



## Learning from Text with Spaced Remembering Practice

Rebecca Finney

## Learning from Text with Spaced Remembering Practice

**ABSTRACT**

The aim of the experiment was to determine the best ways for students to learn from text with spaced remembering practice. The within-participants design involved university students aged 18-22. Participants read a chapter in an initial study session and were tested on half the idea units immediately. The text was then split up into four sections and a condition was assigned to each part: expanding practice, uniformly spaced practice, single early test and single delayed test. Over the next 15 days, on days 2, 5, 7, 10 and 15, spaced online study sessions were completed which tested recall of the text and gave feedback. Four weeks later participants completed a short answer test to measure long-term recall of the whole text. More tests led to better long-term recall, but the effects of the time of the initial test and whether material had been immediately tested were non-significant. A single retrieval practice is more beneficial if taken early and an immediate test takes place after initial presentation. Conclusions were drawn that multiple retrieval attempts will increase long-term recall more than a single retrieval practice but one type of spacing of practice tests (either expanding or uniform) is not superior.

<b>KEY WORDS:</b>	<b>RETRIEVAL PRACTICE</b>	<b>LEARNING FROM TEXT</b>	<b>MEMORY</b>	<b>LEARNING FROM TESTS</b>
-------------------	---------------------------	---------------------------	---------------	----------------------------

## Introduction

### Learning from Text with Spaced Remembering Practice

Ebbinghaus (1885) was the first to experimentally demonstrate that rehearsal decreases forgetting. Following on from this, Thorndike's 'law of disuse' (1914) proposed that without rehearsal or revisiting, the association between the cue and the target (i.e. the thing to be remembered) would weaken. Bjork and Bjork (1992) expanded on this by highlighting the important distinction between the retrieval strength and the storage strength of a memory representation. Retrieval strength refers to how accessible the memory is at a certain time, often influenced by environmental cues, whereas storage strength describes how well established the memory is with related representations. If a memory has higher storage strength then the rate of decay, or weakening of the association, is slowed down. On the other hand, if something has low storage strength, such as newly presented information, then the rate of decay will be quite fast.

Bjork and Bjork (1992) propose that recall is determined by retrieval strength and storage strength has no direct effect on recall performance. Therefore, it is wrong to view retrieval failure as a weakness of the whole memory system. In fact, Bjork and Bjork (1992) suggest that losing retrieval access to information which is disused is actually an adaptive process due to the enormous storage capacity of long-term memory. For instance, when asked to give a current telephone number, it would not be useful for an individual to recall every phone number they have ever had. Therefore, decay of old phone numbers, which are no longer used, is part of an adaptive process, which prioritises and speeds up retrieval of current information. Furthermore, disused information is still stored and accessible with effortful retrieval if required in the future.

However, although there may be no limit of storage capacity, retrieval capacity may be limited due to several factors. For instance, following subsequent study as retrieval strength for newly studied items increases, there will be a corresponding loss in retrieval strength across previously studied items as effortful processing focuses on the new information. In addition, research has indicated that the ability to remember information may be largely cue dependent, with levels of recall increasing if the same cues are present as at initial encoding. For instance, Smith (1986) found that student's memory for information depended not only on study cues but also in the classroom they studied. Therefore, multiple opportunities for rehearsal in different contexts may help to reduce these types of restraints and strengthen retrieval capacity, as information is repeatedly processed and recall does not rely on one particular cue. For instance, Bjork and Bjork (1992) report that rehearsal in the form of successful recall of an item or simply re-studying can increase retrieval.

This finding that rehearsal is crucial to successful recall is crucial to efficient educational practice. Students are required to continuously take in information and are then expected to recall a large amount of learnt material after a prolonged time. In recent years, there has been increased interest in study skills and the best ways for student to learn and retain information (e.g. Hadwin & Winne, 2009). For instance, research such as Dempster (1997) has indicated that practice testing can improve later recall of material. Consequently, there has been much discussion and

further empirical work into the best ways for educators to give tests and for students to use testing as an aid to study. Much of this research has focused on the benefits of distributed compared to massed practice, the optimum delay between initial study and the initial practice session and the type of spaced practice schedule over a prolonged period.

### **Testing and Generation Effect**

Research into the effects of tests on learning has implied that tests do more than simply test; they promote effective learning through a process of effortful retrieval. Fritz (2011) defines testing and generation effects as 'a benefit in later remembering as a result of producing an answer to a demand at an earlier time' (p. 200). Over the last few years, applied cognitive psychology has shown an increasing focus on the testing effect, although it has had a presence in psychological research for a long time. James (1890) suggested that the benefits of testing are based on the value of active repetition; effective learning happens through the practice of effortful retrieval. Similarly Mace (1932) states that 'active repetition is very much more effective than passive repetition' (p.39). For instance when studying for an exam, active repetition may refer to an attempt to recall information without direct reference to material, whereas passive repetition may simply involve re-reading given information. Abbott (1909) supported this theory, reporting that correctly recalling items on an intervening test led to generally better recall on a delayed test than if the items had just been restudied. Consistent findings have been replicated with a range of tests (e.g. free-recall, short answer, free recall etc.) and an array of learning materials (e.g. word lists, texts) many times since this early empirical work (e.g., Carrier and Pashler, 1992; Karpicke and Roediger, 2008; Spitzer, 1939).

There have been two hypotheses to explain the testing effect, as discussed by Dempster (1997). Amount of Processing Hypothesis is one theory that has been offered to explain the benefits of intervening tests. This proposes that tests act as simply additional study sessions by allowing more time for the information to be processed before the final test. However, the superiority of practice tests over extra passive study sessions suggests that the beneficial effects of testing are not just because of multiple presentations. Alternately, the Retrieval Hypothesis suggests that it is the actual process of retrieval, which accounts for the positive effects of tests, rather than just the time spent on processing the information. Bjork (1975) explains this in terms of Craik and Lockhart's level of processing theory (1972). As with encoding, Bjork (1992) proposes that retrieval under conditions that require little effort can be detrimental to long term retention due to only a surface level of processing. On the other hand, retrieval of information, which involves complex processes, and higher levels of engagement will lead to easier retrieval in the long term. This is illustrated by Glover (1989), who found that cued recall tests had a smaller facilitative influence on later recall than a free recall test, but a greater influence than a recognition test.

The process of generation has also been linked with effective practice. For instance, Chi, de Leeuw, Chiu and Lavancher (1994) reported that students who spontaneously generated self-explanations of given examples were the most successful at solving problems in a later test. In contrast to re-reading, practice tests require generation and therefore deeper levels of processing and a greater memory

for information. Research applying testing and generation effects to educational settings has supported these claims. Jones (1923) reported that scores were 50% higher on criterion tests occurring three days to eight weeks later when practice tests immediately followed a lecture. More recently, Fritz and Morris (2003) obtained similar results by embedding practice tests in a second year university statistics course. At the end of each lecture, students were given either a short answer test or a review of the key points. Results showed that having a test was more beneficial than just reviewing a list of key points.

