Journal of Applied Research in Memory and Cognition 3 (2014) 183–188

Contents lists available at ScienceDirect

Journal of Applied Research in Memory and Cognition

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

Is the benefit of retrieval practice modulated by motivation?

Sean H.K. Kang[a,*], Harold Pashler[b]

[a] Dartmouth College, Department of Education, Hanover, NH 03755, USA
[b] University of California, San Diego, Department of Psychology, La Jolla, CA 92093-0109, USA

ARTICLE INFO

Article history:
Received 18 February 2014
Received in revised form 15 May 2014
Accepted 24 May 2014
Available online 10 June 2014

Keywords:
Retrieval practice
Testing effects
Motivation

ABSTRACT

Retrieval practice tends to produce better long-term learning than rereading, but laboratory studies have typically used arbitrary material that subjects may not care to learn. The observed advantage of retrieval practice may be exaggerated because low motivation may result in deficient processing during (usually passive) rereading. Thus, when subjects are motivated to learn the material, the type of study strategy (whether retrieval practice or rereading) might be less important. To test this hypothesis, we conducted 3 experiments in which we manipulated the incentives (using monetary bonuses or time savings) for learning Swahili–English word pairs. Items that had undergone retrieval practice were better recalled than reread items on a final test 2 days later, but this effect did not interact with incentive level. These results provide some reassurance that lab findings from the testing effects literature likely generalize to real-world situations in which motivation to learn may be greater.

1. Introduction

Laboratory research on human memory has long shown that a memory test is not a neutral event that merely measures the contents of memory (e.g., Abbott, 1909; Lachman & Laughery, 1968; Tulving, 1967). On the contrary, taking a test typically enhances long-term learning and retention more than rereading of the target information (e.g., Carrier & Pashler, 1992; McDaniel, Wildman, & Anderson, 2012; Rohrer, Taylor, & Sholar, 2010), and processes involved in memory retrieval have been implicated (e.g., Carpenter, 2009; Kang, McDermott, & Roediger, 2007; Karpicke & Zaromb, 2010). This testing effect is often referred to in the research literature as the benefit of *retrieval practice* (for recent reviews, see Carpenter, 2012; Rawson & Dunlosky, 2012). Given that retrieval practice would appear to be a useful instructional tool, in recent years there have been calls for greater application of retrieval practice in the classroom (e.g., Dempster, 1992; Karpicke & Grimaldi, 2012; Roediger, Agarwal, Kang, & Marsh, 2010).

Literally hundreds of experiments have found retrieval practice beneficial for learning—across a wide range of study materials and diverse student populations. It is worth noting, however, that the overwhelming majority of these studies were conducted in the lab, and one can reasonably wonder whether lab findings will generalize to real-world classrooms (e.g., Efklides, 2012; Lundeberg & Fox, 1991). One difference is that in lab studies, the material is usually arbitrary from the subject's point of view, i.e., it is assigned by the experimenter with little regard for the subject's actual interests or goals (e.g., Swahili foreign vocabulary). By contrast, when students study in real life it is typically in pursuit of some larger goal, perhaps to master some content area that is of interest to the student or to do well on an upcoming exam. Thus, it seems likely that the motivation to learn the material for a typical subject in a lab experiment is substantially lower than for students in most real-world situations.

1.1. Motivation and learning

Motivation refers to the condition that initiates and/or maintains a person's goal-directed behavior. It is generally assumed that there are powerful links between motivation, learning, and academic achievement (e.g., Deci, Vallerand, Pelletier, & Ryan, 1991; Dweck, 1986; Lepper, Greene, & Nisbett, 1973). Motivation is generally thought to facilitate learning through several means, such as increasing the attention the individual pays to the materials (as compared to competing stimuli in the environment) and by promoting the adoption of effortful encoding strategies.

How does motivation bear on the enhancement of learning through retrieval practice? It is conceivable that a relatively passive study strategy like rereading (the usual control against which

* Corresponding author. Tel.: +1 603 646 9051; fax: +1 603 646 3968.
E-mail addresses: Sean.H.Kang@dartmouth.edu, seankang@wustl.edu (S.H.K. Kang), hpashler@ucsd.edu (H. Pashler).

http://dx.doi.org/10.1016/j.jarmac.2014.05.006
2211-3681/

retrieval practice is compared) might be particularly vulnerable to lapses in attention (see Szpunar, Khan, & Schacter, 2013). Researchers in the past have raised concerns about reading control conditions being susceptible to subjects' failing to attend to the materials for the entire presentation duration. For instance, Pressley, Symons, McDaniel, Snyder, and Turnure (1988) when assessing the benefits of elaborative interrogation for learning decided to use a more conservative control condition than in past studies: subjects had to read the information (sentences) aloud repeatedly for the entire time that it was presented. They found that the more active reading condition led to better learning compared to the relatively passive reading control used in a previous study (Pressley, McDaniel, Turnure, Wood, & Ahmad, 1987). Other studies have had reading controls that required subjects to copy by hand (i.e., write out) the target information (e.g., deWinstanley & Bjork, 2004), in an effort to ensure that the information would be properly attended to. Moreover, research on learning from prose (via reading) has shown that motivation interacts with text readability: comprehension of hard-to-read texts is generally poorer than easy texts, but this difference is smaller or eliminated in motivated subjects (Fass & Schumacher, 1978; Klare, 1976).

