The Testing Effect and Cramming: Investigating the Presence of the Testing Effect in a Shorter Retention Interval

by

CHRISTIAN JOSEPH M. ARENAS

A THESIS

Presented to the Department of Psychology
and the Robert D. Clark Honors College
in partial fulfillment of the requirements for the degree of
Bachelor of Science

June 2014

An Abstract of the Thesis of

Christian Arenas for the degree of Bachelor of Science
in the Department of Psychology to be taken June 2014

Title: The testing effect and cramming: Investigating the testing effect in a shorter retention interval

Approved: _____

David Nash Unsworth

The presence of the testing effect in a retention interval of 8 hours is investigated in this study. Undergraduate students were told to remember a word list and were either given a test to study or re-read the word list before taking a recall test for these words 8 hours later. There was no difference in word recall accuracy for those who were given a test to study and who re-read the word list to study. The results showed that the testing effect was not present for a retention interval of 8 hours. This null effect is atypical to the robustness of the testing effect found in other studies that have a retention interval of at least 1 day.

Acknowledgements

I would like to thank Professor Nash Unsworth for being my primary advisor, for helping me and guiding me throughout this whole thesis process. Thank you for allowing me to use your lab (and credits) to conduct my study – without it I will have no data for this study. Also, thank you for taking the time to answer my many questions throughout this whole process. I doubt that I could have done this thesis without your help and guidance.

I would like to thank Professor Goldberg and Matthew Robison for taking the time to be part of my Thesis Defense Committee.

Finally, I would like to thank my parents, without whom I would not be here right now. Thank you for your unwavering support throughout my college career. I especially thank you for your continuous encouragement throughout this process, both during the ups and the downs. Maraming Salamat!

Table of Contents

Introduction

"Memory, my dear Cecily, is the diary that we all carry about with us" – Miss Prism in Oscar Wilde's *Importance of Being Earnest*

Memory, as Oscar Wilde puts it, is like a "diary that we all carry about with us." However, unlike a diary, it stores a lot more information and doesn't require you to write in it. So, how do we "write" and retrieve information from this diary? Psychologists think that this is done through the process of memory encoding and recall. Memory encoding is defined as the process of converting information so that it can be stored in the brain (Reisberg, 2013). Once encoded into memory, the process of retrieving this information from memory when it is required is called memory recall (Reisberg, 2013).

Memory encoding and memory recall are integral to many tasks that we do. Take a student taking a test for example. Before taking the test, he encodes the needed information by taking notes in class, listening to his professor, and then studying his notes before his big test. During the test, he recalls the necessary information required by the questions he needs to answer. However, there are times when he will be unable to recall some of the answers, even if he made sure that he re-read his notes before the tests. Why could he not recall all the answers? Was re-reading his notes not enough? Are there any tips or tricks he can do to make this happen less often? These are some of the questions cognitive psychologists are keen to answer.

Psychologists Roediger and Karpicke proposed that the testing effect is a tool that can be used to improve memory recall. The testing effect is the phenomenon that tests enhance later retention of information more than additional study of the material,

even when tests are given without feedback (Roediger & Karpicke, 2006a). Typically, tests are given out in class to check comprehension and to assess the knowledge of a particular material. However, Roediger and Karpicke suggested that tests can be useful as a means of studying as well. In fact, their findings suggested that testing yourself is more useful for improving your memory recall in a later test than re-reading your notes to study. So, the student can use the testing effect and give himself a test while studying, instead of re-reading his notes, to improve his memory recall for his big test.

Research on the testing effect

Two of the earliest studies conducted to investigate the testing effect were done by Gates (1917) and Spitzer (1939). However, it should be noted that at this time the term testing effect was not coined yet but the effect was still observed in their studies. Both studies were large scale studies involving school children that investigated effects of testing (Gates called it recitation) on information retention. Gates (1917) showed that the longer a child does recitation, the better the child performs in a later recall test. Likewise, Spitzer (1939) showed that students who were given a practice test closer to the presentation of information performed better in a recall test (given 63 days later) than those who were given a practice test much later (e.g. 21 days since studying). So, both of these early studies showed that students who were given a test as a way of studying led to better performance in a later memory recall test. Spitzer's study (1939) further suggested that the practice test is more effective when it is given much closer to the initial presentation of information. Thus, both of these studies showed that the testing effect was present, even in early research.

