The Testing Effect in Institutional Video Safety Training: A Comparison of Testing, Pretesting and the Influence of Feedback in Long Term Retention

Julio Ludowieg Cassinelli

Master thesis Psychology, specialization Occupational Health Psychology

Department of Psychology

Faculty of Social and Behavioural Sciences – Leiden University

Date: 13 August, 2015

Student number: s1471856

Supervisor: Jop Groeneweg

Abstract

This study focuses on the Testing Effect in realistic safety video trainings with minimal human intervention (such as company inductions). Testing, accompanied by Feedback and Pretests, are compared in 4 different learning conditions (in a 2 X 2 design) in order to observe mnemonic retention in the long and short term. In addition, it was also attempted through the pretests to prime retention of specific information. Subjects were tested in two occasions, one after watching a safety in height video and another after one week. The results were analyzed by means of an ANOVA analysis. The results yielded significant effects for Feedback in the long term retention. However, no relevant results were found for pretesting and primed learning in the short and long term.

Testing Effect

The testing effect, also known as retrieval practice, occurs when a test is applied to a person's study regime. This has been found beneficial for increasing one's ability to retain information especially when tests involve mainly recall (Dempster, 1996; Toppino & Cohen, 2009). This effect has been found to produce a better retention than restudying the same information for an equivalent time frame (Roediger & Karpicke, 2006a). Additionally to this, the Testing Effect has a delayed benefit over mnemonic retention. The effects are only seen after some time has elapsed. Wheeler, Ewers and Bounanno (2003) found that restudying had better results when free-recalling after a time span of five minutes. Nevertheless, this declined as time passed. In contrast to this persons who were tested had a better retention, in relation to the restudy group, reaching its highest difference after seven days. Similar results have been found by Toppino & Cohen (2009), Roediger & Karpicke (2006b) and Carpenter, Pashler, Wixted & Vul (2008). Also, it was found that the response of the subject while attempting to recall was faster when being tested than when restudying, even when this information came from the short and long term memory (Van den Broek, Segers, Takashima & Verhoeven, 2014).

Roediger and Karpicke (2006b) distinguished that testing has mediated and direct effects. In the former, testing can facilitate performance indirectly by improving further study (e.g. allowing choosing a study strategy) and in the latter, testing per se generates better retention.

Another aspect in regard to the testing effect is if it promotes transfer of knowledge. This is if testing can allow a person to construct new responses and answer dissimilar questions. According to Roediger & Butler (2011), this is true when comparing testing and restudying. They explain that testing can promote a more effective knowledge transfer from one domain to another (For example applying

knowledge from echolocation of bats to submarine sonar operation). This implies that testing can promote far transfer and not only an immediate simplistic one (Barnett & Ceci, 2002)

It's not yet clear by which mechanisms testing affects memory retention. Nevertheless, the yielded results of the studies point that retrieval practice during a test affects later retention through different mechanisms than those being used during studying. A possible explanation of why retrieval practice can enhance long term memory retention could be that a retrieval attempt can be thought as an effort in the reprocessing of memory. The more effort done, in the retrieval of the information, implies more reprocessing. This probably involves more elaboration of the memory trace and retrieval routes making it easier to recover (Toppino & Cohen, 2009). Wheeler et. al. (2003) also proposed that studying may strengthen a representation in the memory but retrieval practice strengthens the retrieval process per se increasing the memory resilience and thus reducing memory loss over time.

Testing Effect Enhancers: Retesting

The finding that retrieval practice boosts mnemonic retention brings into subject a question: What are the optimal conditions for retrieval (testing) so it can enhance retention over the long term? Craik & Watkins (1973) findings indicate that a short delay between retrievals can help achieving an errorless free recall which could lead to a better retention. This comes from the idea that if a person makes an error in retrieval this would be learned and thus making the learning of the correct answer more difficult.

However, tests which were easy, with none or little error in answering had less impact over long term recall than more difficult ones that had more wrong answers (Pyc & Rawson, 2009).