The vast majority of prior studies on retrieval practice have compared retrieval practice against passive rereading, a control condition that might suffer disproportionately when motivation is weak (for the reasons mentioned above). Part of the observed advantage of retrieval practice over rereading in lab studies might therefore be due to the former requiring a higher level of engagement with the material (e.g., the subject is typically asked to make an overt response within a given time limit) than the latter. It is thus plausible that increasing motivation to learn the material would have a greater effect on learning via rereading than via retrieval practice, and in turn reduce the benefit of retrieval practice over rereading.

1.2. Present study

We examined whether learner motivation modulates the benefit of retrieval practice. Specifically, we tested the hypothesis that when learner motivation is high, the advantage of retrieval practice over rereading would be attenuated, relative to when learner motivation is low. In the first two experiments learner motivation was manipulated using monetary bonuses; in the third experiment we used time savings (i.e., subjects could leave the experiment early) to motivate learning.

2. Experiment 1

2.1. Method

2.1.1. Subjects

Thirty-eight undergraduates from the psychology subject pool of a large public university participated for course credit. All subjects were proficient in English and had no prior knowledge of Swahili.

2.1.2. Stimuli

Fifty-six Swahili words paired with their English translations were selected from the Nelson and Dunlosky (1994) norms.[1] The word pairs were randomly assigned to conditions for each subject, with 7 word pairs assigned to each of 8 conditions (see Section 2.1.3).

2.1.3. Design

Three factors were manipulated within subjects: type of review (retrieval practice vs. reread), amount of review (2 vs. 4 trials), and incentive level ($0.05 vs. $0.30 monetary bonus for correct final recall of the item).

2.1.4. Procedure

Subjects were seated at computer terminals and informed that they would study a series of Swahili words paired with their English translations, one pair at a time, and that their aim was to learn the word pairs so as to be able to recall the English equivalent when cued with a Swahili word. All 56 Swahili–English word pairs were first presented once for study, each appearing for 8 s on the computer screen (with a 1-s blank screen between each pair), with the ordering of pairs randomized for each subject.

After subjects had seen all the items once, they were told that additional review would occur in one of the two forms for each item: either re-presentation of the Swahili–English pair for study (reread) or flashcard-style testing with feedback (retrieval practice). Additionally, subjects were informed that each item was associated with a monetary value—either $0.05 or $0.30—that they would receive as a bonus if they could correctly recall the English equivalent of the Swahili word on a final test 2 days later, and that the monetary value would be displayed above each item during review. In the reread condition, each trial consisted of a Swahili–English word pair being presented for 8 s followed by a 1-s blank screen (i.e., identical to the initial study presentation). In the retrieval practice condition, each trial began with a Swahili word being presented for 6 s, during which time the subject was instructed to recall and type the English equivalent if they could, followed by 2 s of the intact Swahili–English word pair (i.e., corrective feedback), and finally a 1-s blank screen. Type of review was blocked, with the items assigned to the reread condition reviewed first followed by items assigned to the retrieval practice condition, or vice versa (order of type of practice was counterbalanced across subjects). In each block, items assigned to 2 vs. 4 trials of practice and those assigned to low ($0.05) vs. high ($0.30) monetary value were all intermixed, and there were 4 consecutive cycles of review, with items assigned to 2 and 4 trials of review appearing in the first 2 cycles and only items assigned to 4 trials of review appearing in the last 2 cycles. The order of presentation in each cycle was randomized for each subject, with the constraint that the final 2 items in one cycle would not be the first 2 items in the next cycle. After completing their review of the items, subjects were dismissed.

Subjects returned 48 h after the start of the first session for a final test on all the items. The Swahili words appeared on the screen one at a time (in a random order for each subject) and the subject tried to recall and type the English equivalents (the monetary value associated with each item was not displayed). The final test was self-paced and no feedback was provided. After completing the test, subjects were thanked, debriefed, and awarded the appropriate monetary bonus based on their performance on the final test.

