



Is recitation an effective tool for adult learners?

Kathleen B. McDermott*, Farah Naaz

Washington University in St. Louis, United States



ARTICLE INFO

Article history:

Received 2 April 2014

Received in revised form 24 June 2014

Accepted 27 June 2014

Available online 8 July 2014

Keywords:

Recitation

Testing effect

Learning

Covert retrieval

ABSTRACT

The effects of recitation on subsequent recall were examined in 4 experiments modeled after those of Gates (1917). Participants underwent a study phase, a recitation phase, and a test phase. During the recitation phase participants were to attempt to recall the previously studied material and then to restudy it when they could not recall any new information. They were encouraged to switch between recalling and restudying. The proportion of the total acquisition time that was spent in recitation was varied. Unlike the classic findings reported by Gates (1917) in schoolchildren, there was no consistent evidence that recitation enhanced learning in these adult learners.

© 2014 Society for Applied Research in Memory and Cognition. Published by Elsevier Inc. All rights reserved.

Retrieval practice is well-established to be an effective strategy for learning (McDermott, Arnold, & Nelson, 2014; Roediger & Karpicke, 2006a). Indeed, many articles and reviews covering the benefits of retrieval practice begin with an historical overview highlighting the classic study of Gates (1917)—entitled *Recitation as a Factor in Memorizing*—as one of the foundational reports demonstrating how retrieval practice can benefit student learning.

By recitation, Gates was referring to a specific type of self-testing wherein a student had the to-be-learned information in front of them and tried to recall the information covertly, glancing back to the memoranda whenever retrieval began to fail and then returning to covert recall. Gates was motivated to empirically test the claim of Francis Bacon (1620/2000), who had asserted that “If you read anything over twenty times you will not learn it by heart so easily as if you were to read it only ten, trying to repeat it between whiles, and when memory failed looking at the book.” Gates was also interested in knowing when in the learning process recitation should be introduced; that is, perhaps only after significant study would such self-testing be beneficial.

Gates reviewed the sparse literature to date, noting that Katzaroff (1908) had shown that adults learning nonsense syllable pairs recalled more when they spent time in recitation in addition to reading (relative to spending all the acquisition time in the reading phase). A more careful look at this procedure suggests that the study actually employed overt cued recall during this “recitation” phase and is not, therefore, recitation, at least as defined by Gates and in the present report.

Thorndike (1914) gave 28 adults 4 sets of vocabulary words to learn and demonstrated no benefit of recitation. That is, for the sets learned only by reading and re-reading, people later recalled 82%; for those learned by reading and self-testing, people later recalled 72%. Thorndike did not draw strong conclusions from this null effect, however. He concluded “The experiment was too crude and too slight to give numerical results worth presenting in detail.”

The one study cited by Gates that seems most promising in setting up his own study was one reported by Kuhn, 1914, who showed that adults learning verses, words, or nonsense syllables learned better when they were able to engage in recitation (relative to just reading and re-reading). Exactly how to interpret these results is in question, however, in that the dependent variable was the number of repetitions engaged by the subject “until he was confident of his mastery of the material” (Gates, 1917, p. 6). That is, recitation’s effects on objective recall probability are still in question.

Gates’s (1917) results were more robust, although even here one can see limitations to the generalizability of the results. Depicted in Fig. 1 are a subset of his findings, on an immediate test with nonsense materials (nonsense syllables) and meaningful materials (biographies). All children were given 9 min to learn a set of materials. What differed was how that time was spent. In the 5 conditions depicted in Fig. 1, students spent the entire time studying (the 100:0 condition) or varying amounts of time reciting after initial study (with the ratio of percent study time to percent recitation time varying from 80:20, 60:40, 40:60, and 20:80). Gates observed that more recitation led to greater recall, both immediately (shown in Fig. 1) and after a delay (not shown). A second empirical finding can be seen from this figure, as well: The recitation effect was much greater for nonsense syllables than for biographies.

The motivation behind the present studies was to explore the optimal configurations of study and recitation within adult

* Corresponding author at: Department of Psychology, CB1125, Washington University, St. Louis, MO 63130-4899, United States. Tel.: +1 3149358743.

E-mail address: kathleen.mcdermott@wustl.edu (K.B. McDermott).