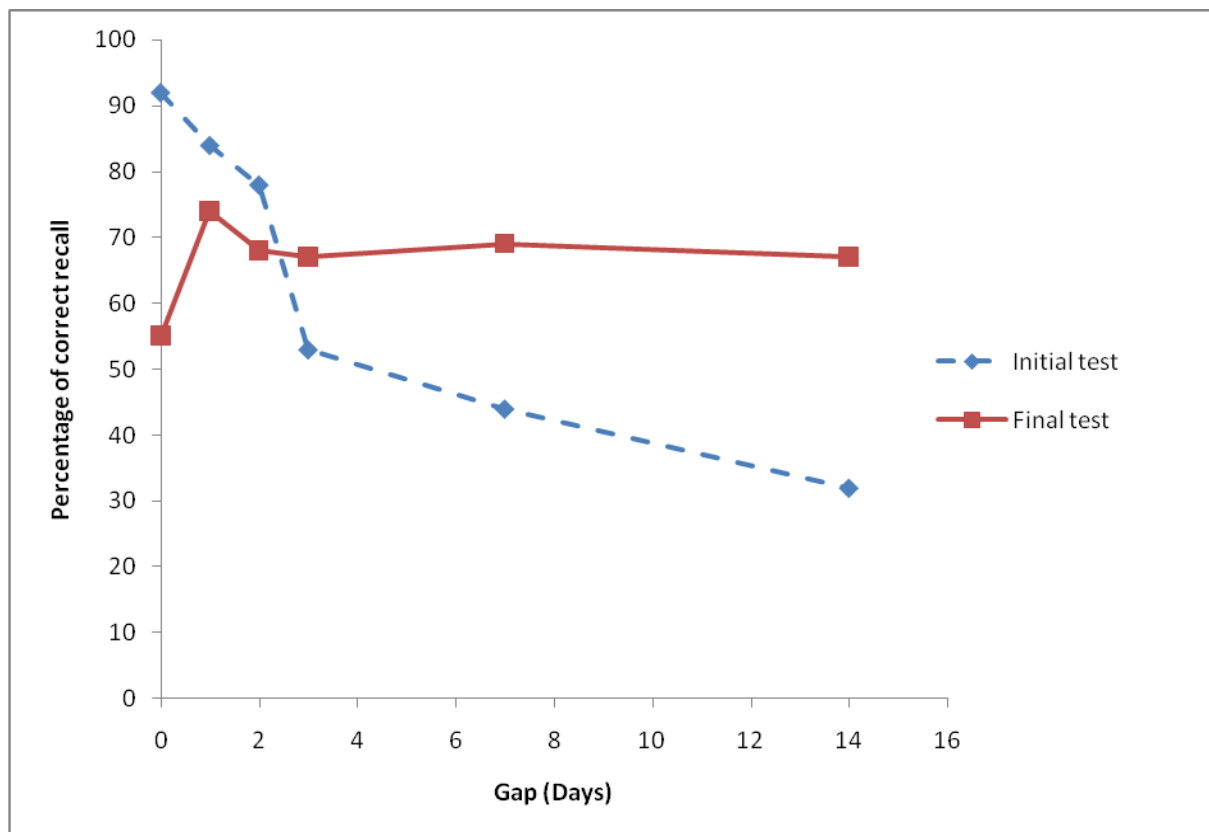
### **Spacing Effect (Distributed vs. Massed Practice)**

In light of the large amount of empirical work supporting testing effects, there have been criticisms that educators are not fully recognising the significant effect testing can have on the learning process (e.g., Bjork, 1975; Dempster, 1997). However, there is much debate as to how tests should actually be delivered: Should pupils be tested on everything once at the end of each term (massed) or should they have a series of shorter tests spaced throughout the term (distributed)? In experiments investigating the benefits of massed versus distributed practice, the same amount of time is given to rehearse the information, it is just presented differently. For instance, a text may be assigned for half of the participants to review in a single session, whereas for the other half this time may be divided into several smaller review sessions. Despite massed and distributed items receiving the same amount of study time, spaced items are generally better recalled than massed items. The spacing effect (e.g., Delaney, Vekoeijen and Spigel, 2010) refers to the finding that memory performance is superior when practice sessions are separated by other events or items (distributed practice) than when practice sessions follow immediately after one another (massed practice). Ebbinghaus (1885) recognised the positive effects of distributing practice compared to massed practice and these original findings have been repeatedly confirmed and extended since. Reviews by Dempster (1997) and Raaijmakers (2003) have indicated that the spacing effect occurs with a large range of testing procedures, materials and participant populations.

The Spacing Effect has been applied to educational practice, with many academics supporting learners spacing their study practice. For instance, Rohrer and Taylor (2007) investigated distributed learning in terms of shuffling mathematics problems in textbooks. Following mastery of solving one kind of problem, college students were given subsequent practice problems in either a massed format in a single session (as do standard textbooks) or spaced across multiple sessions. Rohrer and Taylor (2007) reported that those receiving spaced practice problems performed better on a test given 1 week later. Nevertheless, most distributed practice studies have taken place over a very short period of time, with a spacing gap of seconds and minutes rather than days, hours or months. Therefore, there is little valid evidence to apply the spacing effect to real-life contexts and educational practice. However, research such as Karpicke and Roediger (2010) and Cepeda, Coburn, Rohrer, Wixted, Mozer & Pashler (2009) has recognised this hole in the literature and conducted several experiments with practice sessions over a period of days and test delays weeks and in the case of the latter up to 6 months later. Both experiments reported benefits of spacing study and practice sessions.

## Time of initial learning

Research with varied schedules of testing has suggested that certain positioning of practice tests may be more beneficial than others. Early work by Spitzer (1939) investigated whether the time of the first test had an effect on learning from text with primary school aged children. Each child had to read a factual article and then was tested at various intervals one or more times. It was reported that if the initial test occurred 1 and 7 days after reading final test scores were 15 to 30 percent higher than if the initial test took place after 14 or 21 days. Karpicke and Roediger (2007) found similar advantages to an early first test, reporting enhanced short-term accessibility during the learning phase on an immediate test. However if the first test was delayed Karpicke and Roediger (2007) reported that recall was higher on a test 2 days later than the early initial test condition. Examining the effect of the spacing gap on recall of foreign words, facts and names, Cepeda et al. (2009) reported an optimal gap improved final recall by 150%. The pattern of results from their first experiment is shown in Figure 1. Final test marks showed a clear benefit of spacing practice, with an optimal gap of 1 day. As the gap then increased from a single day, final test marks began to show a relatively steady decline.



**Figure 1: How retention is affected as gap is increased from 5 minutes to 14 days for a test delay of ten days**

Bjork (1999) suggested that key to effective learning is 'desirable difficulty'. Although some conditions may produce weak learning to begin with, long-term retention is often greater. Therefore, initial performance is not necessarily reflective of learning over time. For instance, an immediate massed test may be more beneficial in the short-term, but spaced repetitions are more desirable for delayed recall. Bjork (1999)

suggests that initial difficulty leads to more effortful learning overall and therefore is more beneficial in the long-term. Consequently, early initial practice may be more beneficial if learning is to be tested soon after practice, but if recall is required after a longer period of time a delayed initial practice test is most effective. However if the level of difficulty is set too high, that is, if the initial test is too delayed, recall will decrease as forgetting occurs before initial learning takes place.

### **Types of Spacing/Distributed Practice**

Other research has investigated the most beneficial spacing practicing schedule when there is more than one practice test. This area of research has aimed to determine the optimum recall delay between each practice test. Landauer and Bjork (1978) suggested expanding retrieval practice is the best way to space practice. This involves retrieving information soon after study and then gradually increasing the spacing between repeated practice sessions. The success of expanding spaced schedules is based on a combination of retrieval success and retrieval difficulty, due to the number of tests, the time of the initial test and the delays between each practice session. An early test promotes retrieval success by consolidating learning and gradually increasing the space between practices increases retrieval difficulty, replicating the benefits of a delayed test. Landauer and Bjork (1978) reported that in various tasks of remembering names the expanding spaced schedule produced significantly higher final test results and therefore an overall greater level of learning than uniform spaced schedules.

However since then there has been a limited amount of research, which directly compares the effectiveness of expanding and uniform spaced schedules. In a review by Balota, Duchek and Logan (2007) it was reported that in a lot of studies it was not clear whether it was the particular spacing schedule of testing, which had positive effects on retention, or whether it was repeated testing or spaced testing. A large amount of empirical work has tended to focus on the benefits of expanding spaced schedules over massed retrieval practices. Rea and Modigliani (1985) investigated the effects of expanding test schedules with schoolchildren as they learned multiplications and spellings. However the control condition was massed testing, not equally spaced testing and therefore their study showed a general superiority of spacing rather than specifically of expanding spaced practice. Similarly other research has compared expanding testing schedules to simply expanding study sessions. Conclusions were drawn that expanding schedules were better when they involved tests rather than re-study sessions, yet there was no type of uniform spaced condition. For instance, Morris and Fritz (2000, 2002) investigated a technique for learning the names of a medium sized group of people based on an application of expanding retrieval practice. Greater recall of people's names was reported following the 'name game' compared to the no-retrieval condition, after delays of 30 minutes and 2 weeks. However this simply demonstrates the presence of a testing effect, highlighting the benefits of active recall rather than demonstrating a specific benefit of expanding retrieval practice.