[1] Some readers might find it ironic that we used Swahili–English word pairs as the study stimuli in the present study, given that we had earlier raised as a potential concern the typical use of arbitrary material in memory experiments (i.e., motivation to learn the material would likely not be high). We decided to use word pairs for two main reasons: (i) a sizeable proportion of previous retrieval practice studies has used word pairs (e.g., Carrier & Pashler, 1992; Karpicke & Roediger, 2008) and we wanted our results to be relatable to prior findings, and also (ii) manipulating the type of study material to increase motivation (e.g., having less vs. more interesting texts) would introduce item differences as a potential confound. Our strategy therefore was to use arbitrary study material (in continuity with the bulk of past studies) coupled with incentives to increase learner motivation.

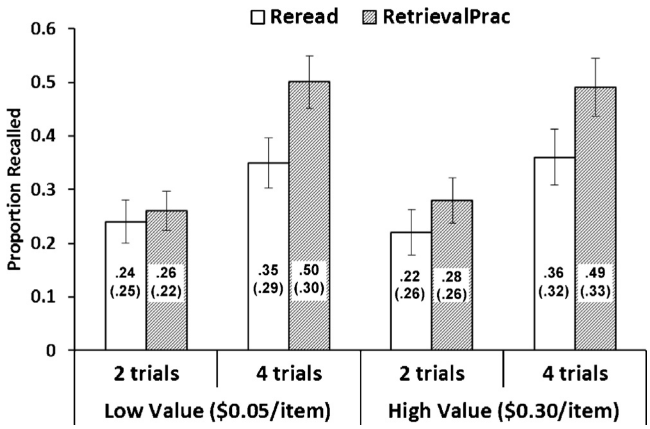

Fig. 1. Final test performance in Experiment 1 as a function of type of review, amount of review, and incentive level. *M* and *SD* for each condition are listed within the respective bars. Error bars indicate standard errors of the means.

2.2. Results and discussion

The α-level for all analyses was set at .05. Subjects' responses on the final test were scored for accuracy (responses that contained minor spelling errors were still counted as correct). Fig. 1 shows mean final test performance as a function of condition. As can be seen from the figure, retrieval practice produced performance superior to rereading and 4 review trials yielded better retention than 2 trials. In addition, the benefit of retrieval practice appeared to increase with more trials of practice, but crucially the magnitude of the benefit did not seem to depend on the level of incentive.

The above observations were confirmed by a 2 (type of review) × 2 (amount of review) × 2 (incentive level) within-subjects ANOVA, which revealed main effects of type of review [$F(1, 37) = 11.71$, $p = .002$, $MSE = .58$, $\eta_p^2 = .24$] and amount of review [$F(1, 37) = 88.96$, $p < .001$, $MSE = 2.35$, $\eta_p^2 = .71$], as well as a significant interaction between these two factors [$F(1, 37) = 7.38$, $p = .01$, $MSE = .20$, $\eta_p^2 = .17$]. With regard to our main hypothesis, there was no interaction between type of review and incentive level [$F < 1$].

Our results clearly demonstrate the advantage of retrieval practice over rereading, consistent with many previous findings (e.g., Carrier & Pashler, 1992). However, contrary to our hypothesis, we found no evidence that the effect was modulated by learner motivation. Whether the incentive to learn was high or low, the size of the retrieval practice benefit did not differ significantly. It should be noted, however, that we did not find a main effect of incentive level [$F < 1$], which leaves open the possibility that our incentive manipulation was perhaps not effective in altering subjects' motivation to learn the items. If this was the case, then the experiment was not an adequate test of our hypothesis.

Our failure to find an effect of incentive level may seem surprising in light of studies showing that young and healthy older adults demonstrate better memory for items of higher value. When tasked with memorizing a list of items varying in point values and asked to maximize the total point value of items they can later recall, the items recalled at test tend to be ones that are of highest value (Castel, Balota, & McCabe, 2009; Castel, Benjamin, Craik, & Watkins, 2002; Castel et al., 2011). At first blush, it would seem that our subjects should have utilized the information about the monetary value of each item during review to maximize their earnings on the final test. But upon closer inspection, the experiments by Castel and colleagues used much shorter lists (e.g., 12 words), and free recall was assessed usually immediately after presentation of a list. It seems plausible that working memory played a much larger role in driving performance in those studies, relative to our experiment in which the final test was administered after a 2-day retention interval (see also Eysenck & Eysenck, 1982). Also, strategic control of attention during study is at least in part guided by the individual's metacognitive monitoring of the learning situation (e.g., Metcalfe & Finn, 2008; Nelson & Narens, 1994); it is possible that with 56 word pairs to study there were just too many items for subjects to keep track of the monetary value of each item (since the high- and low-value items were randomly intermixed), and so they ended up treating all items the same (disregarding the monetary value that accompanied each item). Our next experiment was aimed at avoiding this potential problem by making the incentive scheme more straightforward: incentive level was manipulated between subjects in Experiment 2, with subjects either informed or not informed about the monetary bonus they would receive for each item correctly recalled on the final test.