The testing effect is also present in more contemporary studies. Roediger and Karpicke (2006b) presented their participants with two prose passages. After their initial study, participants were told to either re-read the passages, the restudy condition, or they were told to write down as much of the passage as they can remember (a free recall task), a task similar to their final recall task, the testing condition. Then, they were given a free recall task 5 minutes, 2 days or 1 week after their restudy (the retention interval) where they have to recall the passages they have learned. They found that there was no difference between the proportion recalled when the retention interval was 5 minutes. However, for those tested 2 days or 1 week later, participants in the testing condition recalled significantly more than those in the restudy condition. Their results showed that the testing effect is present on quite a complicated task, the recall of prose passages.

Additionally, the testing effect has been observed both in the lab and in the real world. First, Agarwal, Karpicke, Kang, Roediger, & McDermott (2008) showed that students performed better in a memory recall task when they either did a closed-book or open-book practice test when compared to students who only studied without testing. Next, Carpenter, Pashler, Wixted, &Vul (2008) showed that participants performed better on a memory recall task of Swahili-English word pairs and obscure facts when they tested themselves as a form of memory rehearsal – a testing effect. Moreover, the testing effect was present for 8th graders that were given a memory recall task for US history facts (Carpenter, Pashler, & Cepeda, 2009). Also, the testing effect was found for college students learning linear functions (Kang, McDaniel, & Pashler, 2011). Additionally, the testing effect was present when it was investigated in an online college

course (McDaniel, Anderson, Derbish, & Morrisette, 2007) and in an actual cognitive psychology course for college students (Hattikudur, & Postle, 2011). Finally, Meyer & Logan (2013) showed that the testing effect was present for people of many ages as they found it in university students, young adults that were not university students, and middle-aged to older adults. In all, it can be seen that the testing effect is present in a variety of situations, people, and locations and not just in the lab.

Explaining the testing effect

Although we have established the robustness of the testing effect, little is known how or why it works. Most research has focused on the presence of the testing effect on various items and situations. But, not many have tried to determine how it improves a person's performance on a memory task. So, I would like to propose some theories that may help explain how the testing effect works.

First, the testing effect may affect the person's attentional control during the encoding and recall process. Specifically, the testing effect reduces mind wandering during the studying of the material (encoding). Unsworth, McMillan, Brewer, & Spillers (2012) showed that mind wandering seem to occur more frequently in class or while studying (educational contexts). So, Szpunar, Moulton, & Schacter (2013) in their review suggested that interspersing periods of instructions with some quizzing can promote student attention and thus reduce mind-wandering. Although Szpunar et al.'s (2013) review does not relate specifically to the testing effect, there is an interesting parallel that can be made between their review and the testing effect. They suggested that quizzes be included during the encoding of information which is very similar to the testing effect. This is similar because for the testing effect to be present, the information

is first given to the person, then they are given a test as a way of studying during encoding. Thus, based on Szpunar et al.'s (2013) review, when a person uses a test to study he is more likely to pay attention to what he is learning and encode the information better which will help him in the later memory recall test later (such as a test).

Second, the testing effect may improve a person's performance on a memory recall task because of the context dependent memory hypothesis. According to this hypothesis, people perform better on memory recall tasks if their context (e.g. location) during recall is the same as their context during encoding. This is best shown by Baddeley & Godden's (1975) classic study on context dependent memory. In their study, divers had to learn a list of words either on land or under water. Then, they were given a recall test on these words either on land or under water. Baddeley & Godden (1975) found that divers who were tested in the same environment where they learned the words (e.g. learned on land, tested on land; learned underwater, tested underwater) performed better in the recall task than those who learned and were tested in two different environments (e.g. learned on land, tested underwater). Their results showed a clear context dependent memory. So, this is relevant to the testing effect because using a test to encode information, especially if this test is similar to the later memory recall test, provides a similar context for both the encoding and retrieval processes which is very much in line with the context dependent memory hypothesis.