Yet in 1992, Shaughnessy and Zechmeister were able to replicate the findings of Landauer and Bjork (1978), demonstrating a small but positive effect of expanding over uniform retrieval schedules shortly after presentation. However generally empirical evidence investigating the effectiveness of different types of spaced

practice schedules has yielded mixed results. In a series of follow up experiments Cull, Shaughnessy and Zecmeister (1996) presented mixed results of the superiority of expanding schedules. In only 2 out of the 5 experiments a significant positive effect of expanding over uniformly spaced practice was reported. Furthermore, Carpenter and Delosh (2005) reported that equally spaced schedules were in fact better than expanding. Support for the superiority of uniformly spaced practice schedules have tended to be reported where long-term retention is assessed after a significant delay. For instance Cull (2000) found there to be greater recall in the uniform condition after a delay of 3 and 8 days.

### **Learning from text**

Although there has been a lot of research in this area, it has tended to focus on word pairs and remembering short chunks of information. Although this type of study can be applied to learning a foreign language (e.g. Morris & Fritz, 2006), on the whole there is limited work focusing on learning from longer texts. Furthermore, although some recent studies have investigated the effect of spaced rereading (Rawson & Kintsch, 2005) and retrieval practice (Roediger & Karpicke, 2006) there has been a lack of research exploring how different types of spacing practice influence learning from text.

However Karpicke and Roediger (2010) recognised this gap in empirical work and investigated the effects of varied spaced practice on learning from text. In experiment 1 undergraduates read two short text passages displayed on a computer screen. In this same session participants practiced remembering the material three times without feedback. Between the practices participants completed a filler task. How the practice recall tests were spaced in the initial learning session was manipulated with four conditions: immediate-expanding, immediate-equal, delayed-expanding and delayed-equal. In a free recall test one week later those who completed an immediate initial recall test performed better than those who had taken a delayed initial test. However there was no reported significant effect for the actual spacing of practice tests.

Experiment 2 sought to replicate the results of Experiment 1 but correct several of its problems. For instance, Karpicke and Roediger (2010) identified a need to have a more distinct separation of the position of the first test from the spacing of the repeated test. Consequently, in experiment 2 the position of the first test and the spacing of practice schedules were factorially crossed so that there were 4 practice schedules: 0-1-2-3, 0-2-2-2, 2-1-2-3 and 2-2-2-2. Control conditions were also included: no practice, a single immediate practice and a single delayed practice. Practice conditions were also crossed with a feedback manipulation. Scores on the final test a week later indicated a significant testing effect in that passages, which had not been previously tested, were less well recalled than passages, which had been tested four times. Feedback was also reported as beneficial to the learning process. For the groups who received feedback, having a delayed first test produced significantly greater recall whereas for the non-feedback groups the position of the first test did not have a significant effect on recall. In addition whether the practice schedule was expanding or uniform had no significant effect.



Overall Karpicke and Roediger (2010) drew out four main conclusions. Firstly that there is a testing effect, as taking a test was found to enhance long-term recall better than re-reading the text. Secondly repeated retrieval attempts produced greater long-term retention than taking a single test. Thirdly testing with feedback, such as restudying the information, produces better retention than testing without feedback. Finally, expanding and uniform spaced practice schedules produce no significant differences in terms of final recall.

The present research project aimed to replicate Karpicke and Roediger's (2010) study with a few changes. This study aimed to investigate whether the actual spacing schedule of practice tests, in terms of the number of practice tests and the time of the initial practice test, has a significant effect on long-term recall. Taking into consideration Karpicke and Roediger's (2010) findings, feedback was included for all practice tests. However the study also aimed to look at whether an immediate test as a comprehension check, without feedback, would affect final recall, despite any differences in the following practice schedules.

In addition the present research addressed some key methodological criticisms, which limit the extent to which Karpicke and Roediger's (2010) findings can be validly applied to a real-life educational context. Firstly with Karpicke and Roediger (2010) the intervals between practices and the timescale of the whole study were trivial and unrealistic when compared to real undergraduate study schedules. The study sessions and filler tasks were each only four minutes long and the whole practice schedule only took an hour to complete. The final test took place one week later and only took 8 minutes. University students are expected to study over a much longer period of time, often having an hour's lecture on a particular topic and then are required to sit an exam on it several weeks later. In the present research practice sessions were spaced over 15 days and the final test took place four weeks after the initial study session and took on average 30 minutes to complete. Participants were given as much time as the needed to complete each practice session (test and feedback). This reportedly took approximately 15-20 minutes to complete each session.

Secondly Karpicke and Roediger (2010) did not use materials, which were an appropriate level of complexity to replicate undergraduate study. The texts used were brief passages taken from Encyclopedia Britannica on single topics just over 150 words in length. Therefore the actual material, which participants were required to practice and recall was considerably shorter, contained fewer concepts and was overall easier to understand and learn than texts, which undergraduates normally encounter. For this research the text used was a chapter from an academic book, 4000 words long. In addition the way participants initially studied the text was adapted to be more realistic. In Karpicke and Roediger (2010) the text was always read from a computer screen, however as this may not have been participants usual study method this may not have reflected realistic learning. In this study participants were encouraged to read the text in their usual style, for instance, underlining, highlighting or making notes to control such confounds.

Thirdly, although Karpicke and Roediger (2010) reported that feedback had an effect, the validity of these claims can be questioned, as it was not realistic feedback. Karpicke and Roediger (2010) gave participant feedback by allowing participants one

minute to re-read the text. In reality, students would not be limited to such a short period of time to re-read a text following a practice test. Participants were therefore given as much time as they need to re-read the text following the practice test.

Overall this research aimed to investigate the best practice conditions to learn from text, with a particular focus on undergraduate study. It looked at whether having more practice tests would improve recall, whether the time of the initial test influenced learning and whether having an immediate (massed) test led to greater recall four weeks later. In line with Karpicke and Roediger (2010) and past research on testing effects, it was predicted that participants would perform better on the final test following three practice tests rather than a single practice test. In addition, it was hypothesised that an immediate open-book test of material would result in higher long-term recall, at least for the tested material. Finally, due to inconclusive past evidence to form a basis for a hypothesis, no firm predictions were made on the influence of the time of the initial test and subsequent spacing of practice schedules. However due to the more realistic time frame of this study compared to Karpicke and Roediger (2010) it was possible that results would differ. For instance, as the initial test only took place after a number of days rather than minutes, it may have had a greater influence over long-term retention as recall on the first test would be difficult whether part of an expanding or uniformly spaced practice schedule. A far larger amount of forgetting may have occurred by the time the delayed initial test of the uniformly spaced schedule was taken, with the difficulty of the delay not being as desirable as seen in other studies.

## **Method**

### **Participants**

Forty-eight participants (22 males, 26 females) were recruited who were all undergraduate students at Lancaster University aged between 18 and 22. Twenty-one studied in the Faculty of Science and Technology, 26 studied in the Faculty of Arts and Social Sciences and 1 in the Management School. Participants were selected using an opportunity sample, with the researcher asking fellow students. Before the experiment began, contact details were collected from participants including email address and mobile number, so that the researcher could send out URL links and reminders about practice sessions. All participants were treated according to the British Psychological Society ethical code.

### **Design, Materials and Procedure**

A within-group design was used. The independent variables were the number of tests taken (one or three), the type of retrieval practice (early/expanding or delayed/equally spaced) and whether material was immediately tested or not. The dependent variables included the immediate open-book comprehension test scores, practice test scores as well as the final short-answer test scores four weeks after the initial study session.

The text used for study was the chapter 'Managing Pain' from *In the Mind's Eye-enhancing human performance* (Druckman and Bjork, 1991), see Appendix A. The text was approximately 4000 words in length and of similar difficulty and style to

readings university students are often set. The text had 3 different sections: Aspects of Pain, Factors of Pain and Treatments. However, as the section on treatments was longer, for purposes of counterbalancing it was considered as two sections. The 4 different sections of the test were assigned to the 4 conditions: Expanding practice schedule, uniform practice schedule, single early test and single delayed test. The assignment of parts of the text was counterbalanced across participants. Overall there were 8 groups of participants, labelled group 1a, 1b, 2a, 2b, 3a, 3b, 4a, 4b for counterbalancing purposes. Participants were assigned to groups 1 to 4 alternatively, with 12 participants in each group. Each group was then split into 2 sets, the first 6 in-group A and the last 6 in Group B.