3. Experiment 2

3.1. Method

3.1.1. Subjects

Fifty-nine undergraduates from the subject pool participated for course credit, none of whom participated in the previous experiment.

3.1.2. Stimuli

The same 56 Swahili–English word pairs were used as in Experiment 1. The word pairs were randomly assigned to conditions for each subject, with 14 word pairs assigned to each of 4 within-subjects conditions (see Section 3.1.3).

3.1.3. Design

Two factors were manipulated within subjects: type of review (retrieval practice vs. reread) and amount of review (2 vs. 4 trials). Incentive level (either informed or not informed about the $0.25 monetary bonus for correct final recall of each item) was manipulated between subjects, with subjects randomly assigned to either condition.

3.1.4. Procedure

The procedure was identical to that for the previous experiment except for the following two aspects. First, after the initial study presentation of all the items and prior to the review of the items, subjects that were assigned to the incentive-aware condition were informed that they would receive $0.25 bonus for each item they could correctly recall on the final test in 2 days; subjects that were assigned to the incentive-unaware condition were not informed of the bonus. Second, during the review of the items no monetary value was displayed with each item.

3.2. Results and discussion

Mean final test performance is shown in Fig. 2. As can be seen from the figure, the pattern of results was strikingly similar to that of the previous experiment. Just as in Experiment 1, there was an advantage of retrieval practice over rereading, and 4 review trials produced better retention than 2 trials. Also, the magnitude of the retrieval practice advantage appeared to increase with more trials of practice, but like before the effect of retrieval practice did not seem to interact with incentive level.

A 2 (type of review) × 2 (amount of review) × 2 (incentive level) mixed ANOVA confirmed the above observations. There were main effects of type of review [$F(1, 57) = 53.81$, $p < .001$, $MSE = .98$, $\eta_p^2 = .49$] and amount of review [$F(1, 57) = 128.34$, $p < .001$, $MSE = 2.08$, $\eta_p^2 = .69$], as well as a significant interaction between these two factors [$F(1, 57) = 20.71$, $p < .001$, $MSE = .31$, $\eta_p^2 = .27$]. Again, there

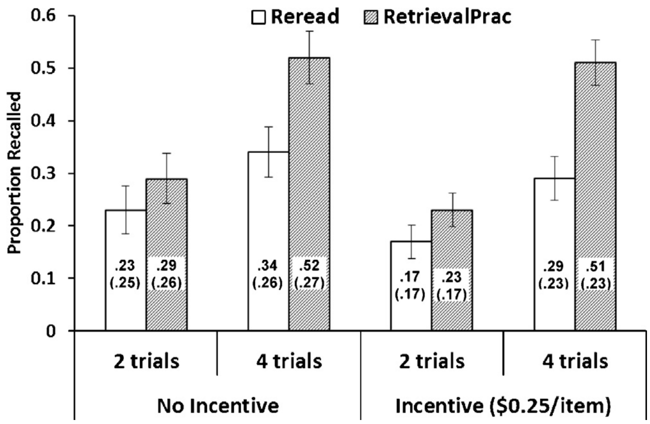

Fig. 2. Final test performance in Experiment 2 as a function of type of review, amount of review, and incentive level. *M* and *SD* for each condition are listed within the respective bars. Error bars indicate standard errors of the means.

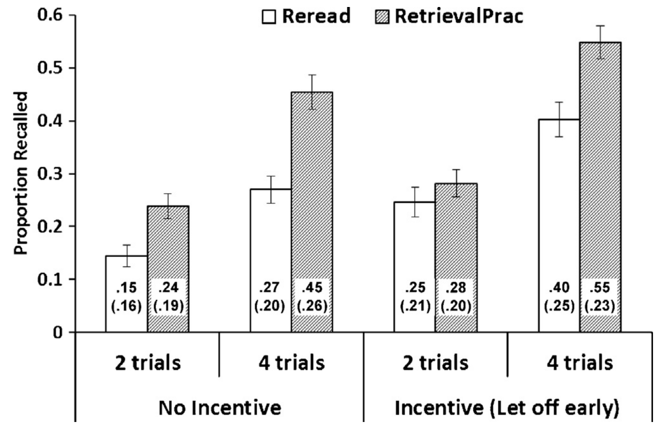

Fig. 3. Final test performance in Experiment 3 as a function of type of review, amount of review, and incentive level. *M* and *SD* for each condition are listed within the respective bars. Error bars indicate standard errors of the means.

was neither a main effect of incentive level [$F < 1$] nor an interaction between type of review and incentive level [$F < 1$].