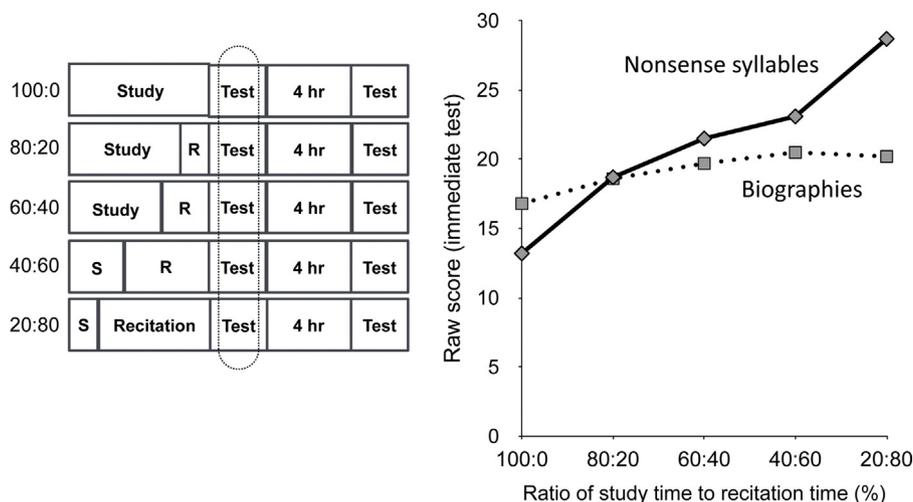


Fig. 1. Left panel. Procedures used by Gates (1917). Students worked with materials for 9 min, always beginning by reading the materials. Most groups then performed recitation for varying amounts of time, from 100% of the time (the 100:0 condition, referring to 100% studying and 0% recitation) to 20% of the time (the 20:80 condition). Right panel: Selected results reported by Gates (1917). As the ratio of study time to recitation decreased, subsequent recall increased; this pattern was more pronounced for nonsense syllables than for biographies. Data from Tables XVII, p. 36, and XXI, p. 40, average of grades 4, 6, and 8.

learners. As a starting point, we sought to establish the basic phenomenon, using conditions similar to some of Gates's more extreme conditions to establish the finding within a sample of adult subjects. We began with unrelated words in Experiment 1 and (as will be seen) failed to achieve any differences as a result of the proportion of time spent self-testing. In Experiments 2–4 we moved progressively closer to the procedures of Gates but consistently failed to find any benefits of recitation.

1. Experiment 1

All subjects were given a fixed amount of time (5 min) to learn a set of English words; the independent variable was how this time was spent. Specifically, some subjects spent 70% of their time studying, followed by 30% of the time in self-testing (the 70:30 condition). Other subjects spent 30% of their time studying, followed by a self-testing phase (70% of their time, the 30:70 condition). To the extent that this self-testing is effective, the group spending more time undergoing this process would be expected to recall more on the final test. Note that this procedure differs from that used by Gates in part because the cued recall phase was overt, not covert. As Gates defined recitation (a definition adopted here), recitation involves *covert*, self-testing with self-administered feedback. Experiments 2–4 will use recitation proper.

1.1. Method

1.1.1. Participants

Sixty-four participants were recruited from an online subject pool (Amazon Mechanical Turk) and compensated \$1 for every 15 min of participation; the experiment took 13.2 min on average (range = 11.2–20.4). Fifty participants (mean age 37.0 years, range 18–60 years; 21 males) were included in the data analyses. The data from 14 subjects were excluded because they reported noting down the study words ($n = 9$) or not understanding the instructions ($n = 5$).

1.1.2. Design

Participants were randomly assigned to one of two learning groups. After applying the exclusion criteria, 24 participants remained in the 70:30 group and 26 participants in the 30:70 group. All participants went through the 4 phases in the experiment:

study, self-testing with feedback, Tetris (a game used as a distractor to prevent rehearsal), and a final free recall test. The proportion of time spent in study and in self-testing was varied across conditions.

1.1.3. Materials

Thirty unrelated English words were selected from the norms of Nelson, McEvoy, and Schreiber (2004). The words had a mean concreteness rating of 6.15 (range 5–7) and mean word length of 6.20 (range 5–9).

1.1.4. Procedure

Participants were informed that the experiment would consist of 4 parts: study, self-testing, Tetris, and a memory test. Specifically, they were informed that they would attempt to learn 30 words and that they would then take a test on those words (looking back to the studied words as needed). They would then play Tetris for 3 min and take a final test on the 30 studied words.

The 30 unrelated English words were presented on the computer's display (simultaneously, with 3 columns of 10 words each). The instruction "STUDY" appeared at the top of the screen, as depicted in Fig. 2. Subjects were asked to encode the words in preparation for an upcoming memory test. This phase lasted 1.5 or 3.5 min, depending on condition, and was followed by the self-testing phase. For this phase and all phases, a countdown timer on the bottom right of the screen informed subjects how much time remained in the current phase.

Self-testing began with a blank screen, which prompted participants to type in as many of the words as they could recall (Fig. 2). Further, they were told "When you cannot recall any more words, click on the 'Study Words Again' button. . .to see all of the words again. Study all the words briefly, and then try recalling them again by clicking on the 'Practice Recalling the Words Again' button." They were to iteratively switch between recall and encoding until their time was up (3.5 or 1.5 min, depending upon condition). They were further informed that the "goal should be to learn as many words as you can for the final test. We encourage you to try to test yourself and restudy the words as often as possible during this time."

All participants then played Tetris for 3 min, after which they took the free recall test (2 min). Here, they were instructed to type in as many words as they could remember in any order they pleased.