Roediger & Butler (2011) indicate that performing five to seven retrieval practices seems to be the optimal amount to improve long term retention regardless of the timing of the final evaluation. In addition to this, the time lapses between each retrieval attempt are more effective over long term

memory when six minutes have passed rather than attempting them after only one minute, regardless of the amount of retrievals attempted. Short delayed retrievals might generate a rote rehearsal making it easier to recall but with not much mnemonic benefit creating a ground-floor performance (Craik & Watkins, 1973). On the other hand, leaving a larger time gap generates greater effort to recall something and thus this seems to be the intervening factor for long term retention (Karpicke and Roediger, 2007). These findings are in line with Pyc & Rawson (2009) which found that difficulty can promote long term retention in retrieval practice.

Testing Effect Enhancers: Feedback

Another enhancer of long term retention is feedback. Although retrieval practice helps achieving better long term retention, even without any feedback, giving the right answer after a retrieval attempt augments the mnemonic benefit (Bangert-Drowns, Kulik, C.C., Kulik, J.A. & Morgan, 1991; Kulhavy & Stock, 1989). This occurs due that feedback gives the correct answer which in turn increases learning enabling the person to correct their mistakes (Pashler, Cepeda, Wixted & Rohrer, 2005) and maintain the correct responses (Butler, Karpicke & Roediger, 2008).

Another aspect that affects feedback is the timing. The benefit of feedback is commonly thought to be the strongest when it is given immediately after test (Kulik, J.A.& Kulik, C.C., 1988; Skinner, 1954). However more recent experiments regarding this contradict the previous findings. These studies found that delayed feedback boosts the ability to retain information over longer periods of time (Butler, Karpicke & Roediger 2007; Metcalfe, Kornell & Finn 2009). Other studies have found that just by giving immediate feedback just after the complete test has a higher effect on retention versus giving the feedback directly in every answer (Wheeler et. al., 2003). The advantage of delayed feedback has been observed even when giving it after one day. Another particular aspect worth mentioning is that delayed

feedback increases the probability to retain a correct response in the long term but is has minimal effects over error correction and in maintaining an erroneous response (Smith & Kimball, 2010).

Testing Effect Enhancers: Retrieval Practice Techniques

Logan & Balota (2008) studied the difference between two different techniques of retrieval practice. One of these is the expanding retrieval which attempts to gradually expand the time gap of retrievals (e.g., first retrieval attempt after 4 minutes after the learning phase, then the next after 8 minutes, and the next after 16 minutes and so on). The other is equal interval retrieval which attempts retrievals in a fixed interval of time (e.g., all retrieval attempts are done every 24 hours). Between these two procedures there has been found differences in regard to long term retention. Expanding retrieval seems best when retention within the same day is the objective. Nevertheless, it is not as effective as equal interval retrieval in longer term retention (i.e. when there's a delay of a day or more) (Karpicke & Roediger, 2007; Logan & Balota, 2008; Cull, 2000). The reason for this different effectiveness seems that in the expanding retrieval procedure the timing of the first retrieval is attempted shortly after learning which makes the person do less effort in recovering the information from the memory. Whereas, in the equal interval method this is delayed thus requiring a greater effort on the first test (Pyc & Rawson, 2009). Even if some errors are made (Roediger & Karpicke, 2010; Pashler, Zarow & Triplett, 2003).

Pre Testing Effect

The literature gives us support that the testing effect after learning is beneficial for long term retention. This is attempting retrieval of the information, previously learnt, from the memory enhances its permanence. However, attempting retrieval of the information before learning helps retention after the learning process? Rothkopf & Bisbicos (1967) found that asking participants questions that had numerical answers led to a better retention of all the numerical aspects from a text. This is explained

due that the participants directed more their attention to this type of information and therefore where able to memorize it more effectively. Nevertheless, cognitively speaking the testing effect has proven to be a long term memory enhancer even when there has been no opportunity to restudy the information which rules out that the benefit of testing can only by an attention process only (Richland, Kornell & Kao, 2009). Presley, Tanenbaum, Mc Daniel & Wood (1990) distinguished the effects between attention and pre testing. In their study the presented one group of subjects with passages and asked about its content, the other was asked about if these were well written. The participants who were asked about the content were better in retaining. Nevertheless, the majority of the answers were easy enough for the test group to answer in the pretest so the retrievals were successful. In another study, done by Richland et. al. (2009), this variable was controlled by giving a pretest which would be difficult to know without previous training. They found that the pretesting, in contrast to restudy, had a positive effect over retention even without a successful retrieval of information. These effects could even be seen after one week. They also explain that these findings indicate that pretesting enhances retention no only by merely directing attention but that it seems to strengthen the retrieval routes and also that failed retrieval attempts may reduce or suppress errors instead of reinforcing them. Another conclusion from this study was that pretested items don't impair the retention of those that were included in the test (applied after the learning session). Finally, they give two possible explanations for these results: one is that it is possible that retrieval attempts strengthen retrieval routes between the questions and a correct answer. A first thought someone could have is that failed retrieval could strengthen dead ends and not retrieval routes. On the other hand, it seems also plausible that failing in a retrieval attempt could weaken errors creating a positive setting which could facilitate retention. The second explanation relates to deep processing. They explain that pre-questions without retrieval might set in motion a process in which a person starts reasoning about the question which allows to mind set to acquire the information. For example: "How does a plane fly?" this question can trigger the person's imagination