The results of the present experiment replicate the findings from Experiment 1. Retrieval practice was a more potent learning strategy than rereading, and the effect was magnified with additional practice trials. Importantly, the efficacy of retrieval practice did not vary as a function of incentive level. But, again, whether or not subjects were incentivized to learn did not affect overall performance, even when the incentive scheme was simplified and manipulated between subjects (cf. Experiment 1). It is possible that the monetary bonus was too small to increase the motivation of subjects in the incentive-aware condition (the maximum one could earn was $14 for perfect performance on the final test) and/or subjects had difficulty judging how many items they would remember in 2 days, which might have led to discounting of the potential bonus. In order to provide a stronger test of our hypothesis that learner motivation modulates the relative efficacy of retrieval practice, we used a different incentive in the next experiment—the opportunity to save time (through early release from the experiment).

4. Experiment 3

4.1. Method

4.1.1. Subjects

120 college students from the subject pool participated for course credit. None of them participated in the previous two experiments.

4.1.2. Stimuli

The same Swahili–English word pairs were used as in the previous experiments. The word pairs were randomly assigned to conditions for each subject, with 14 word pairs assigned to each of 4 within-subjects conditions (see Section 4.1.3). Also, an additional 12 pairs were selected from the Nelson and Dunlosky (1994) norms to be used as filler items. These filler items were randomly intermixed with the other items during the initial presentation and the review phase (3 filler pairs assigned to each of the 4 within-subjects conditions), and were presented for a brief final test at the end of the first session. The filler items were excluded from the final test in the second session (2 days later).

4.1.3. Design

The overall design was identical to that of the previous experiment. Only the type of incentive was different: instead of using monetary bonuses, the present experiment offered subjects the chance to save time (i.e., leave early from the experiment). Subjects were either made aware or not of the incentive.

4.1.4. Procedure

The procedure was basically the same as that for Experiment 2 with the following exceptions. First, subjects that were assigned to the incentive-aware condition were informed just prior to the review phase that the duration of their second session would be determined by how well they learned the word pairs in the first session; specifically, subjects were told that at the end of the current session they would be tested on a subset of the items, and if they managed to recall correctly at least 5 out of 12 of the test items their second session would last less than 10 min, but if they failed to meet that criterion their second session would last the full hour. Subjects that were assigned to the incentive-unaware condition were not informed that good performance would lead to time savings in the second session. Second, all subjects were presented with 68 word pairs to learn in the first session, 12 of which were filler items that were used to determine whether the subject qualified for the time savings in the second session. Third, at the end of the first session (after review of all the items), all subjects received a self-paced cued recall test of the 12 filler items. Subjects that scored at least 5 correct underwent a second session that was identical to that in the previous two experiments—i.e., a self-paced final cued recall test on the 56 target items. Subjects that did not surpass the threshold were also given the same final test in the second session, but after the test they were given additional review of the items as a time-filler task.

4.2. Results and discussion

Mean final performance is displayed in Fig. 3. In many respects, the pattern of performance was similar to those in Experiments 1 and 2. The key difference for the present results is that the time savings incentive did appear to improve performance.

A 2 (type of review) × 2 (amount of review) × 2 (incentive level) mixed ANOVA corroborated the above observations. There were main effects of type of review [$F(1, 118) = 61.15$, $p < .001$, $MSE = 1.58$, $\eta_p^2 = .34$], amount of review [$F(1, 118) = 206.44$, $p < .001$, $MSE = 4.34$, $\eta_p^2 = .64$], and incentive level [$F(1, 118) = 8.47$, $p = .004$, $MSE = 1.02$, $\eta_p^2 = .07$]. Also, type of review interacted significantly with amount of review [$F(1, 118) = 20.72$, $p < .001$, $MSE = .30$, $\eta_p^2 = .15$]. No other effect was statistically reliable.[2]

We again found that retrieval practice enhanced long-term learning more than rereading and that the benefit was not modulated by learner motivation. The incentive manipulation in the present experiment raised overall final performance (unlike in the previous two experiments), but without reliably interacting with type of review.

5. General discussion

There has been a wealth of evidence—mostly from lab experiments—that retrieving information from memory promotes long-term learning and retention more than rereading of the target information (Carpenter, 2012). This phenomenon would appear to have great application in educational practice. However, lab studies of retrieval practice have generally asked subjects to learn material without the intrinsic or extrinsic rewards usually present in real-world educational contexts. As pointed out above, retrieval practice (a strategy that requires active responding) might be especially beneficial compared to more passive strategies (e.g., rereading) when learner motivation is low. When the learner is highly motivated to learn, on the other hand, perhaps the study strategy employed makes less of a difference. We manipulated the incentives for learning in the present study to investigate the hypothesis that the advantage of retrieval practice over rereading would be attenuated when learner motivation is relatively high. In three experiments we repeatedly found a superiority of retrieval practice over rereading for long-term retention—a benefit that increased with more trials of practice, consistent with prior research (e.g., Karpicke & Roediger, 2008). However, contrary to our hypothesis, we did not find any evidence that learner motivation modulates the benefit of retrieval practice.