In all experiments, participants received questionnaires after the primary tasks. These questionnaires asked demographic

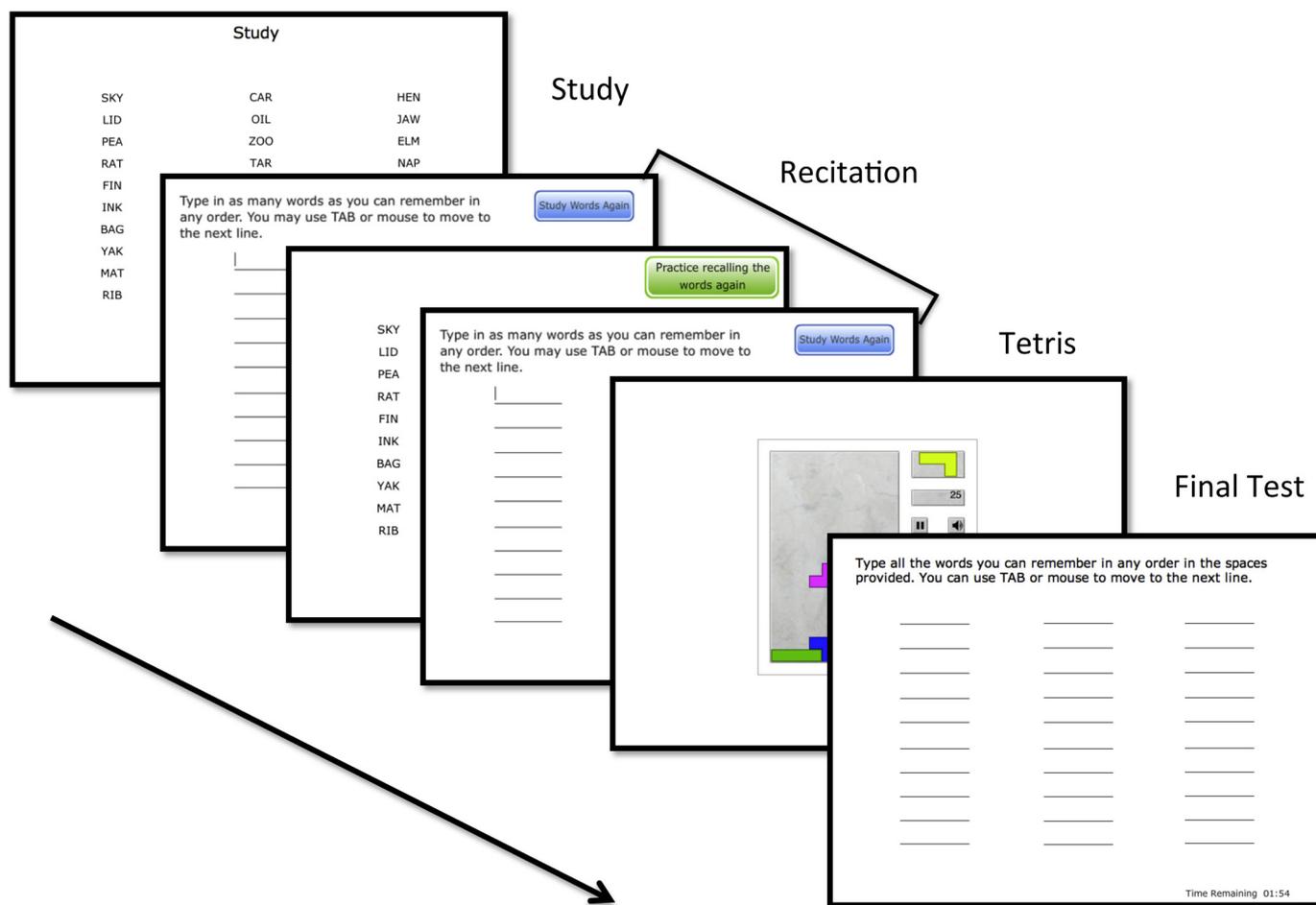


Fig. 2. Procedures of Experiment 1. Specifically, screenshots of the study, recitation, distractor task, and final test phase are shown.

questions and questions regarding the subject's level of participation in the experiment (e.g., expertise in English, level of interest, effort, difficulty involved in the study, whether there was anything confusing about the study, and whether they wrote down the study words). These questions were included to ensure that participants understood the task and as a check to exclude participants who did not follow the instructions.

1.2. Results and discussion

During the self-testing phase, the 70:30 group took approximately 2 tests ($M=1.8$, $SD=0.8$), whereas the 30:70 group took approximately 3 tests ($M=2.7$, $SD=1.6$). Subjects did appear to attempt recall on the first test; those in the 70:30 group recalled 40.7% of the words; those in the 30:70 group 38.9%. Similar performance was observed on tests 2 and 3.

On the final test, both the learning groups recalled approximately 54% of the words, as can be seen in Fig. 3 (left panel). Specifically, the 70:30 group recalled 54.9% ($SD=17.2$) whereas the 30:70 group recalled 53.8% ($SD=16.0$). A t -test revealed no significant difference between the groups, $t(1,48)=0.216$, $p=0.83$.

