Spring 2016

Practical Implications of Learning from Unsuccessful Retrieval Attempts

Ann C. Rossmiller
acr47@zips.uakron.edu

James R. Houston
University of akron, main campus

Practical Implications of Learning from Unsuccessful Retrieval Attempts

Ann Rossmiller

The University of Akron

Abstract

Recent findings suggest that retesting oneself facilitates better learning than studying alone. Building off previous experiments where correcting participants has significantly increased correctness, the current study furthers our understanding about learning from unsuccessful retrieval attempts by manipulating the frequency of correction. Using a set of 42 associated word pairings, each participant was exposed to two blocks where they would memorize the word pairs. This was followed by two quizzing blocks and a final exam block where participants were asked to write down the associate to the stimulus presented on screen. Frequency of correction was manipulated during the quizzing blocks where the participant had their answers checked and corrected by the proctor after every word pair, every third, or every seventh, depending on their condition. There was no correction during the final exam block. After completing the experiment, each participant completed both the Mill-Hill vocabulary scale and the OSPAN task for baseline comparison. Results indicated no significant difference between any of the correction frequency groups but did find that participant's second quizzing and final exam scores increased significantly from the initial quiz score.

Current studies that analyze the effects of studying and testing on learning have found that testing promotes better long term retention of material. In their article, Roediger and Karpicke (2006) review recent findings on this topic and claim that testing not only provides better long-term retention than studying, but that the retesting of material provides greater learning. In fact, recent findings suggest that an effective way to promote learning in schools is to exponentially increase the time students spend retrieving information and reduce the time spent lecturing. This is because self-testing can be a more effective learning mechanism than usual study methods. The effect can be seen in an experiment conducted by Wheeler, Ewers, and Buonanno (2003) where participants were either placed in a study only condition or a repeated test condition in order to learn a list of words. They found that even though studying provided better recall initially, participants in the study only group's retention dropped a week later. However, those in the testing condition had little to no loss of retention between initial recall and recall after one week.

In their study, Knight et al. (2012) expand upon the idea of learning from testing and analyzed whether or not being corrected on a failed retrieval attempt (i.e. an incorrect answer) could facilitate learning. Participants were counterbalanced into one of two stimuli conditions; related word pairs and unrelated word pairs. After participants in each condition were introduced to each word pair individually during a study block, they were asked to recall the paired word corresponding to the word prompted on screen and then were corrected immediately after each task.

Results from this study show that correcting associated word pairs promotes learning but correcting unassociated word pairs actually inhibits learning. In order to rule out study time being the factor causing this, Knight et al. (2012) repeated the task with the same conditions and

extended the study time during the first block so that participants could gain familiarity with unassociated pairs. However, the same effect persisted from the first study showing that study time was not a factor affecting performance. As a result, the authors argued that related word associates are less subject to interference during retrieval compared to unassociated word pairs, leading to learning effects in related, but not unrelated word conditions.

In their discussion section, Knight et al. (2012) mentions how their results lack real world application since they corrected the participant after every trial. My between-subjects study expands on their findings by manipulating the frequency of correction in order to find a more practical learning effect. To recreate the learning effect for associated word pairs, the current study has one condition where a participant is corrected after every trial and will only be using associated word pairs as stimuli. The two new conditions are denoted by providing corrections after every third trial or providing corrections after every seventh trial. Also, as previously explained, repetition facilitates better learning so we can assume that the second quizzing block will have higher accuracy than the first and that the final testing block with have the highest accuracy.

We are most concerned with analyzing the differences between overall score improvement between the three frequency conditions. We expect to see at least one significant difference between the ‘correction every trial’ condition and the ‘correction every seventh trial’ condition based off the study by Wheeler, Ewers, and Buonanno (2003). Their results suggested that repeated study conditions promoted great short term recall but poor long term retention so we would expect that by studying the failed retrieval attempts more frequently, greater short term retention will persist.

Method

Participants

Twenty-two undergraduate students from The University of Akron participated in one of the three conditions of the study. Five extra credit points toward their psychology class was offered in return for their participation.

Materials and Apparatus

E-prime 2.0 Professional was used to create the program that presented the word list on screen during the two quizzing blocks and the questions prompted on screen for the quizzing and final exam blocks. All quizzing and final exam answers were recorded on their respective paper answer sheets provided by the experimenter. Participants were given three *Scientific American* articles to read as distractors between certain sections. The first was given between the second study block and the first quiz. The second was administered between the first and second quiz, and the last, between the second quiz and final exam. Every participant was given the same article in the same order. 42 word pairs selected from a previous study.

Procedure

Because of the study's design, this computer-based task was administered in person, one at a time, with the researcher present. First, the participant was exposed to a study block where each associated word pair was presented on screen for 3 seconds. Next, the participant was exposed to a second study block that presented the same 42 word pairs again. The only difference between the two study blocks was word pair order, which was random across and within participants. Participants were then asked to read over a magazine article immediately following the study block in order to provide a break and discourage rehearsal between sections. After the finishing the article, the participant began the first of two quizzing blocks.

Based off of recent findings discussed in Roediger and Karpicke (2006), we decided that having two quizzing sections would be beneficial because retesting has shown better learning. As previously discussed, the quizzing blocks were where participants were either corrected after every trial, after every third, or after every seventh trial. On screen, the first word appeared from each pair, followed by four question marks, '????', and the participant had to record the corresponding associated word on a provided answer sheet.