In Experiments 1 and 2, we used cash incentives to motivate learning but we failed to obtain a main effect of incentive level, suggesting that learner motivation was unchanged by our incentive manipulation (and thus an inadequate test of the hypothesis). We were initially puzzled by the lack of effect of cash incentives, particularly given that individuals do exhibit value-directed remembering (e.g., Castel et al., 2002, 2009, 2011). There are, however, a couple of key procedural differences between the present experiments and those by Castel and colleagues that could account for the apparent difference in findings: they used 12-item study lists and memory was assessed immediately after the presentation of each list. It is therefore likely that working memory contributed substantially to value-directed remembering (i.e., subjects could selectively maintain high value items in working memory until the recall test), and not so for final test performance in the present experiments (number of to-be-remembered items exceeded working memory capacity; 2-day retention interval). Additionally, financial incentives failing to improve performance are not without precedent in the memory literature. For instance, in Nilsson's (1987) study subjects were presented with lists of words and were given both immediate and final free recall tests; offering a cash incentive of $10 for the best performer (in a group of 10 subjects) failed to improve performance, relative to a control group that was offered no incentive, even though the incentive group reported higher motivation to do well (see also Ngaosuvan & Mäntylä, 2005).

Why then did the prospect of being released early from the experiment succeed in motivating subjects in Experiment 3 to learn the material? The incentive scheme may have played a role. Experiments 1 and 2 employed what has been described as a piece-rate scheme—a predefined amount of money is paid for each unit of output—whereas Experiment 3 used a quota scheme by which a reward is received when a certain target level of performance is reached (Bonner, Hastie, Sprinkle, & Young, 2000). There is ample evidence that specific, challenging (but achievable) goals tend to increase motivation and effort and thus improve performance (Locke & Latham, 2002). Perhaps it was this goal-setting aspect of the quota incentive scheme that was responsible for boosting motivation and, in turn, learning.

5.1. Effect of motivation on retrieval practice

Our results indicate that learner motivation did not modulate the *benefit* of retrieval practice for learning, which properly understood means that the difference in final performance between the retrieval practice and control (reread) conditions was equivalent regardless of incentive level. The invariance of the magnitude of the retrieval practice advantage over rereading across different levels of motivation suggests that the beneficial effect of retrieval practice is probably not driven by relatively lower attention or engagement in the control condition.

Learner motivation did, however, have an effect on retrieval practice, at least in Experiment 3 in which subjects that were aware of the incentive showed better learning (in all conditions, including the retrieval practice condition). In other words, learning via retrieval practice is susceptible to variations in motivation level, and retrieval practice is not a panacea for low motivation (although it does improve learning relative to rereading).

How might learner motivation influence the (absolute) effectiveness of retrieval practice? One obvious candidate is the amount of attention the learner devotes to the learning task. During explicit learning, one needs to pay attention in order to encode the information. If the learner's mind is wandering and s/he is not focused on the task at hand, then clearly there will be decrements to learning. This account applies, of course, to all study strategies, not just retrieval practice. There are other potential explanations that are specific to retrieval practice. Even though inducements at retrieval usually do not improve the quantity of accurate recall on a given test (Koriat & Goldsmith, 1994; Roediger & Payne, 1985), they potentially influence the retrieval process in ways that promote later learning and retention. Some plausible candidate mechanisms (none of which are mutually-exclusive) include: (i) greater effort or time expended in the search of memory space (Chan, McDermott, & Roediger, 2006); (ii) enhanced elaborative activation of related information, which then serves as effective mediators for subsequent retrieval (Carpenter, 2009); and (iii) greater likelihood of switching from less effective to more effective mediators during retrieval failure (Pyc & Rawson, 2012), which in turn increases test-potentiated learning of the corrective feedback (Arnold & McDermott, 2013).

[2] In the interest of full transparency, we would like to disclose that when we presented this study at the Annual Meeting of the Psychonomic Society in 2012 we had preliminary data from only 57 subjects. The pattern of results with just the subset of $N = 57$ was identical (i.e., the same effects were statistically significant) except that the main effect of incentive level at that juncture was significant only at the one-tailed level [$t(55) = 1.79$, $p = .04$, one-tailed, $d = 0.47$]. We decided to run more subjects in the experiment until we had roughly double the N, in part because there appeared to be a trend (from looking at a graphical display of the means) towards an interaction between type of review and incentive level [$F(1, 55) = 1.14$, $p = .29$, $\eta_p^2 = .02$], and we felt that having a much larger sample size would be necessary before we could confidently conclude that the benefit of retrieval practice was not modulated by learner motivation. It has been demonstrated that repeatedly running statistical tests and incrementing the sample size after each test (in the hopes of finding a significant result) inflates the overall Type I error rate, but the choice to increase the sample size was not made in any biased effort to obtain a significant finding (the "p-hacking" practice criticized by Simmons, Nelson, & Simonsohn, 2011).