Thirteen participants remained on the test screen for the entire duration of the self-testing phase (n : 70:30=8; 30:70=6); they never returned to the study screen to review the full set of items. Removing these subjects from the analysis results in no change in the conclusions—52% recall on the final test for the 70:30 group and 53.0% recall for the 30:70 group. Further, if we restrict the

analysis to people ($N=8$) who took 4 or more initial tests in the 70:30 condition, the performance is still the same ($M=52.5\%$).

The procedure of using explicit recall (typing responses) was not the way Gates implemented (or defined) recitation, so in an attempt to more closely approximate his procedures, to get closer to “true” recitation (as opposed to explicit testing with feedback), and to gain more control of the instructional phase, we conducted Experiment 2.

2. Experiment 2

Experiment 2 was conducted in the laboratory at Washington University in St. Louis, and participants were instructed in-person by the experimenter to ensure that they understood the task. Procedures followed those of Experiment 1 with the following exceptions. The self-testing phase consisted of a blank screen for covert practice of recalling the material, and participants could practice recall and study the material as often as they needed to. The retention interval was 10 min. The duration for the final test was increased from 2 to 5 min to ensure that participants had sufficient time to record all the words they remembered.

2.1. Method

2.1.1. Participants

Participants were recruited through advertisement on the campus of Washington University in St. Louis. Participants could opt for either cash or research credit as compensation for their time (\$5 or

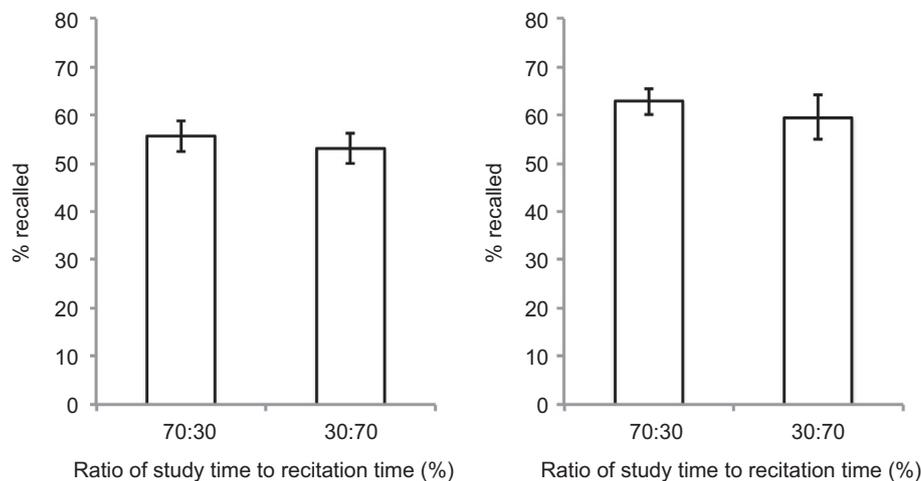


Fig. 3. Mean percent correct performance on the final test in the two recitation groups in Experiment 1 (left panel) and Experiment 2 (right panel). Error bars display the standard error of the mean.

0.5 credit for 30 min of participation). Volunteers under 18 years of age were allowed to participate in this experiment and all subsequent experiments for research credit, but their data were not included in the final analyses.

Fifty-seven participants volunteered for the study. Fifty-two participants were included in the final data analyses (21 males). Data from one subject were lost (technical issue), and 4 subjects were excluded from analysis because they did not follow instructions ($n=2$), reported being confused about the study ($n=1$), or reported fewer than 10 years of experience with English language ($n=1$). The mean age of the participants included in the final data analyses was 19.3 (range 18–22).

2.1.2. Design

As in Experiment 1 there were two learning groups, and everyone spent an equivalent amount of time with the material during the initial acquisition phase; we varied the ratio of study: recitation time (70:30, and 30:70). Participants were randomly assigned to one of the learning groups. After applying the exclusion criteria there were 28 participants in 70:30 group and 24 participants in 30:70 group. All participants went through the 4 phases in the experiment: study, recitation, Tetris, and a final free recall test.

2.1.3. Materials

Materials for the experiment were the same 30 unrelated English words as Experiment 1.

2.1.4. Procedure

For this and subsequent experiments, before starting the experiment participants were instructed with the help of screenshots taken from different phases of the experiment. In addition to the initial verbal instructions, brief instructions occurred throughout the experiment to remind participants of their task. In the study phase participants were presented with 30 English words. In the recitation phase participants were presented with a blank screen and instructed to recall all of the words they could remember. Specifically, “you will try to recall all of the words to yourself (in your mind). You won’t be able to write down the words or say them aloud. Just practice saying the words in your mind.”

When they could not remember any more words they could study the words again by clicking on a button. After this brief restudy phase, they were to return to the blank screen and practice recalling the words again. They were encouraged to cycle through recalling and restudying as many times as they felt necessary with

the goal of learning as many words as possible. The overall time spent with the material was 5 min for both groups. The 70:30 group studied the words for 3.5 min and used recitation (self-paced restudy) for 1.5 min. The 30:70 group studied the words for 1.5 min and used recitation for 3.5 min. Following the recitation phase all participants played Tetris for 10 min to distract them from rehearsing the items and then took a free recall test (5 min).