When the participant reached the correction screen, the researcher checked over their answers either saying 'Actually,' and then telling them the correct pair, or 'Yes,' and then reiterating the correct pairing. After the participant had finished recalling every pair on the list, the participant was given a second article to read over, acting as another distractor between the two quizzing blocks. Then, the task was repeated, again, for a second quizzing block with the same frequency of correction and the same method of correction. As with the study blocks, word pair order was randomized across participants and across quizzing blocks for each participant. Immediately following the second quizzing block, the participant was asked to read a third article to prevent carryover effects into the final testing block.

During the final testing block, participants had to write down the word associated with the presented stimulus on screen just as they did during the quizzing blocks. However, in this section there were no corrections given so the participant was not aware of how well they are doing.

Scoring

Each quiz and final exam received their own score by summing up of the number of correctly recalled word associates. The maximum possible points earned is 42 per quiz and final exam. To deal with spelling mistakes, the experimenter asked the participant to pronounce the

word associate they had written down and if it was the correct pronunciation, it was counted as correctly recalled (i.e. ‘shiney’ and ‘shiny’).

Results

After the study blocks, participants were able to recall 19.41 word pairings (on average) despite correction frequency. Subsequently, participants had an average recall of 30.64 and 37.59 word pairings for the second quizzing and final exam blocks, respectfully. In order to see if these effects indicated reliable improvement and if this improvement was affected by correction frequency, we performed a factorial mixed ANOVA contrasting correction frequency performance over testing attempt. Correction frequency (1, 3, or 7) was entered as a three-level between-subjects factor and attempt (first quiz, second quiz, final exam) was a three-level within-subjects factor. The main effect of attempt was significant, $F(2,38) = 84.529$, $p < .001$. Significant improvements were identified between all attempts with improvements reached with each additional attempt. The main effect for correction frequency was not significant, $F(2,19) = .673$, $p = .522$. No differences were identified across correction frequencies. Critically, the attempt by correction frequency interaction was not significant, $F(4,38) = .265$, $p = .899$. Thus, improvements made with additional attempts did not depend on the frequency of correction. See Figure 1 for a representation of learning effects across condition.

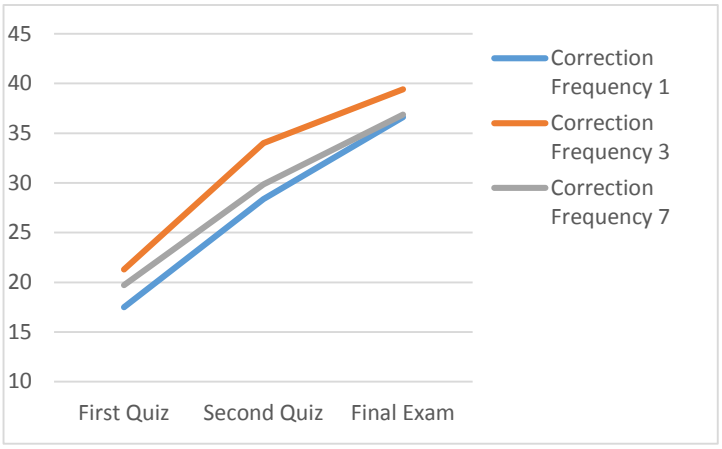

Figure 1. Correctly-identified word pairs across attempt and correction frequency.

Discussion

Regardless of condition, this study found a learning effect which resulted from increased practice and correction. However, our hypothesis that increasing the frequency of correction would promote the highest overall improvement scores failed. In the end, this study found that the frequency of correction had no effect on word-recall abilities.

It may not be that our hypothesis itself was wrong, but that our design was flawed. First and foremost, we had a total of 22 participants which resulted in approximately seven participants per correction frequency. With more data, we may see results that follow a trend that matches our hypothesis.

If one was to replicate this study, we would suggest using only one study block as opposed to two and increasing the associated word pairing list from 42 to a number closer to 60. Most participants were able to recall 21 or more pairings during the first quiz, meaning, they either had too much initial exposure to the pairings, or there were too few pairings so the task was too simple. Also, with more pairings to recall, this will increase variability between

improvement scores and give future attempts a greater difference between frequency of correction conditions. We also suggest to broaden the size difference between frequency conditions from one, three and seven to one, 10 and 20. This may increase mean frequency difference and show a trend where correction after every trial will have the highest scores and correction after every twentieth will have the lowest scores. Most importantly, futures studies so ensure that they include a 'no corrections given' condition where the participant is only presented with the study block after every quizzing section. This is pertinent to reiterate that corrections do improve learning better than studying alone, and to see if all frequency of correction types support this finding.

Researching learning from self-correction and unsuccessful retrieval attempts is important not only to learn about cognition and memory but also to take these results and implement them in the classroom. It is important to build off of the design of this study and use its results as a stepping stone for future research involving learning and education approaches.

References

Knight, J. B., Ball, H. B., Brewer, G.A., DeWitt, M. R., Marsh, R. L. (2012). Testing unsuccessfully: A specification of the underlying mechanisms supporting its influence on retention. *Journal of Memory and Language, 66,* 731-746.

Roediger, H. L., III, & Karpicke, J. D. (2006). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science, 1*(3), 181-210.

Rohrer, D., & Pashler, H. (2010). Recent research on human learning challenges conventional instructional strategies. *Educational Researcher, 39*(5), 406-412.

Wheeler, M. A., Ewers, M., & Buonanno, J. F. (2003). Different rates of forgetting following study versus test trials. *Memory, 11*(6), 571-580.