5.2. Practical implications

The present study was prompted by a concern that as robust as the benefit of retrieval practice was in the literature, the extant effect sizes could be exaggerated due to low motivation to learn in lab studies. In particular, the arbitrary nature of the to-be-learned material in these studies likely means that the subjects had little incentive to maximize their learning, unlike in most real-life learning situations in which students purposefully try to learn (even if the goal is merely to pass an exam). In cases where learner motivation is relatively low, it is plausible that passive learning conditions (e.g., rereading) are especially vulnerable to lapses in attention or mind-wandering, which would artificially augment the advantage of retrieval practice (which requires active responding). This idea is disconfirmed by our results. The results are encouraging for the idea that practicing retrieval enhances learning more than rereading regardless of whether motivation is high or low, bolstering previous recommendations that educators and students should adopt retrieval practice as a strategy for efficient and durable learning.

Conflict of interest

The authors declare that they have no conflict of interest.

Acknowledgments

This work was supported by the Institute of Education Sciences (US Department of Education, Grant R305B070537 to H. Pashler), the National Science Foundation (Grant BCS-0720375, H. Pashler, PI; and Grant SBE-0542013, G.W. Cottrell, PI), a Walter and Constance Burke Research Initiation Award (S. Kang, PI), and a collaborative activity award from the J.S. McDonnell Foundation. We acknowledge the contribution of the following individuals: Joshua Martinez, Stephanie Ho, and Kevin Nguyen programmed the experiments; Noriko Coburn, Spencer Chu, and Vivian Chen assisted with data collection. Part of this research was presented at the 53rd Annual Meeting of the Psychonomic Society, Minneapolis, MN, in November 2012.

References

Abbott, E. E. (1909). On the analysis of the factor of recall in the learning process. *Psychological Monographs*, *11*, 159–177.

Arnold, K. M., & McDermott, K. B. (2013). Test-potentiated learning: Distinguishing between direct and indirect effects of tests. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *39*, 940–945.

Bonner, S. E., Hastie, R., Sprinkle, G. B., & Young, S. M. (2000). A review of the effects of financial incentives on performance in laboratory tasks: Implications for management accounting. *Journal of Management Accounting Research*, *12*, 19–64.

Carpenter, S. K. (2009). Cue strength as a moderator of the testing effect: The benefits of elaborative retrieval. *Journal of Experimental Psychology: Learning, Memory & Cognition*, *35*, 1563–1569.

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

Carrier, M., & Pashler, H. (1992). The influence of retrieval on retention. *Memory & Cognition*, *20*, 633–642.

Castel, A. D., Balota, D. A., & McCabe, D. P. (2009). Memory efficiency and the strategic control of attention at encoding: Impairments of value-directed remembering in Alzheimer's disease. *Neuropsychology*, *23*, 297–306.

Castel, A. D., Benjamin, A. S., Craik, F. I., & Watkins, M. J. (2002). The effects of aging on selectivity and control in short-term recall. *Memory & Cognition*, *30*, 1078–1085.

Castel, A. D., Humphreys, K. L., Lee, S. S., Galván, A., Balota, D. A., & McCabe, D. P. (2011). The development of memory efficiency and value-directed remembering across the life span: A cross-sectional study of memory and selectivity. *Developmental Psychology*, *47*, 1553–1564.

Chan, J. C. K., McDermott, K. B., & Roediger, H. L. (2006). Retrieval-induced facilitation: Initially nontested material can benefit from prior testing of related material. *Journal of Experimental Psychology: General*, *135*, 553–571.

Deci, E. L., Vallerand, R. J., Pelletier, L. G., & Ryan, R. M. (1991). Motivation and education: The self-determination perspective. *Educational Psychologist*, *26*, 325–346.

Dempster, F. N. (1992). Using tests to promote learning: A neglected classroom resource. *Journal of Research and Development in Education*, *25*, 213–217.

deWinstanley, P. A., & Bjork, E. L. (2004). Processing strategies and the generation effect: Implications for making a better reader. *Memory & Cognition*, *32*, 945–955.

Dweck, C. S. (1986). Motivational processes affecting learning. *American Psychologist*, *41*, 1040–1048.

Efklides, A. (2012). Commentary: How readily can findings from basic cognitive psychology research be applied in the classroom? *Learning and Instruction*, *22*, 290–295.

Eysenck, M. W., & Eysenck, M. C. (1982). Effects of incentive on cued recall. *Quarterly Journal of Experimental Psychology*, *34*, 489–498.