The current study

Based on the results of the studies investigating the testing effect, the testing effect has some interesting implications on education and learning. First, it provides a

different perspective on testing, where instead of originally only being a way to measure the information a person can recall from their memory, it can also act as a way to learn and remember this information for future recall tests. Also, it provides educators and students another useful tool to help them teach and study a wide variety of information whether it is word pairs, prose passages, and even math functions. Thus, it is a worthy endeavor to further investigate the testing effect because of its perceived benefits to teachers and students alike.

As previously mentioned, the studies investigating the testing effect showed the testing effect is fairly robust. However, it should be noted that in all of these studies, the researchers investigated the testing effect with a retention interval of at least one day or more. The retention interval is defined as the time between the initial presentation of information (encoding) and the memory recall test for this information (retrieval). So, for a retention interval of one day, the participant is given a memory recall task one day after the information is presented. Thus, a shorter retention interval that is less than a day has not yet been thoroughly investigated. So, this study seeks to further test the robustness of the testing effect and see if it is still present with a much shorter retention interval of eight hours.

This study may be of further benefit to a student as it can be connected to cramming. For this study, cramming is defined as the act of studying for an exam by memorizing facts at the last minute. Thus, by definition cramming is learning something in a short retention interval because you study the necessary information quite close to the memory recall test. So, there are parallels with cramming and the short retention interval of 8 hours that is being investigated for the testing effect by this study.

Moreover, when a student is cramming it is less likely that the student is studying for the test a few days before it, but it is more likely that they are studying for their test on the day of the test and with only a few hours to spare, which fits well with the 8 hour retention interval that will be investigated in this study. As every student can attest, they have crammed for a test at least once in their career as a student. So, investigating if the testing effect is present in a short retention interval may prove useful to students because it may be a useful tool for studying when the student has no choice but to cram for their test. Although, it should be mentioned that this study does not promote cramming, as studying for a test well in advance is definitely more helpful than cramming. Instead, it seeks to see if the student has no choice but to cram, will the testing effect be a useful way of studying.

Thus, the aim of this study is to see if the testing effect is present in a shorter retention interval than what was studied in the existing body of research. Specifically, it aims to see if performance on a memory recall task, with an 8 hour retention interval, will improve if participants are given a test to study (the testing effect) when compared to those who were told to just re-read the information. The dependent variable for this study is the performance on the memory recall test given 8 hours after the initial presentation of the information (the retention interval). This is operationalized by giving the participants a word list to remember and study, and then giving them a free recall task to test how many words from the list are correctly remembered. The more words remembered correctly (a higher accuracy), the better their memory recall is. Although a free recall task of a word list seems artificial, the use of this free recall task is still useful because it provides an easy test to see the presence of the testing effect. Thus, it is a

good task for a proof of concept, and once the presence of the testing effect is determined with the free recall task, it can be extended and investigated with more realistic tests that mimic tests found in classes. The independent variable for this study is the presence of the testing effect, which is done by making the subject either re-read the word list to study (the study condition – no testing effect) or do a free recall task to study (the test condition), both of these methods of studying will be done by the participant immediately after the word list is initially presented.

For this study, an independent samples t-test will be conducted to investigate the presence of the testing effect. So, the primary hypothesis for this study is that participants who used a test to study will perform better in the later memory recall test compared to participants who just re-read the word list to study. This is in line with the findings of previous studies which suggest that the testing effect is present, albeit in this case in a much shorter retention interval. The alternate hypothesis is that there will be no difference in performance in the memory recall test between the participants who used a test to study and those who re-read the list to study. This suggests that the testing effect is not present in the shorter retention interval as testing yourself did not improve memory recall performance.

Method

Participants

A total of 32 University of Oregon students participated in the study (16 for each condition). The majority of the participants (n = 17) were recruited through the subject pool in exchange for course credit. While the rest are a convenience sample (n = 15), with six coming from the Clark Honors College. 11 were male participants and 21 were female. Participant ages ranged from 18 to 31 years old (M = 21.03, SD = 3.15).