### **Initial Study Session.**

A personalised information sheet was given to each participant at the initial study session. This contained the participant's group number, participant number and standardised instructions. This included an outline of what the experiment involved in the initial study session, the further study sessions, the final test and what dates each stage would take place on. Each participant was also given a printed copy of the text and asked to read it in their usual reading style, such as note taking, highlighting, underlining etc. Participants were given as long as they needed to read the text.

Then all participants completed an open-book comprehension test to consolidate learning. Two matched sets of questions, which tested the same idea units, were written; these are referred to as Set A and Set B in this report. Both sets consisted of 32 matched questions, with 8 questions for each section of the text. In order to control for item effects half of the participants practiced with each set and were later tested with the unpracticed set. The immediate open-book comprehension test consisted of half of the items they had just read. Half of the participants initially answered even number questions and the other half answered the odd questions. Both sets of questions were therefore divided into odd and even questions so that there were 4 different test papers; Set A Odd, Set A Even, Set B Odd and Set B Even. Each set of questions were kept in the same order as they featured in the text. Participants assigned to A groups were given the Set A questions and B groups the Set B questions. Assignment to either odd or even questions was decided by alternating allocation down the list of participants for each group. Participants were given lined paper to write down their answers to the questions and asked to hand in all the materials they had used when they had finished.

### **Practice Sessions.**

Practice schedules were dependent on four conditions: expanding practice schedule, uniform practice schedule, single early practice test and single delayed practice test, see Table 1. The expanding practice schedule involved practice on days 2, 7, 15 and the equally spaced practice schedule involved practice on days 5, 10 and 15. The single early practice test took place 2 days after the initial study practice to match the expanding schedule and the delayed practice test was given after 5 days to match the uniform spaced schedule. Each condition was then applied to a discrete part of the studied text. In order to control for order effects and other differences between

sections, the assignment of condition to a section of the text was counterbalanced across participants as shown in Table 2.

**Table 1**  
**Practice test schedules: the 2 x 2 design produced four schedules**

Time of Practice Test	Number of Tests	
	1	3
Early	Single Early Test	Expanding Practice Schedule
Delayed	Single Delayed Practice Test	Uniform Practice Schedule

**Table 2**  
**Counterbalanced assignment of conditions to sections of the text for each group of participants**

Group	Expanding	Uniform	Early Test	Delayed Test
1	Aspects	Treatment 2	Treatment 1	Factors
2	Factors	Aspects	Treatment 2	Treatment 1
3	Treatment 1	Factors	Aspects	Treatment 2
4	Treatment 2	Treatment	Factors	Aspects

Forty different online review activities were provided using online questionnaire software (SNAP); this included one for each of the 5 study days (days 2, 5, 7, 10 and 15), for each of the 8 groups of participants. All participants completed the same activities; the groups differed only in terms of which specific questions were included. On day 2, all participants answered questions associated with the early-single test and the early-expanding tests; on day 5 they answered questions assigned to the delayed-single test and delayed-uniform test; Day 7 questions related to the early-expanding tests were answered; on day 10 they answered questions associated with delayed-uniform; and on day 15 participants answered questions assigned to early-expanding and delayed-uniform. See Table 3 for the practice and testing schedule for all participants.

The online study sessions included all 8 questions for each of the sections being tested. The order of the questions was randomised in sets of 4. See Table 4 for order of questions. At the beginning of each study session the first question always asked for a participant number so that each participant's set of answers could be

identified by the researcher. Feedback in terms of presentation of the relevant section of text followed each set of questions. Participants were unable to return to the questions after seeing the text.

Each participant was emailed a URL link on the evening before a practice session was due to take place corresponding to their specific group and set, that is, group 1a, 1b, 2a, 2b and so forth. A reminder that they would need to complete the study session on the following day before midnight was also included in the email. At 9pm on the day of the study session those participants who had not yet submitted their answers received a reminder text to complete the study session before the end of the day. All participants completed each study session. When completing the online practice tests, participants simply had to type in their answers, click 'next' to move forward and then 'submit' to send their answers to the researcher.

**Table 3**  
**Practice and testing schedules (Day 0 – Day 28)**

Condition	Initial Study	1 <sup>st</sup> practice	2 <sup>nd</sup> Practice	3 <sup>rd</sup> Practice	Final test
Expanding practice	Read text and take open-book short answer test on half of target items  DAY 0	Test – Review After 2 days  DAY 2	Test – Review After 5 more days  DAY 7	Test – Review After 8 more days  DAY 15	Free recall followed by short answer test four weeks after initial study  DAY 28
Uniform practice		Test – Review After 5 days  DAY 5	Test – Review After 5 more days  DAY 10	Test – Review After 5 more days  DAY 15	
Single early practice		Test – Review After 2 days  DAY 2	----	----	
Single delayed practice		Test – Review After 5 days  DAY 5	----	----	

**Table 4**  
**Order of questions for online study sessions**

Questions	Aspects	Factors	Treatment 1	Treatment 2
1-4	2,3,4,1	1,2,3,4	3,1,2,4	4,2,3,1
5-8	7,8,6,5	6,7,8,5	5,6,7,8	8,6,5,7

### **Final Test.**

Participants attended a session 4 weeks after the initial study session and were given a closed book test. The complete lists of 32 questions were used for both Set A and Set B in the order that they appeared in the text. Participants who had been given Set A questions in the initial test and practice sessions were tested using Set B questions and vice versa. This was in order to consider retention in terms of generalising knowledge rather than simply memorisation of answers, which had already been given. At the top of the question paper were also some brief instructions, asking participants to answer questions on a separate piece of paper labelled with their name and participant number and then to hand in both papers when they had finished. After finishing the final test participants were then given a full debrief.

### **Pilot Research**

Four students were recruited for the pilot research, 3 studied at Leeds University and 1 at York University. All were aged between 20 and 22. The pilot research aimed to assess the clarity and level of difficulty of the question sets in order to eliminate the possibility of ceiling or floor effects. Participants were emailed the text to read and instructed to complete it in their usual reading style, for example highlighting or making notes and asked to time how long it took to read, although they could take as much time as they needed. Two participants were sent the whole set of A questions and 2 participants were given Set B questions. Both consisted of 37 questions at this stage. Participants were told to try firstly to answer the questions without consulting the text on a document labelled 'no text'. They could then go back and answer any questions that they had been unable to answer on a separate document labelled 'with text'. Following the pilot research the sets of questions were cut down to 32 questions and some of the questions were re-worded. It was also decided that the initial study test would involve an open book test rather than a closed book test as participants achieved very low scores when attempting to answer the questions without the use of the text.

### **Scoring Procedure**

Tests were given two different marks: a strict score and a lenient score. A strict mark was awarded for an answer, which was completely correct and contained all the key details. A lenient mark was awarded if the answer contained some key details but was lacking in certain aspects. The initial study practice test was out of a maximum

of 16 marks, 4 marks for each section and the final test was out of a possible 32 marks with a maximum of 8 marks for each section.

## Results

Final test scores are summarised in Table 6 (strict) and 7 (lenient). A high-test score indicated high retention of the information of that section of the text.