2.2. Results and discussion

Both learning groups recalled approximately 60% on the final test (Fig. 3, right panel). The 70:30 group recalled 62.9% ($SD=14.4$) while the 30:70 recalled 59.6% ($SD=22.3$). A *t*-test revealed no significant difference between the two groups on recall on the final test, $t(1,50)=0.637$, $p=0.527$.

On average, the 70:30 group attempted recall a couple times during recitation ($M=2.1$, $SD=1.4$) while the 30:70 attempted recalling the words about 3 times ($M=3.2$, $SD=2.7$). Overall the two learning groups spent 85% and 80% of the 5 min allotted for “study+recitation” phases just studying the material. It appears that when adults are instructed to learn something with the help of retrieval practice, they may spend some time practicing recall of the material, but they tend to focus most of their time in studying the material (Karpicke, Butler, & Roediger, 2009; Karpicke, 2009).

3. Experiments 3 and 4

The results of Experiment 2 suggest that adult learners tend to focus more on studying the material than on practicing recalling it when given the option to do both. In an effort to encourage more switching between covert retrieval and restudy, the change was made to a serial recall test (instead of free recall test), which was what Gates (1917) used. Here, subjects would need to recall as many words as possible and then look back at the study material when they could no longer remember the subsequent items.

Further, Gates’s (1917) results suggested that learners benefited the most from recitation when they learned nonsense material. Therefore, in Experiments 3 and 4 pronounceable nonsense words were used as stimuli, and participants were instructed to recall the words in the same order as they had seen them initially.

In addition to changes mentioned above, a “study only” group (100:0) was added to the experiments, and the recitation conditions were modified to be 80:20 (eighty percent of time in study, twenty percent in recitation) and 20:80 (twenty percent of time in study,

eighty percent in recitation). The two experiments use the same material and procedure except for the interval between recitation and the final test. In Experiment 3 the final test was conducted immediately after the recitation phase, whereas in Experiment 4 the final test was conducted after a delay of 10 min. As will be seen, the motivation for Experiment 4 was to attempt to replicate a hint of an effect seen in Experiment 3, but this attempt failed.

3.1. Method

3.1.1. Participants

The participants were recruited through an advertisement on the campus of Washington University. Participants could opt for either cash or research credit as a compensation for their time (\$5 or 0.5 credit for 30 min of participation).

Eighty-one participants volunteered for Experiment 3. Seventy participants were included in the final data analyses (26 males). The data from 11 subjects were not included as they reported being under 18 years of age ($n = 3$), did not follow the instructions ($n = 2$), reported being confused about the study ($n = 1$), or reported fewer than 10 years of experience with the English language ($n = 2$). Three participants were outliers in terms of age (age > 2 standard deviations from the group mean). The mean age of the participants included in data analyses was 19.5 (range 18–26).

One hundred and twenty-eight participants volunteered for Experiment 4. One hundred and eleven participants were included in the final data analyses (39 males). The data from 17 subjects were not included as they reported being under 18 years of age ($n = 1$), did not follow the instructions ($n = 2$), reported being confused about the study ($n = 3$), or reported fewer than 10 years experience with English language ($n = 8$). Two participants were outliers in terms of age (age > 2 standard deviation of the group mean). Data from one participant could not be used because the subject had participated in a previous study using the same stimuli. The mean age of the participants included in data analyses was 19.7 (range 18–29).

3.1.2. Design

Both Experiments 3 and 4 consisted of three learning groups: study only (100:0), and two recitation groups (80:20, and 20:80). Participants were randomly assigned to one of the learning groups. Because the procedural variations are minor, the materials and procedures are considered here together.

In Experiment 3 there were 25 participants in the 100:0 group, 22 participants in the 80:20 group, and 23 participants in the 20:80 group. All participants went through the 3 phases in the experiment: study, recitation, and a final serial recall test.

In Experiment 4 there were 36 participants in the 100:0 group, 35 participants in 80:20 group, and 40 participants in 20:80 group. All participants went through the 4 phases in the experiment: study, recitation, Tetris, and a final serial recall test.

3.1.3. Material

The nonsense words were generated using The English Lexicon project (Balota et al., 2007). The 20 words included in the experiments were pronounceable 4 letters nonsense words (e.g., FUSP, JENO, CADE, HEAB).

3.1.4. Procedure

Participants were instructed by the experimenter about the various phases of the experiments using screenshots. Brief written instructions were also included in the program to remind them of the task. Participants were informed that they might or might not receive the recitation part of the experiment. Participants were presented with 20 nonsense words during the study phase for about 6 min in the 100:0 group, for 4.8 min in the 80:20 group, and for 1.2 min in the 20:80 group. They were instructed to read through

the list of nonsense words from beginning to the end as many times as they could during the time provided for studying the material. In the recitation phase participants were presented with a blank screen and instructed to recall all of the nonsense words from beginning to the end in the same order as they had seen them (covert serial recall). When they could not remember the next word in the list they were to study the words again by clicking on a button that took them to the study screen. They were instructed to go back to the blank screen and practice recalling the words again after studying the words briefly. They were encouraged to cycle through recalling the words and restudying as many times as they felt necessary with the goal of learning as many words as possible. In Experiment 3 participants took the final test immediately after the recitation stage. In Experiment 4 the final test was given 10 min after the recitation stage. Participants played Tetris during this interval. The final test was presented for 3 min and consisted of blank spaces for subjects to type the nonsense words in the same order as they had studied them (serial recall).