Procedure

Participants arrive in the lab at 9 am for the first part of the study. Then, they were randomly assigned either to the test or study condition. All participants were told that they will be shown a list of words and that their task was to remember the words as best as they can for a later test. The list is comprised of 40 nouns selected from the Toronto word pool (the same word pool from Unsworth, Spillers, & Brewer (2012) was used). The whole task is done on a computer in the lab. Before beginning, all participants were given a practice test to familiarize themselves with the task. Then, the actual task began. First, all participants were shown the word list. Each word was shown for 3 seconds, and then a blank screen appears for 0.75 seconds before the next word is shown. So, in total the whole 40 word list is shown for around 150 seconds.

Then, participants in the study condition were told that they will be shown the word list again and to try and do their best to remember these words for a later test. As before, each word was shown for 3 seconds, and then a blank screen for 0.75 seconds before the next word is shown. The whole list is shown for a total of 150 seconds.

While, the participants in the test condition were given a free recall task that was exactly the same as their later memory recall task and were told to recall as many of the words as they can remember from the list, after the word list was presented to them. During recall, participants were shown three question marks (???) indicating that they can now begin the free recall task. They type in as many words as they can remember from the original list. They type one word at a time, and then pressed Enter after each response. The free recall task is done after 150 seconds, the same amount of time as the re-study time in the study condition. After the study period for each condition was finished (i.e. 150 seconds after), the participants were told to come back to the lab at 5 pm, 8 hours later (the retention interval) for the second part of the study where they will be given a memory recall test.

For the second part of the study, all participants returned at 5 pm (8 hours later) and were assigned the same room and computer they used in the first part. Then, they were told that they will be given a free recall task and to recall as many of the words as they can remember from the word list presented in the first part of the study. This free recall task was exactly the same as the one given to the participants in the testing condition, but for a longer time. Each participant was given 10 minutes to recall as many words from the list as possible. They were shown three question marks (???) in the middle of the screen to indicate that the free recall task has begun. They type in one response at a time, and press enter after each response. They were given ten minutes to recall in order to reach asymptotic recall. After ten minutes, they were told that the task is done and were thanked for their participation and debriefed.

Results

The primary hypothesis for this study is that participants who were given a free recall task as a way to study will remember more correct words (a higher accuracy) in a later free recall test of memory than those participants who were just presented the words again to study. The alternate hypothesis is that there will be no difference between the number of correct words remembered for those given a test to study or re-read the word list (same accuracy).

In this study, an independent samples t-test was conducted to test these hypotheses. An independent samples t-test revealed that the number of correct words remembered from the list (accuracy) by participants who used a test to study (M = 10.76, SD = 4.96) is not significantly more than the number of correct words remembered from the list (accuracy) by participants who re-read the word list (M = 13.00, SD = 7.88), $t(30) = 0.972$, $p = 0.339$. These results suggest that there is no difference in accuracy between the two groups of participants. Specifically, the results suggest that using a test to study does not improve the accuracy of words remembered.

Discussion

In this study, it was predicted that participants who were given a test to study (the test condition) will have a higher accuracy compared to participants who re-read the word list to study (the study condition). However, the results of the study showed that there is no significant difference between the accuracy of the two groups. Thus, the primary hypothesis is not supported, and the alternate hypothesis is accepted. In fact, the study group had a slightly higher mean accuracy compared to the test group, but this difference was not significant. Therefore, we can conclude that there is no testing effect for a retention interval of 8 hours.

Interestingly, the findings of this study go against the typical findings of most studies investigating the testing effect. Unlike the studies mentioned before which investigated the presence of the testing effect in a variety of situations (see Roediger & Karpicke, 2006b; Carpenter et al., 2008 for example), this study found no testing effect in a shorter retention interval. So, even though the testing effect is fairly robust, it is not always present in every situation. Although the testing effect is useful tool for a student to use when studying, it does not seem to be useful if they have no choice but to cram (a shorter retention interval).