**Table 6**

**4 week delayed test mean strict scores. For each cell the maximum possible score is 4. Standard deviations are in parenthesis. N=48**

Initial Practice	Number of Practice Sessions	
	One	Three
Material with immediate test		
Early (Day 2)	1.83 (1.08)	1.73 (1.12)
Delayed (Day 5)	1.71 (1.29)	1.85 (1.13)
Material without immediate test		
Early (Day 2)	1.62 (1.28)	2.02 (1.14)
Delayed (Days 5)	.90 (1.00)	1.98 (1.25)

**Table 7**

**4 week delayed test mean lenient scores. For each cell the maximum possible score is 4. Standard deviations are in parenthesis. N=48**

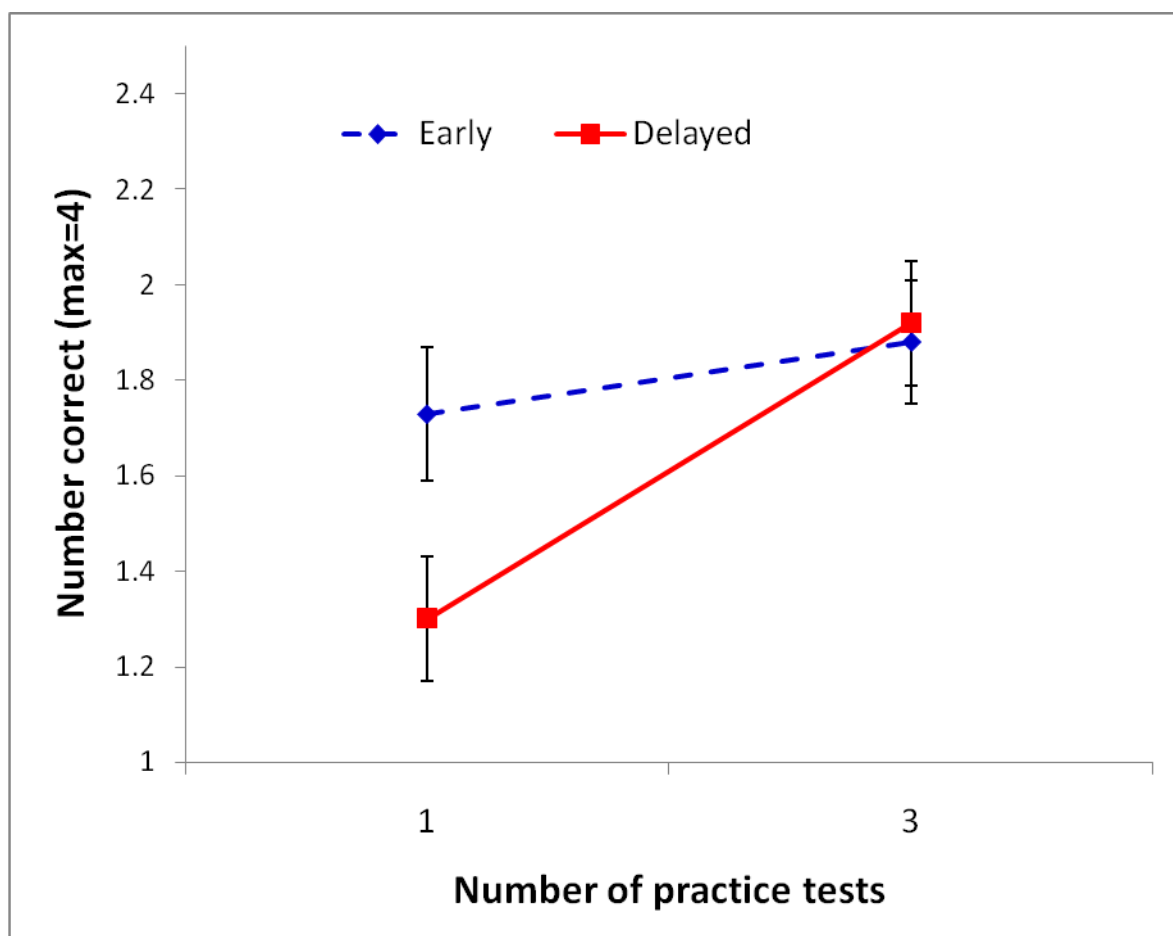
Initial Practice	Practice Sessions	
	One	Three
Material with immediate test		
Early (Day 2)	2.40 (1.03)	2.48 (1.01)
Delayed (Day 5)	2.44 (1.22)	2.58 (1.15)
Material without immediate test		
Early (Days 2)	2.10 (1.13)	2.63 (1.10)
Delayed (Days 5)	1.65 (1.10)	2.37 (1.25)

A three way within subjects ANOVA investigated the effects on final test scores of three factors: number of practice tests, time of initial practice test and whether material was immediately tested or not. Strict scores were used in the final analysis as they were thought to reflect a more accurate score of whether specific information

had been firmly locked into long-term memory. Lenient scores gave an indication of whether general information could be recalled but strict scores showed whether a deep level of learning had taken place.

The number of practice tests taken had a significant effect on long-term recall,  $F(1, 47) = 7.48$ ,  $MSE = 1.87$ ,  $p = .009$  with a large effect size of partial  $\eta^2 = .14$ . With three practice tests participants scored higher on the final test and so appeared to have remembered more. The main effect of the time of the initial practice test did not have a significant effect on the final test score,  $F(1, 47) = 3.49$ ,  $MSE = 3.57$ ,  $p = .068$ , partial  $\eta^2 = .07$ . The effect of whether material was immediately tested or not was also non-significant,  $F(1, 47) = 2.58$ ,  $MSE = .85$ ,  $p = .115$ , partial  $\eta^2 = .052$ .

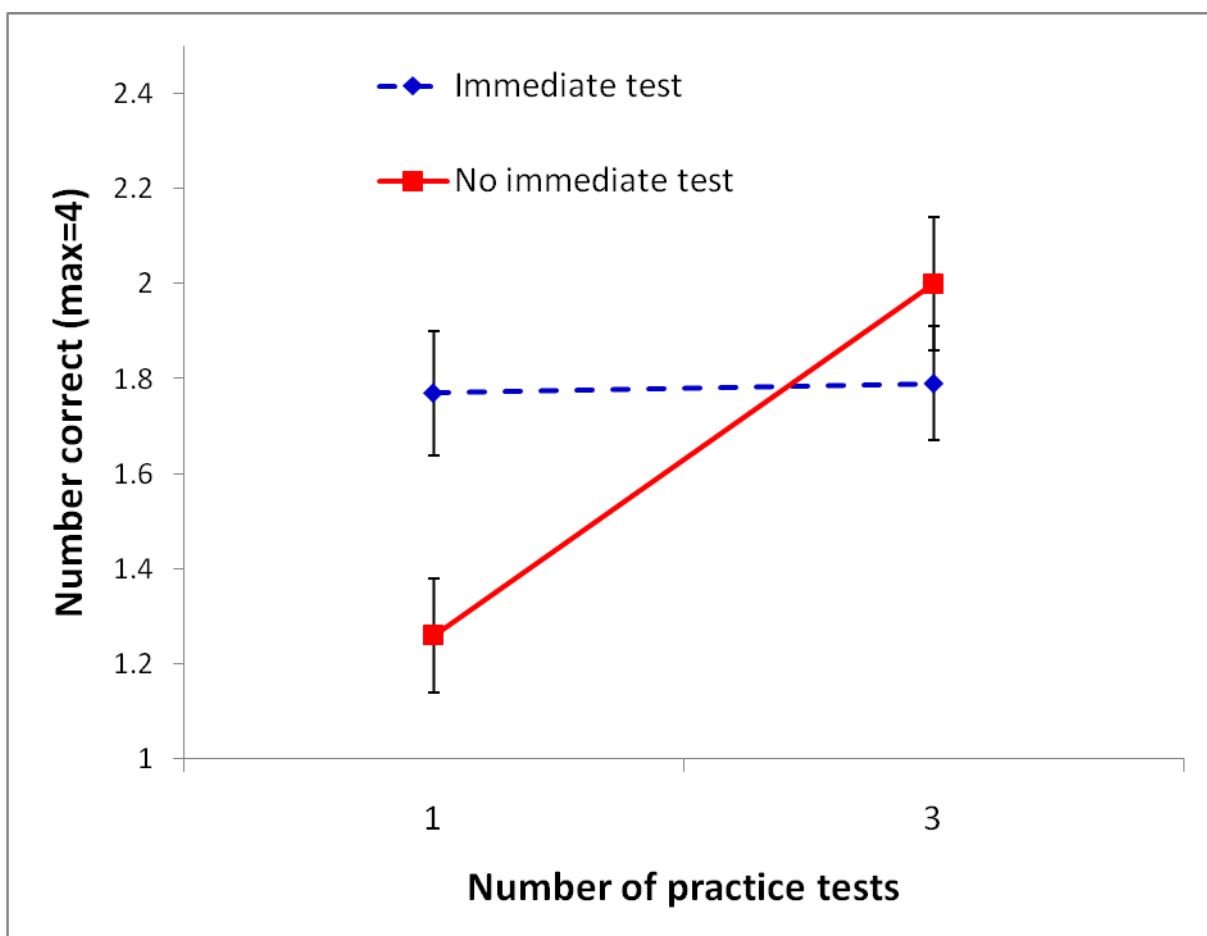
The interaction between the number of practice tests and the time of the initial practice test (shown in Figure 2) was significant,  $F(1, 47) = 4.14$ ,  $MSE = 1.27$ ,  $p = .048$  with a medium effect of partial  $\eta^2 = .08$ . Where participants had three practice tests, the time of the initial practice test did not have a significant effect on final recall. However when participants only had one practice test the time of the test made a significant difference. Higher scores on the final test were achieved when the single test was early rather than delayed. Overall this suggests that when students only take one practice test they will remember more if it is taken earlier rather than later.