Fass, W., & Schumacher, G. M. (1978). Effects of motivation, subject activity, and readability on the retention of prose materials. *Journal of Educational Psychology*, *70*, 803–807.

Kang, S. H. K., McDermott, K. B., & Roediger, H. L. (2007). Test format and corrective feedback modify the effect of testing on long-term retention. *European Journal of Cognitive Psychology*, *19*, 528–558.

Karpicke, J. D., & Grimaldi, P. J. (2012). Retrieval-based learning: A perspective for enhancing meaningful learning. *Educational Psychology Review*, *24*, 401–418.

Karpicke, J. D., & Roediger, H. L. (2008). The critical importance of retrieval for learning. *Science*, *319*, 966–968.

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.

Klare, G. R. (1976). A second look at the validity of readability formulas. *Journal of Literacy Research*, *8*, 129–152.

Koriat, A., & Goldsmith, M. (1994). Memory in naturalistic and laboratory contexts: Distinguishing the accuracy-oriented and quantity-oriented approaches to memory assessment. *Journal of Experimental Psychology: General*, *123*, 297–315.

Lachman, R., & Laughery, K. R. (1968). Is a test trial a training trial in free recall learning? *Journal of Experimental Psychology*, *76*, 40–50.

Lepper, M. R., Greene, D., & Nisbett, R. E. (1973). Undermining children's intrinsic interest with extrinsic reward: A test of the overjustification hypothesis. *Journal of Personality and Social Psychology*, *28*, 129–137.

Locke, E. A., & Latham, G. P. (2002). Building a practically useful theory of goal setting and task motivation: A 35-year odyssey. *American Psychologist*, *57*, 705–717.

Lundeberg, M. A., & Fox, P. W. (1991). Do laboratory findings on test expectancy generalize to classroom outcomes? *Review of Educational Research*, *61*, 94–106.

McDaniel, M. A., Wildman, K. M., & Anderson, J. L. (2012). Using quizzes to enhance summative-assessment performance in a web-based class: An experimental study. *Journal of Applied Research in Memory and Cognition*, *1*, 18–26.

Metcalfe, J., & Finn, B. (2008). Evidence that judgments of learning are causally related to study choice. *Psychonomic Bulletin and Review*, *15*, 174–179.

Nelson, T. O., & Dunlosky, J. (1994). Norms of paired-associate recall during multitrial learning of Swahili–English translation equivalents. *Memory*, *2*, 325–335.

Nelson, T. O., & Narens, L. (1994). Why investigate metacognition? In J. Metcalfe, & A. P. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 1–25). Cambridge, MA: MIT Press.

Ngaosuvan, L., & Mäntylä, T. (2005). Rewarded remembering: Dissociations between self-rated motivation and memory performance. *Scandinavian Journal of Psychology*, *46*, 323–330.

Nilsson, L. G. (1987). Motivated memory: Dissociation between performance data and subjective reports. *Psychological Research*, *49*, 183–188.

Pressley, M., McDaniel, M. A., Turnure, J. E., Wood, E., & Ahmad, M. (1987). Generation and precision of elaboration: Effects on intentional and incidental learning. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *13*, 291–300.

Pressley, M., Symons, S., McDaniel, M. A., Snyder, B. L., & Turnure, J. E. (1988). Elaborative interrogation facilitates acquisition of confusing facts. *Journal of Educational Psychology*, *80*, 268–278.

Pyc, M. A., & Rawson, K. A. (2012). Why is test-restudy practice beneficial for memory? An evaluation of the mediator shift hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*, 737–746.

Rawson, K. A., & Dunlosky, J. (2012). When is practice testing most effective for improving the durability and efficiency of student learning? *Educational Psychology Review*, *24*, 419–435.

Roediger, H. L., Agarwal, P. K., Kang, S. H. K., & Marsh, E. J. (2010). Benefits of testing memory: Best practices and boundary conditions. In G. M. Davies, & D. B. Wright (Eds.), *Current issues in applied memory research* (pp. 15–49). East Sussex, UK: Psychology Press.

Roediger, H. L., & Payne, D. G. (1985). Recall criterion does not affect recall level or hypermnesia: A puzzle for generate/recognize theories. *Memory & Cognition*, *13*, 1–7.

Rohrer, D., Taylor, K., & Sholar, B. (2010). Tests enhance the transfer of learning. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *36*, 233–239.

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, *22*, 1359–1366.

Szpunar, K. K., Khan, N. Y., & Schacter, D. L. (2013). Interpolated memory tests reduce mind wandering and improve learning of online lectures. *Proceedings of the National Academy of Sciences*, *110*, 6313–6317.

Tulving, E. (1967). The effects of presentation and recall of material in free-recall learning. *Journal of Verbal Learning and Verbal Behavior*, *6*, 175–184.