and knowledge to find an answer, in this way activating a set of information which could help in the learning process)

The Present Research

This research focuses on the Testing Effect and its different mnemonic retention enhancers within an institutional video safety training setting. Testing, Pretesting and Feedback are used to create four different learning conditions which are compared to observe which one allows greater long term retention of the information (all these are retested after 7 days of delay). On the other hand, due that this research focuses on a practical situation, such as induction training, time constraints were considered in the different experiments (e.g. retesting multiple times and delaying feedback for hours, as literature recommends, might not be a plausible option due that many companies have to pay for this extra time, thus making training more expensive). Therefore what this research attempts is to find which plausible setting, for an induction or training context, has the best effects over memory retention in the long term.

In addition, this research tries to find the degree of retention through video training with the minimal human intervention. Because of this, in none of the experimental conditions the experimenter explains video contents or gives oral or written feedback to the participants. The only explanation that is given is through the displayed video and a feedback document (which is a test with the correct answers marked and projected on a screen).

Another goal that this study pursues is to discern if pretesting can actually prime the acquisition and retention of knowledge in the long and short term.

Method

Participants and Design

In this study 107 individuals participated, composed by students from three private universities from Lima, Peru. Subjects were recruited through a general announcement made by the teachers in the classrooms and by directly asking them in the common areas. The participant's sex distribution was 30% male and 70% female. The range of age was from 17 to 40, with 63% of subjects below the group average of 22.73 (SD=4.89). Within the sample 75.6% was born in the capital city of Lima, 23.2% in other parts of the Peru and 1.2% in a foreign country. Also, 75.6% studied Social Sciences, 11% Humanities and Arts, 11% Medical Sciences, 1.2% Engineering and 1.2% Business. Additionally the majority of the participants were not married comprising the 95.1% of the group.

Participants, formed in equal groups, performed two sessions in the experiment. The first had a length of 30 to 45 minutes and the second one from 10 to 15 minutes. Also, each participant received six euro in exchange for their collaboration.

Participants were assigned to four different conditions in a 2 (Pretest-test / test) X 2 (With feedback / No feedback after) design. All these conditions had a re test after one week. This creates the following groups:

Table I

Condition 1		Video	-->	Test	---------->	1 week posttest
Condition 2		Video	-->	Test --> Feedback	-->	1 week posttest
Condition 3	Pretest -->	Video	-->	Test	---------->	1 week posttest
Condition 4	Pretest -->	Video	-->	Test --> Feedback	-->	1 week posttest

Procedure

After all the participants were sitting on their personal table they were explained that the experiment consisted of doing a couple of written tests and watching a video. Also it was explained that the economical reward of six euros was not conditioned to the grade but to the participation in the final exam one week later. Directly after this subjects were asked to fill in the informed consent document and their personal data. Right after this, a pre examination took place (only in the pretest groups). Here they were told to answer all the questions as best as they could even if they did not know the answer. Afterwards they were told that they were going to see a video of approximately 15 minutes and that they should pay attention to the details of the video due that there was an exam immediately afterwards. After watching the video they got an exam to fill in and feedback was given directly afterwards (only in the feedback groups). This was done through projecting the post video test in the room screen with all the correct answers highlighted in red. No oral explanation was done by the experimenter about this document. Finally, once again they were told that after one week there was going to be a final test and that they should do it in order to get the six euro payment. It was also explained that the economical reward is not conditioned to the grade but to the participation in the final exam one week later. After one week participants gathered again in a room in where they sat down and were given the post week test to fill in. Once they finished doing this they handed in the paper and were thanked for their participation. In some cases, if the participant could not attend to the retest session, the post week retest was done through e-mail. All payments were given at a convened place and date within the universities. None of the groups was larger than 30 persons per round. The approximate time length of the procedure across all the conditions was between 30 to 45 minutes.

