.12.001

Morehead, K., Rhodes, M. G., & DeLozier, S. (2016). Instructor and student knowledge of study strategies. *Memory, 24,* 257–271. http://dx.doi.org/10.1080/09658211.2014.1001992

Nelson, T. O., & Dunlosky, J. (1994). Norms of paired-associate recall during multitrial learning of Swahili–English translation equivalents. *Memory, 2,* 325–335. http://dx.doi.org/10.1080/09658219408258951

O'Sullivan, J. T., & Pressley, M. (1984). Completeness of instruction and strategy transfer. *Journal of Experimental Child Psychology, 38,* 275–288. http://dx.doi.org/10.1016/0022-0965(84)90126-7

Pyc, M. A., & Rawson, K. A. (2009). Testing the retrieval effort hypothesis: Does greater difficulty correctly recalling information lead to higher levels of memory? *Journal of Memory and Language, 60,* 437–447. http://dx.doi.org/10.1016/j.jml.2009.01.004

Rawson, K. A., & Dunlosky, J. (2011). Optimizing schedules of retrieval practice for durable and efficient learning: How much is enough? *Journal of Experimental Psychology: General, 140,* 283–302. http://dx.doi.org/10.1037/a0023956

Rawson, K. A., & Dunlosky, J. (2013). Relearning attenuates the benefits and costs of spacing. *Journal of Experimental Psychology: General, 142,* 1113–1129.

Roediger, H. L., & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science, 17,* 249–255.

Schneider, W., & Lockl, K. (2008). Procedural metacognition in children: Evidence for developmental trends. In J. Dunlosky & R. A. Bjork (Eds.), *Handbook of memory and metamemory* (pp. 391–409). New York, NY: Psychology Press.

Schneider, W., & Loffler, E. (2016). The development of metacognitive knowledge in children and adolescents. In J. Dulosky & S. K. Tauber (Eds.), *Oxford handbook of metamemory* (pp. 491–518). New York, NY: Oxford University Press.

Serra, M. J., & Ariel, R. (2014). People use the memory for past-test heuristic as an explicit cue for judgments of learning. *Memory & Cognition, 42,* 1260–1272.

Son, L. K., & Metcalfe, J. (2005). Judgments of learning: Evidence for a two-stage process. *Memory & Cognition, 33,* 1116–1129. http://dx.doi.org/10.3758/BF03193217

Thiede, K. W., & Dunlosky, J. (1999). Toward a general model of self-paced study: An analysis of selection of items for study and self-paced study time. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25,* 1024–1037. http://dx.doi.org/10.1037/0278-7393.25.4.1024

Tullis, J. G., Finley, J. R., & Benjamin, A. S. (2013). Metacognition of the testing effect: Guiding learners to predict the benefits of retrieval. *Memory & Cognition, 41,* 429–442. http://dx.doi.org/10.3758/s13421-012-0274-5

Vaughn, K. E., & Rawson, K. A. (2011). Diagnosing criterion level effects on memory: What aspects of memory are enhanced by repeated retrieval? *Psychological Science, 22,* 1127–1131.

Vaughn, K. E., Rawson, K. A., & Pyc, M. A. (2013). Repeated retrieval practice and item difficulty: Does criterion learning eliminate item difficulty effects? *Psychonomic Bulletin & Review, 20,* 1239–1245. http://dx.doi.org/10.3758/s13423-013-0434-z

Yan, V. X., Bjork, E. L., & Bjork, R. A. (2016). On the difficulty of mending metacognitive illusions: A priori theories, fluency effects, and misattributions of the interleaving benefit. *Journal of Experimental Psychology: General, 145,* 918–933. http://dx.doi.org/10.1037/xge0000177

(*Appendix follows*)

Appendix

Retrieval Practice Instructions

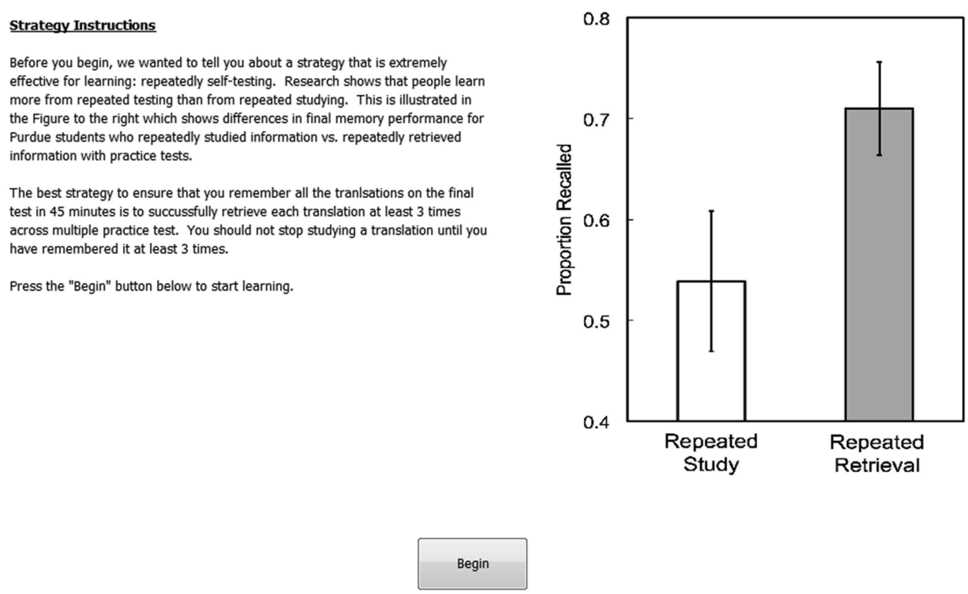

Figure A1. Illustration of the strategy instructions that the retrieval practice instructions group received prior to beginning the initial learning phase of the experiment in Experiment 1 and the initial session of Experiment 2. The retrieval practice instructions group did not receive strategy instructions during the transfer session of Experiment 2.

Received January 6, 2017
Revision received April 17, 2017
Accepted April 28, 2017 ■