One possible explanation for the null effect found in this study is the number of participants in the study. Sixteen participants per condition (for a total of 32 participants) is a fairly small sample size for a study. Also, only seventeen of the 32 participants came from the subject pool and the rest are from a convenience sample which included participants from the Honors College. So, 46% of the participants may not be fully representative of the norm. Moreover, a small sample size is more sensitive

to outliers in the data. It should be noted that this study has a sample size of 32 students and took a small convenience sample is because of time and logistical constraints.

Looking closely at the data, the standard deviation (*SD*) for the study condition is 7.88, a fairly high number which suggests that there are quite a few outliers in this data set. When the word recall accuracy of participants in the study condition is graphed with a normal curve (see Fig. 1), we can see that the accuracy is not normally distributed and slightly skewed to the higher end of the scale. This suggests that the higher mean accuracy of the study group is probably caused by these outliers. Upon closer inspection, three out of the four big outliers were randomly assigned to the study condition. When these three outliers are taken out of the data set for the study condition, the mean accuracy of the test condition actually becomes slightly higher than the study condition, but not significantly so. This suggests that there might be a testing effect for a shorter retention interval, if there were no outliers. However, it is not feasible to take out these outliers because having only 29 participants and 12 participants for the study condition is an incredibly small sample size. So, because it is such a sample size, these three outliers may have skewed the mean accuracy of the study group causing it to have a slightly higher mean accuracy than the test group.