Materials

In this study the materials used were: 1) A video of Safety at Heights (from Coastal International, dubbed to Latin American Spanish), 2) a test in three formats (each format had the order of the

questions and answers randomly changed) with answer sheets, 3) a computer and 4) a projector.

The selected video explains in 15 minutes different subjects of safety while working at heights. These subjects were categorized as: 1) General knowledge of safety at heights, 2) Safety Railings, 3) Harnesses and Personal Protection Equipment (PPE), 4) Hook Points for harnesses and 5) Safety Floor Covers. In the video each of these topics is covered in the mentioned order with a final brief review of all the information.

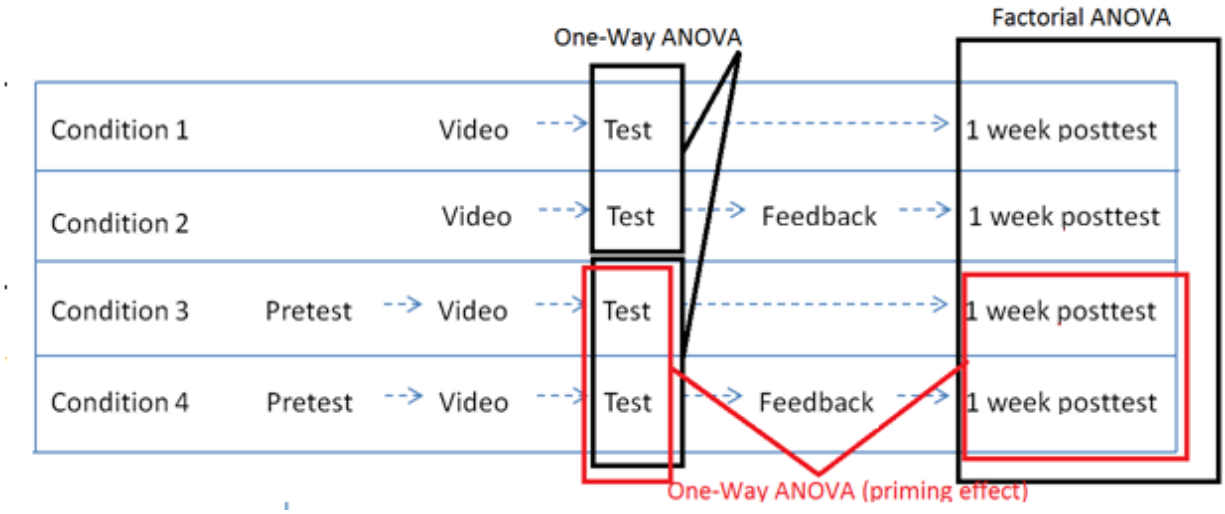
The test was a multiple choice test created from the video content. Its questions and answers were randomized to create two more formats (one format became the pretest, another the after video test and the last the after week posttest). This is done to avoid rote rehearsal retention generated from redoing the exams multiple times. Apart from randomizing the questions and answers, the main difference from these formats was that the pretest instructions indicated to answer the questions according to past experience. In the other formats it said to answer according to what was watched in the video. Also, the focus subject of safety floor covers is replaced with firefighting questions. This allowed observing if it is possible to prime acquisition and retention of some topics through pretests, in the short and long term, in the experimental conditions. These tests count with ten written questions and five multiple choice options per question. Something to remark from these tests is that the questions and answers are not solvable by simple logic. Therefore is should not be possible to solve the test without previous training (e.g. what is the weight that a safety rail should withstand and what height should it be?).

Results

The experiment data was analyzed using three different methods. The first one was a repeated measures ANOVA to check if forgetting was significant within the subjects. The second analysis was 2 X 2 factorial ANOVA. This was done to check for effects of Feedback, Pretest and their interactions in the

long term retention. In addition, a One-Way ANOVA was used in the post video test results to compare the Pretest conditions with the non-Pretest to observe is they had significant differences. Feedback and Pretest included two levels: with and without Feedback and with and without Pretest. Finally, another One Way ANOVA was used to analyze the difference between the five different topics from the video to observe if retention can be primed by a pretest in the immediate and long term.