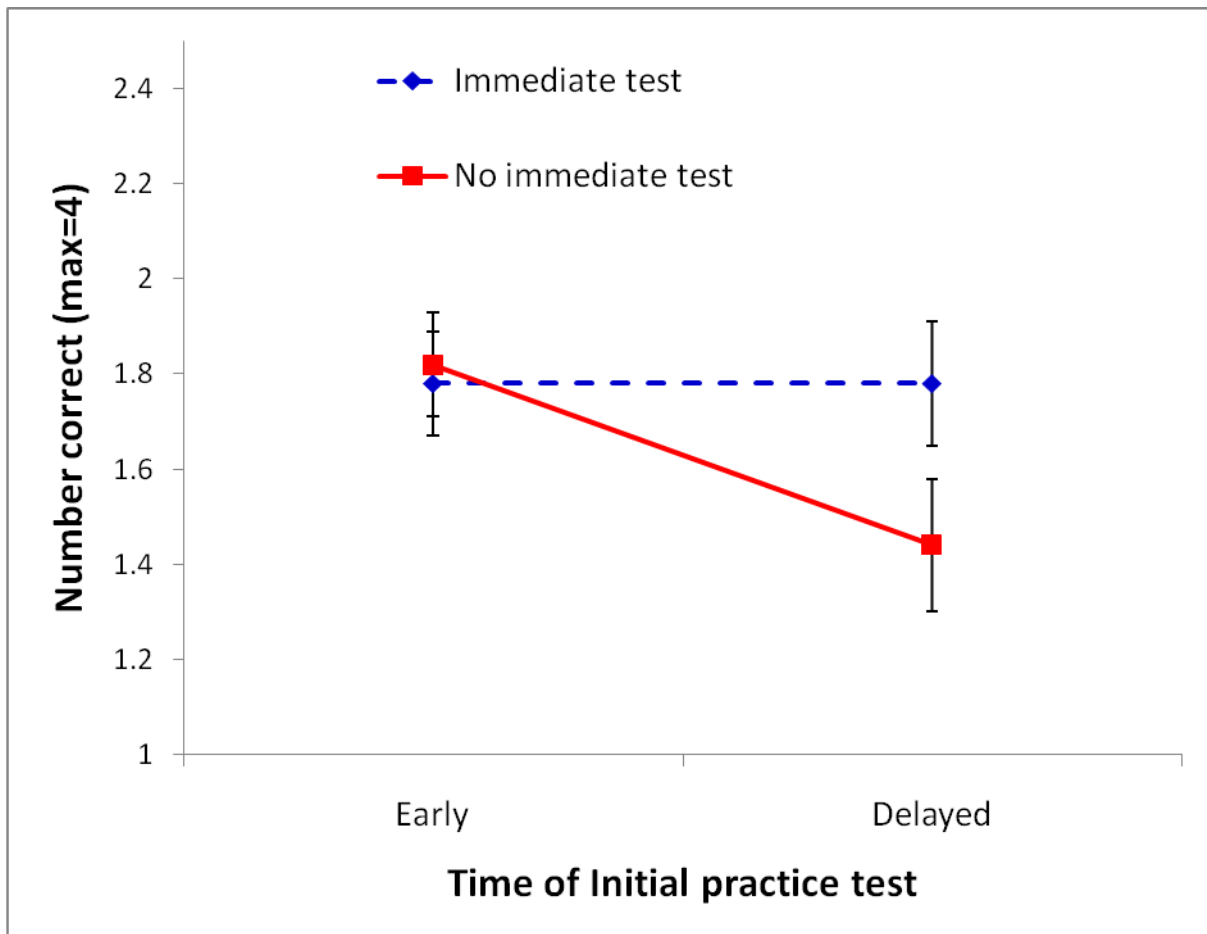
**Figure 2: Interaction between the number of practice tests and the time of the initial practice test. Mean strict scores used, (N=48). Error bars represent  $\pm 1$  SEM**

The interaction between the number of practice tests and whether material was immediately tested (shown in Figure 3) was significant,  $F(1, 47) = 12.08$ ,  $MSE = 1.03$ ,  $p = .001$  with a large effect size of partial  $\eta^2 = .21$ . When participants had three practice tests, whether the material had been immediately tested or not did not have a significant effect on final recall. However when participants had only one practice test, material that had been immediately tested was better recalled than material, which had not. This indicates that when there will only be one practice test, individuals will remember more if they have an immediate test after the initial presentation of the information.



**Figure 3: Interaction between the number of practice tests and whether material was immediately tested. Mean strict scores used, (N=48). Error bars represent  $\pm 1$  SEM**

The interaction between the time of the initial practice test and whether the material had been immediately tested (see Figure 4) was marginally non-significant,  $F(1,41) = 3.13$ ,  $MSE = 1.14$ ,  $p = .3.13$ , partial  $\eta^2 = .06$ . However the data provided an interesting pattern. Although non-significant, when the initial practice test was delayed, material, which had been immediately tested, tended to be remembered better.



**Figure 4: Interaction between the time of the initial practice test and the time of the initial test. Mean strict scores used (N=48). Error bars represent  $\pm 1$  SEM**

The overall interaction between all 3 factors was non-significant,  $F(1, 47) = 1.19$ ,  $MSE = .97$ , partial  $\eta^2 = .025$ .

## Discussion

The aim of the experiment was to further investigate the most effective ways students could learn from text. Past research has explored the benefits of different types of spaced practice schedules but there has been limited work, which has investigated the best ways to learn from texts. Even with the majority of work, which has focused on learning word pairs, there have been inconsistent results. Karpicke and Roediger (2010) examined whether expanding retrieval spaced practice was superior to equally spaced schedules when learning text materials and found that it was not. However they did report that multiple tests were better than a single test.

The hypotheses for this study were that three practice tests would be more beneficial than a single practice and an immediate test would improve long-term recall, tested four weeks later. Due to mixed results in past empirical work a firm prediction was not made on the influence the time of test and subsequent spacing schedules would have on long-term retention. However it was discussed whether results would differ from those of Karpicke and Roediger (2010) given the changes to a more realistic

situation with a more substantial text, a longer initial delay of days rather than minutes and a final test only taking place four weeks after the first study session. Due to the longer period between retrieval attempts there was more time for information to be forgotten and therefore retrieval was more difficult for both expanding and uniformly spaced practice schedules compared to the conditions used in Karpicke and Roediger. The delay for the initial test may have proved too difficult in the uniformly spaced schedule (5 days) and therefore the expanding spaced schedule may have yielded higher long-term retention due to its 'desirable difficulty'.