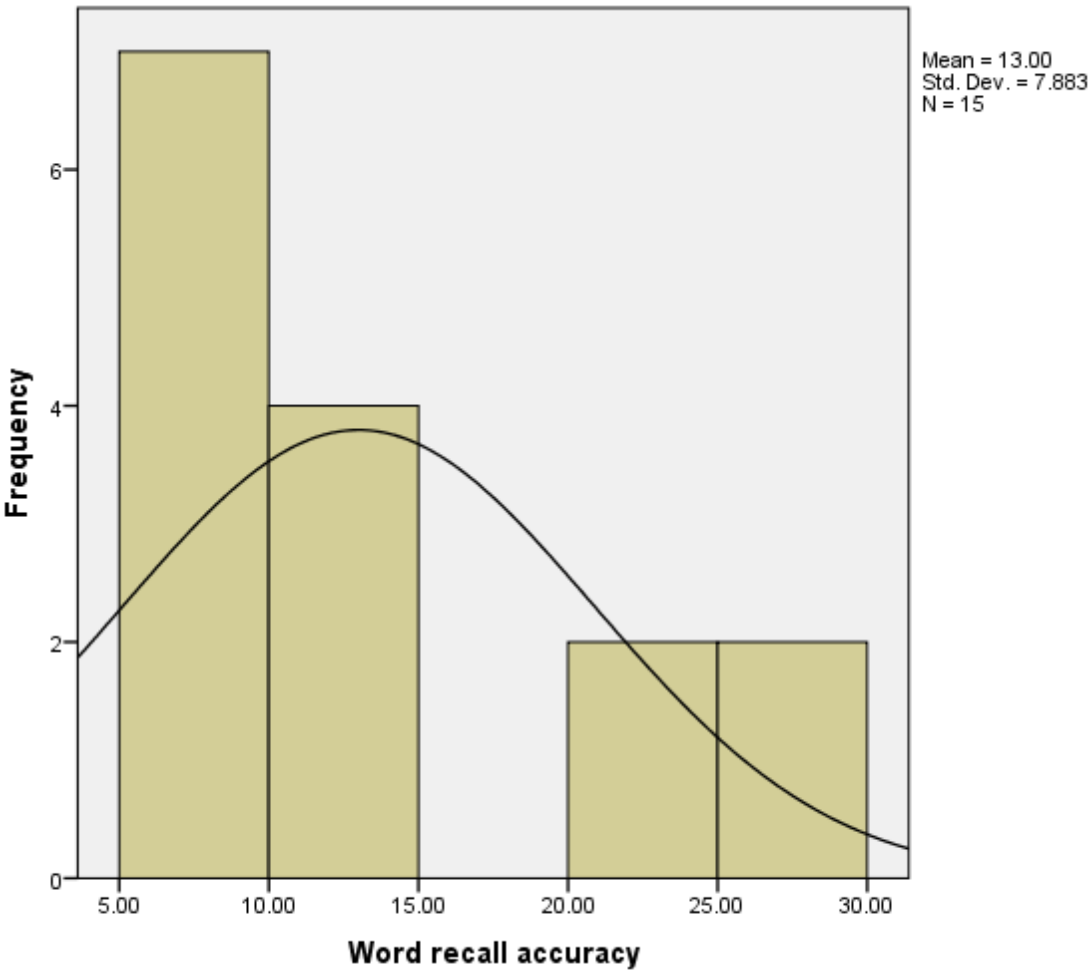

Figure 1 – Histogram of word recall accuracy for study condition with a normal curve

Also, it is possible that there really is no testing effect for a retention interval of at least 8 hours. But, this was not discovered during my research of previous studies because it is unlikely researchers publish a study with a null effect. So, it is possible that previous research has been done investigating the testing effect but it was not published and thus not known.

There are a variety of possible extensions for this study. First, it would be very useful to replicate this study but with a much bigger sample size and without the use of a convenience sample. As pointed out earlier, if the outliers were taken out there seems

to be a slight testing effect. So, with a bigger sample size the data will be less sensitive to outliers and may provide a clearer answer to the question of whether or not the testing effect is present in a shorter retention interval.

Another possible extension for this study is perhaps to do a slightly longer retention interval but making this interval still less than one day. This may be useful because this study has established that the testing effect is not present in a retention interval of 8 hours or less, but it may be present in a later retention interval. So, by doing this a researcher may be able to determine the threshold when the testing effect will be a useful tool for studying. Finally, if a testing effect was indeed found for a retention interval of less than a day, then the robustness of this testing effect can be determined by running it through a variety of situations such as Roediger and Karpicke's (2006b) prose passage, Kang et al.'s (2011) math functions or Hattikudur & Postle's (2011) cognitive psychology class.

Bibliography

Agarwal, P. K., Karpicke, J. D., Kang, S. H. K., Roediger, H. L., & McDermott, K. B. (2008). Examining the testing effect with open- and closed-book tests. *Applied Cognitive Psychology, 22,* 861-876.

Baddeley, A. D., & Godden, D. R. (1975). Context dependent memory in two natural environments: On land and underwater. *British Journal of Psychology, 66,* 325-331.

Carpenter, S. K., Pashler, H., & Cepeda, N. J. (2009). Using tests to enhance 8th grade students' retention of U.S. history facts. *Applied Cognitive Psychology, 22,* 760-771.

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).

Hattikudur, S., & Postle, B. R. (2011). Effects of test-enhanced learning in a cognitive psychology course. *Journal of Behavioral and Neuroscience Research, 9,* 151-157.

Kang, S. H. K., McDaniel, M. A., & Pashler, H. (2011). Effects of testing on learning of functions. *Psychonomic Bulletin Review, 18,* 998-1005.

McDaniel, M. A., Anderson, J. L., Derbish, M. H., & Morrisette, N. (2007). Testing the testing effect in the classroom. *European Journal of Cognitive Psychology, 19,* 494-513.

Meyer, A. N. D., & Logan, J. M. (2013). Taking the testing effect beyond the college freshman: Benefits for lifelong learning. *Psychology and Aging, 28,* 142-147.

Reisberg, D. (2013). *Cognition: Exploring the science of the mind*. New York, NY: W.W. Norton & Company.

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

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

Spitzer, H. F. (1939). Studies in retention. *Journal of Educational Psychology, 30,* 641-656.

Szpunar, K. K., Moulton, S. T., & Schacter, D. L. (2013). Mind wandering and education: From the classroom to online learning. *Frontiers in Psychology, 4*, 1-7.

Unsworth, N., McMillan, B. D., Brewer, G. A., & Spillers, G. J. (2012). Everyday attention failures: An individual differences investigation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38*, 1765-1772.

Unsworth, N., Spillers, G. J., & Brewer, G. A. (2012). Dynamics of context-dependent recall: An examination of internal and external context change. *Journal of Memory and Language, 66*, 1-16.