3.2. Results and discussion of Experiment 3

Performance on the final test showed some hint of recitation, although it was not statistically significant. The 20:80 group recalled approximately 11% more than the other two groups (Fig. 4). The groups with a higher proportion of study time recalled 62% on the final test (100:0 $M = 62.5$, $SD = 28.9$; 80:20 $M = 62.7$, $SD = 24.8$) and the 20:80 group recalled 73.5% ($SD = 26.0$). The main effect of group was not significant, however, $F(2,67) = 1.278$, $p = 0.285$.

The recitation phase data showed that participants in the 80:20 group attempted recalling approximately 6 times ($M = 5.7$, $SD = 2.9$), whereas participants in the 20:80 group attempted recalling approximately 12 times ($M = 12.0$, $SD = 12.2$). Overall the 80:20 group spent 88% of the 6 min studying while 20:80 group spent 68% of the allotted time studying the words.

The results of this experiment are the first to show even a small suggestion of a benefit of recitation. Although it was not a statistically significant effect, the observation that the 20:80 group recalled 11% more than the other groups is at least suggestive that this procedure may elicit a benefit of recitation. The primary goal of Experiment 4, then, was to attempt once again; we also made a change in the procedure (i.e., the introduction of a short delay before the final test), in order to enhance the likelihood of being able to observe a reliable benefit of recitation, but the primary goal was to see whether the small (nonsignificant) effect in Experiment 3 might emerge again under similar circumstances.

3.3. Results and discussion of Experiment 4

Experiment 4 differed from Experiment 3 only in delaying the final test by 10 min. During this delay participants played Tetris. The assumption behind this small change was that introducing a delay between learning and recall may reveal differences between the three learning groups that would not otherwise be seen, given that the benefits of retrieval practice sometimes emerge only after a delay (Roediger & Karpicke, 2006b).

Although some hint of differences among the three learning groups occurred on the final test, these differences were not statistically significant and were in the opposite direction than predicted (Fig. 4). The groups with greater study time recalled approximately 60% of the items on the final test (100:0 $M = 57.5$, $SD = 28.2$; 80:20 $M = 59.5$, $SD = 24.2$) whereas the 20:80 recalled only 51.7% ($SD = 23.2$). This main effect was not reliable, however $F(2,108) = 0.982$, $p = 0.378$.

An examination of the recitation phase data shows that participants in 80:20 group did attempt recall fewer times than those in the 20:80 group ($M = 3.7$, $SD = 2.4$ and $M = 10.0$, $SD = 7.5$,

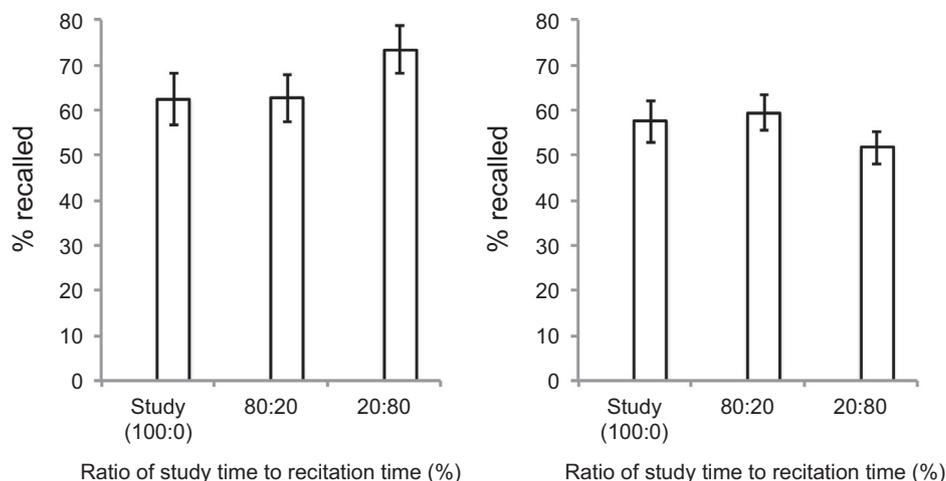


Fig. 4. Mean percent correct performances on the final test in the three learning groups (Experiment 3, left panel; Experiment 4, right panel). Error bars display the standard error of the mean.

respectively). Overall the 80:20 group spent 88% of the 6 min studying while 20:80 group spent 70% of the allotted time studying the words.

In short, although this experiment was designed to complement Experiment 3 by attempting to demonstrate the existence of a hint of an effect in Experiment 3, Experiment 4 demonstrated no benefit of recitation.