There was a significant interaction between the number of practice tests and the time of the initial practice test. Although the time of the initial test did not have a significant influence over long-term recall when there were three practice tests, when students only took one practice test long-term recall was higher when that practice test was early rather than delayed. This may be explained by Bjork's (1999) ideas of desirable difficulty. With a single retrieval attempt, participants only had one chance to actively recall information and therefore a large amount of forgetting may have occurred before the final test four weeks later. Furthermore, if the test was delayed a large amount of forgetting may have taken place before the first practice test. Cepeda et al. (2009) reported that in their research the optimal gap between presentation of information and the initial test for long-term recall was one day. Therefore the delay of five days may have been too difficult for participants to efficiently recall. An early test could have increased the likelihood of retrieval success as less forgetting may have occurred by the time the practice test took place. Material recalled in the first practice test is presumably more likely to be remembered in the long-term than material that was not. Therefore, if successful retrieval is more likely with an early practice test than a delayed practice test, it would be expected that an early test would lead to higher long-term recall. Consequently, the amount of information actively recalled in the practice test is likely to have influenced the level of long-term recall. However when there were three practice tests, performance on the initial test may not have been as crucial, as individuals had several opportunities to actively recall and restudy the text. In order to support these claims further analysis is needed of the individual test scores of each practice test. This interpretation is further supported by Karpicke and Roediger's (2010) observation that in some conditions following an early test final test scores were higher than the delayed test condition due to increased recall on the immediate first test.

A significant interaction was also found between the number of practice tests and whether material was immediately tested. Although there was no significant effect of the immediate test for the material, which had three practice tests, when there was only one practice test material, was better recalled if it had been immediately tested after initial reading. As explained above, material with three practice tests benefitted from several opportunities to actively recall, review information and improve on areas, which they knew they had previously forgotten. However when there was only one practice test there was obviously less chance to review and recall material. Therefore it is possible that active generation of material, due to an immediate test, increased retrieval strength for that information in subsequent practice sessions and overall long-term recall. In line with Craik and Lockhart's level of processing theory (1972), immediately tested material was more deeply processed in the initial sessions and therefore more accessible and easily recalled after a period of disuse.

This effect may have decreased with an increase in practice tests as participants were given several additional opportunities to engage in deep processing and memorising information, which hadn't been tested immediately, as well as strengthening recall of information, which had.

The interaction between the time of initial test and whether material had been immediately tested or not was marginally non-significant. However there was a pattern that suggested that if the initial test was delayed by five days, participants had greater long-term recall if they had been given an immediate test. One explanation for this is that the immediate test helped to consolidate learning of information, which was more likely to be forgotten if not tested or reviewed until five days later. If the initial practice session was undertaken after only two days there was a shorter length of time for forgetting to occur. The influence of an immediate test may not have been as great when the test was early as recall may have been at a higher level regardless.

Overall, as was hypothesised, the number of practice tests had a significant effect on long-term recall. This is in line with Karpicke and Roediger's (2010) findings and Ebbinghaus' theory (1885) that repeated retrieval practice enhances long-term recall more than a single retrieval attempt. Yet despite the apparent benefits of multiple testing, having an immediate test after the initial reading of the text did not have an overall significant effect on recall. The original hypothesis predicted that immediate testing would increase long-term recall, as it would act as an additional opportunity for deep processing of information through generation, increasing overall long-term retrieval strength. However when looking at the individual interactions it is clear that immediate testing does have a significant effect on recall, but that the effect is overwhelmed by the benefits of additional practice. This perhaps suggests that the benefit to multiple testing may be limited, within the time frame of four weeks, in the sense that after a certain number of practice tests recalls will stay at a constant level and not continue to increase as the number of tests increase. However it is possible that if long-term recall was tested after a longer period of time, such as a year later, then more practice would be needed and an immediate test may have had a significant effect on long-term recall.

In addition, the overall effect of the time of the initial practice test was found to be non-significant. This is consistent with Karpicke and Roediger (2010) who reported that when feedback was given, the delay between the initial presentation and the first practice test did not have a significant influence on long-term recall, that is, an early test was not more beneficial than a delayed test. Furthermore support is provided for the conclusions drawn by Karpicke and Roediger (2010) that neither expanding nor uniformly spaced practice schedules are superior. Having more tests significantly increased recall but the positioning of these tests did not and therefore this simply shows that there is a benefit of repeated recall regardless of when it happens. Recent work such as Cull et al. (1996) and Pyc and Rawson (2007) have shown a similar pattern of results.

In addition, although Landauer and Bjork (1985) reported expanding practice schedules to be superior, in contrast to the present study and Karpicke and Roediger (2010), this was without re-presentation of material following each test. Nevertheless, in the second experiment when re-presentation of material was

included, Landauer and Bjork (1985) found that uniform practice schedules produced as high if not higher long-term recall. Consequently whether material is re-presented after practice testing may be a key factor in how effective different spaced practice schedules are. Furthermore in an educational context, in which representations are possible and often recommended, the effectiveness of spaced practice schedules, which involve review after tests, can be considered more relevant than those that do not.

Although Karpicke and Roediger (2010) drew similar conclusions, the design of their experiment limited the validity of application to real-life educational practice. The present study made several adaptations to improve the generalisability of the results. This included having a longer time frame for study taking place over a period four weeks, a text appropriate for an undergraduate level of study, substantial practice sessions and tests and realistic feedback. These improvements enhanced the validity of the conclusions and therefore they should be taken seriously by students and educators. In the last decade there has been an increase interest in the development of study skills with research such as Fleming (2002), indicating a link between using study strategies and high exam performance. This research can be useful in advising students how to space their study for seminars and revision for end of module exams as well as highlighting the value of practice testing. Furthermore, these conclusions may have important consequences for how courses should be structured in the future. For instance, rather than simply having revision sessions at the end of the year, it may be more beneficial to have review, sessions, which include practice tests spaced throughout the course.

However, although the study has a high level of ecological validity when considering the type of text and the timeframe of the whole study, there are still some limitations to application. Although participants were urged to complete all the study sessions, the text was not actually needed for the student's course and their final test mark was therefore unimportant. This lack of motivation to do well and thoroughly learn the information from the text may have affected how participants performed throughout the study. For instance, many participants reported that they skipped or just skim read the feedback text in order to finish the study session more quickly. Although the within-participants design and counterbalancing aimed to control for individual differences and fatigue effects between conditions, these extraneous variables may have influenced the results. In further work perhaps it would be more valid and reliable to imbed different types of practice testing within existing degree schemes in a form of quasi-experiment as did Fritz and Morris (2003). With the motivation to learn the information comprehensively, participants may be more conscientious and actively involved, than in a situation where their final test mark would have no personal long-term consequences.

In addition, the subject participants usually studied may have influenced how well they recalled the information from the text after a delay of four weeks. Although the text was chosen as a general topic, not relating to one particular degree scheme and a possible area of interest for all students, the study sessions may have been more tailored towards certain subjects. For instance, although all students are asked to learn from text at some point in their degree, many Math's and Science student participants reported that studying from long texts and comprehension tests was not part of their regular studying programmers. Although a within-participants design

accounted for individual differences between participants in this study, this questions whether the conclusions about spacing practice can be applied to all educational contexts as content of practice sessions will differ between subjects. This present study has indicated that as long as there are repeated, spaced practice attempts, the type of spacing (either expanding or uniform) does not make a significant difference. However these claims are made in just the context of learning from text and different results may have been found if materials focused on numerical based learning.

Past research has investigated the best ways to study mathematics and statistics, finding spaced practice to be superior to massed practice (Rea & Modigalini, 1985; Rohrer & Taylor, 2007; Smith and Ropkopf, 1984). However there has been limited research, if any, which has explored whether expanding or uniformly spaced practice is superior when learning mathematical equations, statistical methods and problem solving. Often with math's problems a firm grasp of concepts or equations is initially required so that later practice can involve application for problem solving. The different processes involved in mathematical study compared to learning from text may therefore be more suited to one type of practice schedule. The early test of the expanding spaced practice schedules may make it easier to achieve a high level of initial retrieval success and a firm grasp of crucial concepts whilst maintaining the desired difficulty through subsequent spacing. If a high degree of forgetting has occurred by the time of the initial practice, as would be more likely with a uniform schedule, students may lack the theoretical grounding of methods to solve problems in subsequent practice sessions. Although Rea and Modigliani (1985) found a benefit of expanding practice schedules for children learning multiplication facts, this was compared to massed practice and not an equally spaced practice schedule. Further research is therefore needed to determine the most beneficial spacing practices and study strategies for different types of learning.