Table II



After checking for assumptions and outliers it was found that there were some extreme values. These were corrected to reduce the extreme values in order to decrease the disproportionality they create in the results. Following this the ANOVAs were executed.

It was found that there was significantly large differences within individuals in the post video test and post week test $F(1,78)=25.04$, $p<.000$, $MSe=1.33$, $\eta^2 = .24$, which indicates that the participants had forgoten information between the post video test ($M=7.34$, $SD=1.61$) and post week test ($M=6.44$, $SD=1.66$).

In the long term retention analysis it was found that Feedback was significant. Its F ratio was $F(1,78)=21.16$, $p<.000$, $MSe=2.20$, $\eta^2 = .21$, indicating significant large effect differences in the Feedback

conditions (M=7.20, SD=1.54) and the no Feedback conditions (M=5.68, SD=1.42). On the other hand, the results yielded no significant effects for the Pretest conditions and for the interaction between Feedback and the Pretest in the long term retention. However, even if the results in the long term retention did not produce significant results for Pretest, when no Feedback was given the means where higher when the Pretest was done (M=6.00, SD=1.59) than when it was not (M=5.38, SD=1.20).

In the immediate retention the Pretest was not significant which indicates that Pretesting does not have relevant effect over memory in the short term. Even though the results did not yield a significant outcome for this, it should be noted that when no Feedback was given and only a Pretest was done the means where higher when the Pretest was done (M=7.56, SD=1.47) than when it was not (M=7.12, SD=1.74).

In the case of priming the topics (General knowledge, Railings, Harnesses and PPE, Hook points and Floor Covers) in the long and immediate term, the yielded results were non-significant differences for the immediate and long term. This means that there was no evident priming in the short and long term retention.

Table III
Mean scores between groups for short and long term retention

	Post video test score	Mean	SD	Post week test score	Mean	SD
With feedback	With pretest	7.86	1.459	With pretest	7.14	1.621
	No pretest	7.75	1.372	No pretest	7.25	1.482
	Total	7.80	1.400	Total	7.20	1.537
No feedback	With pretest	7.25	1.446	With pretest	6.00	1.589
	No pretest	6.52	1.861	No pretest	5. 38	1.203
	Total	6.88	1.691	Total	5.68	1.422

Discussion

The experiment shows that Testing and afterwards giving Feedback seems to be the most beneficial enhancer for long term retention in a video training context. This finding is in line with the ones from Bangert-Drowns et al.(1991); Kulhavy & Stock, 1989 Pashler et al. (2005) & Butler et al.(2008). We observed that just by showing each solved question for a couple seconds helped the participants retain more answers in the long term. Specifically, giving Feedback in this way helped participants retain approximately 1.5 more answers from an overall of 10 questions.

It seems that Feedback operates as an enhancer because it allows correcting the wrong answers and gives the possibility to have a wider specter of correct answers to retain. This in turn would allow long term retention of more answers. In other words one could say it is easier to remember through time 10 correct answers rather than 7.

On the other hand, it should be noted that it is possible that Feedback in this experiment operated in a more complex way than just increasing the specter of possibilities of retention. More specifically, it might have been moderated by motivational factors. During the experimentation participants who did not receive any Feedback approached the experimenter, specifically in the Video-Test and Pretest-Video-Test groups, to ask for their grade and/or the answers of the test. This could show an intrinsic motivation for Feedback which in turn could moderate the attentional responses and thus enhance the long term retention of the groups observed in this experiment.

In regard to Pretesting, the results did not show any relevance for long and immediate term retention. This indicates that performing a pretraining evaluation, in this type context and with a multiple choice questionnaire, might not result useful. This is not in line with the results found by Rothkopf & Bisbicos (1967), Richland et al. (2009) & Presley, et al. (1990). Although, it should be noted that the mean from the groups that did not receive Feedback, the one that had a Pretest had a higher mean score than its counterpart in the immediate and long term.

This study also attempted to prime subjects in some topics of the video through the Pretest. This resulted in no significant differences for the immediate and long term retention. This does not necessarily mean that priming through a Pretest is not possible. This study used a video in which all content was very similar, the analyzed topics were all related to fall prevention equipment and procedures which might qualitatively not differentiate very much between each other. This in turn might have not allowed observing any differences in the results.