4. General discussion

4.1. Summary

The current experiments call into question the generalizability of the often-cited claim that recitation is beneficial for learning (Gates, 1917). Little evidence for the benefit of recitation was seen in the present experiments, which examined this question in an online sample (Experiment 1) and in the laboratory (Experiments 2, 3, 4), using unrelated words (Experiments 1, 2) and nonwords (Experiments 3, 4), using free recall (Experiments 1, 2) and serial recall (Experiments 3, 4). As can be seen in Table 1, across the 4 experiments the mean advantage for the recitation condition was 1% in the wrong direction.

It should be noted that we have not attempted a direct replication of Gates's procedures. As such, we can make no claims regarding the possible replicability of his results. Here we note some of the differences between his procedure and our own and speculate as to the critical difference(s). The fundamental conclusion, though, is that recitation is not an easily obtained effect in adult learners. Unlike other effects (e.g., the testing effect), the benefits of recitation appear to not emerge under a broad range of circumstances.

4.2. Comparison to Gates's (1917) procedures

As noted, there are some clear differences between the procedures we adopted and those used by Gates. Gates tested

Table 1
The mean proportion of items retrieved across the four experiments.

	Mostly	Mostly	Recitation advantage
Experiment 1	.53	.55	-.03
Experiment 2	.60	.63	-.03
Experiment 3	.72	.62	.10
Experiment 4	.52	.60	-.08

grade-school children within their classrooms (although he did report some evidence that recitation benefitted adults, also). He used a within-subjects design, such that children served as their own controls; over weeks in a classroom, they participated in the various conditions of his study. He did report minor effects for what he called "sense" materials (i.e., meaningful materials, such as biographies, see Fig. 1). Our first two experiments used unrelated words, which fall somewhere between nonsense syllables and biographies on this dimension of meaningfulness. Experiments 3 and 4 used nonsense syllables and serial recall, much like Gates. All our retention intervals were quite short relative to those used by Gates (and, admittedly, those typically used in the classroom).

Perhaps the key difference lies in the instructions given to subjects. For the read condition, Gates told his participants:

"To begin with you are to study by reading this list of words over and over from beginning to end. Remember you are to read only. You should never look away from the paper; never close your eyes to see if you can say the words; in fact never say a single word unless you are actually looking at it, actually reading it. Remember you are to read through from the first to the last every time."

This instruction differs quite a bit from the instructions we gave our subjects, which were to study the materials, attempting to learn them. We did not ask them to simply read from beginning to end.

Gates's recitation instructions were more similar to our own than were the read instructions. Here, he told the children:

"After you have read the words through and through in this way for a while, I am going to give you a signal 'Recite'. When I say 'Recite' you are to hold your paper in front of you so that when you are looking ahead, you look over the top of it and you can see it by glancing downwards like this. Now you are to try to say to yourselves as many of the syllables as you can without looking at the card. When you cannot remember the next word look down at your card and then go on saying as many of them as possible without looking. Glance at the card again whenever you cannot remember. Go through the list from the first word to the last in this way and continue until the word 'Time' is given. Remember you are not to look at the words unless you absolutely have to. When the learning period is over I am going to ask you to write as many of these words as you can."

Gates (1917, p. 30)

Our materials were not presented on physical paper, so the exact instructions differed somewhat, but not as markedly as those for

the read condition. Were Gates's subjects simply very poor at learning the materials under the "read" instructions due to the emphasis on rote, serial reading, with no emphasis on attempting to learn the materials? We cannot answer that question with the present data, but that would appear to be one possible explanation for the discrepant results.

4.3. Comparison to studies of covert retrieval

The repeated failure of recitation to elicit a later recall advantage is curious in light of some related findings on covert retrieval. Consider, for example, a recent set of experiments by Putnam and Roediger (2013). Subjects studied pairs of associated words and then took an initial test; the initial test required either typing, speaking, or merely thinking about the answer. Of interest was the boost in later recall performance obtained from that initial test (i.e., the magnitude of the testing effect). Putnam and Roediger found that the covert recall test (merely thinking the answer) produced as much of a benefit as overt cued recall. Similar findings have also been reported by Smith, Roediger, and Karpicke (2013), by Carpenter, Pashler, Wixted, and Vul (2008), and by Izawa (1976), but Whitten and Bjork (1977) concluded the opposite (i.e., that no testing effect occurred for covert rehearsal). Putnam and Roediger characterize the answer to the question of how covert retrieval influences subsequent memory as "murky," and the present results appear not to solve this puzzle. We do note that a discrete recall test (whether covert or overt) differs from the procedures used in the present recitation experiments, in which self-testing is intermixed with restudying in an unregulated, self-paced fashion. This difference in structure may prove important in understanding the differences between conclusions.