Overall this experiment intended to determine the best ways for students to learn from text. It was concluded that multiple retrieval practices will lead to higher levels of long-term retention than a single retrieval practice. Furthermore when there are multiple practice tests, the time of the initial practice (early or delayed) and the spacing of the practice schedules (uniform or expanding) do not make a significant difference to overall learning. Therefore, in order to improve learning and perform highly on exams, students should aim to actively recall important information on several separate occasions, at whatever times are best for them, to improve. However if it is only possible to actively practice recall once, it is more beneficial to have an early practice and an immediate test following initial presentation. This study focused in particular on the most efficient ways for university students to space practice and therefore further work is needed to see if the same recommendations were suitable for primary school, high school and college aged children. However these valid conclusions have important implications for how university students and educators should use spaced practice testing to learn from text when planning study schedules and developing the structure of university courses.

## References

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

Balota, D. A., Duchek, J. M., & Logan, J. M. (2007). Is expanded retrieval practice a superior form of spaced retrieval? A critical review of the extant literature. In J. S. Nairne (Ed.), *The foundations of remembering: Essays in honor of Henry L. Roediger* (pp. 83-106). New York: Psychology Press.

Bjork, R. A. (1975). Retrieval as a memory modifier. In R. Solso (Ed.), *Information processing and cognition: The Loyola Symposium* (pp. 123-144). Hillsdale, NJ: Lawrence Erlbaum Associates.

Bjork, R. A. (1999). Assessing our own competence: Heuristics and illusions. In D. Gopher & A. Koriat (Eds.), *Attention and Performance XVII. Cognitive regulation of performance: Interaction of theory and application* (pp. 435-459). Cambridge, MA: MIT Press.

Bjork, R. A., & Bjork, E. L. (1992). A new theory of disuse and an old theory of stimulus fluctuation. In A. F. Healy, S. M. Kosslyn, & R. M. Shiffrin (Eds.), *Essays in honor of William K. Estes: From learning processes to cognitive processes* (Vol. 2, pp. 35-67) Hillsdale, NJ: Erlbaum.

Carpenter, S. K., & DeLosh, E. L. (2005). Application of the testing and spacing effects to name learning. *Applied Cognitive Psychology, 19*, 619-636.

Cepeda, N. J., Coburn, N., Rohrer, D., Wixted, J. T., Mozer, M. C., & Pashler, H. (2009). Optimizing distributed practice: Theoretical analysis and practical applications. *Experimental Psychology, 56*, 236-246.

Chi, M. T. H., de Leeuw, N., Chiu, M-H., & LaVancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science, 18*, 439-477.

Craik, F. I. M., & Lockhart, R. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behaviour, 11*, 671-684.

Cull, W. L. (2000) Untangling the benefits of multiple study opportunities and repeated testing for cued recall. *Applied Cognitive Psychology, 14*, 215-235.

Cull, W. L., Shaughnessy, J. J., & Zechmeister, E. B. (1996). Expanding understanding of the expanding-pattern-of-retrieval mnemonic: Toward confidence in applicability. *Journal of Experimental Psychology: Applied, 2*, 365-378.

Delaney, P. F., Verhoeven, P. P. J. L., Spiguel, A. (2010). Spacing and testing effects: A deeply critical, lengthy and at times discursive review of the literature. *Psychology of learning and motivation, 53*, 63-147.

Dempster, F. N. (1997). Using tests to promote learning. In R. F. Dillon (Ed.), *Handbook on testing* (pp. 332-356). London: Greenwood Press.

Druckman, D., & Bjork, R. A. (1991). Managing Pain. In D. Druckman and R. A. Bjork (Eds.), *In the Mind's Eye: Enhancing Human Performance*. Washington, DC: National Academy Press.

Ebbinghaus, H. (1885). *Memory*. Oxford, UK: Dover.

Fritz, C. O. (2011). Testing, Generation and Spacing Applied to Education. In A. Benjamin (Ed.), *Successful remembering and successful forgetting: A Festschrift in honor of Robert A. Bjork* (pp.199-216). London: Psychology Press.

Fritz, C. O., & Morris, P.E. (2003). *Expanding retrieval practice: Investigating the parameters*. Paper presented at the 5<sup>th</sup> Biennial Conference of the Society for Applied Research in Memory and Cognition, Aberdeen. Lancaster, UK.

Glover, J. A. (1989). The 'testing' phenomenon: Not gone but nearly forgotten. *Journal of Educational Psychology, 81*, 392-399.

Hadwin, A. F., & Wine, P. H. (2009). Study strategies have meagre support: a review with recommendations for implementation. *The Journal of Higher Education, 67*, 692-715.

Fleming, V. M. (2001). Helping students learn to learn by using checklists, modified rubrics and email. *Journal of Excellence in College Teaching, 12*, 4-21.

James, W. (1890). *The principles of psychology*. Cambridge, MA: Harvard University Press.

Jones, H. E. (1923). Experimental studies of college teaching. *Archives of Psychology, 10*, 68.

Karpicke, J. D., & Roediger, H. L., (2007). Expanding retrieval practice promotes short-term retention, but equally spaced retrieval enhances long-term retention. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*, 704-719.

Karpicke, J. D., & Roediger, H. L., (2007). The critical importance of retrieval for learning. *Science, 319*, 966-968.

Karpicke, J. D., & Roediger, H. L., (2010). Is expanding retrieval a superior method for learning text materials? *Memory and Cognition, 38*, 116-124.

Landauer, T. K., & Bjork, R. A. (1978). Optimal rehearsal patterns and name learning. In M. M. Gruneberg, P. E. Morris, & R. N. Sykes (Eds.), *Practical aspects of memory* (pp. 625-632). London, UK: Academic Press.

Mace, C. A. (1932). *The psychology of study*. London: Methuen.

Morris, P. E., & Fritz, C. O. (2000). The name game: Using retrieval practice to improve the learning of names. *Journal of Experimental Psychology: Applied, 6*, 124-129.

Morris, P. E., & Fritz, C. O. (2002). The improved name game: Better use of expanding retrieval practice. *Memory, 10*, 259-266.



Morris, P. E., & Fritz, C. O. (2006). *Reduced cuing as an alternative to improving memory using expanding retrieval practice*. Poster presented at the 47<sup>th</sup> Annual Meeting of the Psychonomic Society, Houston, TX.

Pyc, M. A., & Rawson, K. A. (2007). Examining the efficiency of schedules of distributed retrieval practice. *Memory and Cognition*, 35, 1917-1927.

Raaijmakers, J. G. W. (2003). Spacing and repetition effects in human memory: application of the SAM Model. *Cognitive Science*, 27, 431-452.

Rawson, K. A., & Kintsch, W. (2005). Rereading effects depend on time of test. *Journal of Educational Psychology*, 97, 70-80.

Rea, C. P., & Modigliani, V. (1985). The effect of expanded versus massed practice on the retention of multiplication facts and spelling lists. *Human Learning*, 4, 11-18.

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

Rohrer, D., & Taylor, K. (2007). The shuffling of mathematics problems improves learning. *Instructional Science*, 35, 481-498.

Shaughnessy, J. J., & Zechmeister, E. B. (1992). Memory-monitoring accuracy as influenced by the distribution of retrieval practice. *Bulletin of the Psychonomic Society*, 30, 125-128.

Smith, S. M. (1986). Environmental context-dependent recognition memory using a short-term memory task for input. *Memory and Cognition*, 14, 347-354.

Smith, S. M., & Rothkopf, E. Z., (1984). Contextual enrichment and distribution of practice in the classroom. *Cognition and Instruction*, 1, 341-358.

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

Thorndike, E. L. (1914). *The psychology of learning*. New York: Teachers College.

The British Psychological Society (2009). *Code of conduct, ethical principles and guidelines*. (online) Leicester: British Psychological Society. Available at: [http://www.bps.org.uk/downloadfile.cfm?file\\_uid=6D0645CC-7E96-C67F-D75E2648E5580115&ext=pdf](http://www.bps.org.uk/downloadfile.cfm?file_uid=6D0645CC-7E96-C67F-D75E2648E5580115&ext=pdf). (Accessed 2 May 2011).