Further complicating the picture is a set of experiments by McDaniel, Howard, and Einstein (2009), which showed that a method called 3R (read-recite-review) was superior to simply reading when it came to later memory for text passages. Importantly, "recitation" as defined by this 3R method is different than that used in the present studies (and by Gates). Subjects using the 3R method were asked to read a text passage, then set it aside and attempt to recall as much as possible out loud before reading a second time. That is, following the initial study phase, there a discrete test phase, and then subjects overtly recalled (although they did not write anything down), and then they had a re-study phase. In the present approach, recitation was more of a glancing away process, where the learner turned immediately back to the materials (and then away again) in an iterative process, with the recall being completely covert and self-timed. In short, we do not view our results as being in conflict with the conclusion of McDaniel et al. and note that the definition of "recitation" is probably key to when it will and will not be beneficial to the learner.

5. Practical applications

The present results suggest that all forms of self-testing may not be beneficial for an adult learner. Delineating the critical features that separate this recitation approach from the more typical, beneficial retrieval practice awaits future work. Here we offer a few suggestions for how they might differ. Typically, an overt recall test provides structure in that a discrete, fixed amount of time is dedicated to effortful recall, and the participant is held accountable for their performance (in that the overt recall test involves measurable performance). These procedures might encourage more retrieval effort than a situation in which the learner can look back to the

study list as soon as the retrieval process begins to feel difficult, the case with recitation. That is, during recitation, retrieval opportunities might be underutilized due to their difficulty, the lack of accountability, and the ready option of a restudy opportunity.

Further, overt tests often employ retrieval cues whereby one cue corresponds to a single studied item, which also differs from the recitation approach. The retention interval in the present experiments is shorter than in a typical educational setting, so these results do not speak directly to whether recitation of this type would benefit a learner after (say) a day or a week. Research on the testing effect has demonstrated the importance of retention interval in understanding how initial retrieval influences later retrieval, so this would be an interesting empirical question for the future.

In conclusion, the present results demonstrate some limitations to the benefits of self-testing. It will be interesting to explore this concept further to identify when recitation is beneficial for retrieval and for whom it is beneficial.

Conflict of interest statement

The authors declare that they have no conflict of interest.

Acknowledgements

These experiments were supported by the James S. McDonnell Foundation. We thank Katie Arnold, Adrian Gilmore, Steve Nelson, Mark McDaniel, and Roddy Roediger for comments on earlier versions of this work.

References

- Bacon, F. (2000). *Novum organum* (L. J. M. Silverthorne, Trans.). Cambridge, England: Cambridge University Press (Original work published in 1620).
- Balota, D. A., Yap, M. J., Cortese, M. J., Hutchinson, K. A., Kessler, B., Loftis, B., et al. (2007). The English Lexicon project. *Behavior Research Methods*, 39, 445–459.
- Carpenter, S. K., Pashler, H., Wixted, J. T., & Vul, E. (2008). The effects of tests on learning and forgetting. *Memory & Cognition*, 36, 438–448.
- Gates, A. I. (1917). Recitation as a factor in memorizing. *Archives of Psychology*, 6(40).
- Izawa, C. (1976). Vocalized and silent tests in paired-associate learning. *American Journal of Psychology*, 89, 681–693.
- Karpicke, J. D. (2009). Metacognitive control and strategy selection: Deciding to practice retrieval during learning. *Journal of Experimental Psychology: General*, 138(4), 469–486.
- Karpicke, J. D., Butler, A. C., & Roediger, H. L. (2009). Metacognitive strategies in student learning: Do students practise retrieval when they study on their own? *Memory*, 17(4), 471–479.
- Katzaroff, D. (1908). Le Role de la recitation comme facteur de la memorisation. *Archives de Psychologie*, 7, 1224–1250.
- Kuhn, A. (1914). Über Einprägung durch Lesen und durch Rezitieren. *Zeitschrift für Psychologie*, 68, 396–481. Cited in Gates (1917).
- McDaniel, M. A., Howard, D. C., & Einstein, G. O. (2009). The read-recite-review strategy: Effective and portable. *Psychological Science*, 20, 516–522.
- McDermott, K. B., Arnold, K. M., & Nelson, S. M. (2014). The testing effect. In T. Perfect, & D. S. Lindsay (Eds.), *Sage handbook of applied memory* (pp. 183–200). Sage.
- Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (2004). The University of South Florida free association, rhyme, and word fragment norms. *Behavior Research Methods, Instruments, & Computers*, 36(3), 402–407.
- Putnam, A. L., & Roediger, H. L. (2013). Does response mode affect amount recalled or the magnitude of the testing effect? *Memory & Cognition*, 41, 36–48.
- Roediger, H. L., & Karpicke, J. D. (2006a). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science*, 1(3), 181–210.
- Roediger, H. L., & Karpicke, J. D. (2006b). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science*, 17, 249–255.
- Smith, M. A., Roediger, H. L., & Karpicke, J. D. (2013). Covert retrieval practice benefits retention as much as overt retrieval practice. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 39(6), 1712–1725.
- Thorndike, E. L. (1914). Repetition versus recall in memorizing vocabularies. *Journal of Educational Psychology*, 5, 596–597.
- Whitten, W. B., & Bjork, R. A. (1977). Learning from tests: Effects of spacing. *Journal of Verbal Learning & Verbal Behavior*, 16, 465–478.