Limitations

The population that was used for this study was mainly highly educated. All participants were university degree students that most likely have higher abstraction capabilities than individuals with technical or high school degree education. This difference might influence the outcome of the study and therefore the findings might not apply for more concrete educated populations. It should be taken in account that safety training is normally part of a company training for all personnel, but most importantly for the workforce that is most exposed to the workplace risks. These persons normally have high school or technical education which in turn can be more concrete while learning (e.g. in Peru the Instituto Nacional de Estadística e Informática (INEI) in 2013, or translated to English: National Institute of Statistics and Informatics, indicated that by 2012 the Peruvian workforce market was composed by a

16.75% of university degree, 15.15% technical degree and 68.1% of primary and high school degree workers).

Also, the video used for this study is limited to its structure and content. It should be taken into consideration that the results from this study come only from one safety training video with a specific structure and content. The results found in this investigation might differ if the video varies in content and structure.

Further Study

This study used a safety video which explained only how to work safely at heights. It had visual examples and diagrams (e.g. a diagram of how a railing should be structured, a visual example of a worn off harness, etc.) which were used to explain its contents. This was possible due that safety in this activity is visual and concrete. However other safety topics might not be so easy to observe because they are less concrete (as for example electrical safety: explaining the difference of voltage, amperage, electrical resistance of materials and they create safety risks might be a more problematic task) and thus less visual. This might raise the difficulty to understand the subject and hence might create different results. It could be interesting for further studies to consider the level of abstractness as factor that might influence the results.

On the other hand, Kang, S.H.K., McDermott, K.B. and Roediger, H.L. (2007) showed in their study that using short answer questions with posttest feedback improved long term retention in comparison to the multiple choice questions. This study used only the second type. It could be interesting to use short free recall answers in contrast to the multiple choice ones to study the institutional video training retention.

Literature

Bangert-Drowns, R. L., Kulik, C. L. C., Kulik, J. A., & Morgan, M. (1991). The instructional effect of feedback in test-like events. Review of educational research, 61(2), 213-238.

Barnett, S. M., & Ceci, S. J. (2002). When and where do we apply what we learn?: A taxonomy for far transfer. Psychological bulletin, 128(4), 612.

Bjork (Eds.), Handbook of perception and cognition: Memory (pp. 317–344). San Diego, CA: Academic Press.

Butler, A. C., Karpicke, J. D., & Roediger III, H. L. (2008). Correcting a metacognitive error: feedback increases retention of low-confidence correct responses. Journal of Experimental Psychology: Learning, Memory, and Cognition, 34(4), 918.

Butler, A. C., Karpicke, J. D., & Roediger III, H. L. (2007). The effect of type and timing of feedback on learning from multiple-choice tests. Journal of Experimental Psychology: Applied, 13(4), 273.

Carpenter, S. K., Pashler, H., Wixted, J. T., & Vul, E. (2008). The effects of tests on learning and forgetting. Memory & Cognition, 36(2), 438-448.

Craik, F.I.M. and Watkins, M.J. (1973) The role of rehearsal in short term memory. J. Verb. Learn. Verb. Behav. 12, 599–607.

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

Dempster, F. N. (1996). Distributing and managing the conditions of encoding and practice. *Memory*, *10*, 317-344.

Instituto Nacional de Estadística e Informática (2013) Perú: Evolución de los Indicadores de Empleo e Ingreso por Departamento 2004-2012. Retrieved 2015, from http://www.inei.gob.pe/media/MenuRecursivo/publicaciones_digitales/Est/Lib1105/libro.pdf

Kang, S.H.K., McDermott, K.B. and Roediger, H.L. (2007) Test format and corrective feedback modulate the effect of testing on memory retention. Eur. J. Cogn. Psychol. 19, 528–558.

Karpicke, J. D., & Roediger III, 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(4), 704.

Kulhavy, R.W. and Stock, W.A. (1989) Feedback in written instruction: the place of response certitude. Educ. Psychol. Rev. 1, 279–308.

Kulik, J.A. and Kulik, C.C. (1988) Timing of feedback and verbal learning. Rev. Educ. Res. 58, 79–97.

Logan, J. M., & Balota, D. A. (2008). Expanded vs. equal interval spaced retrieval practice: Exploring different schedules of spacing and retention interval in younger and older adults. Aging, Neuropsychology, and Cognition, 15(3), 257-280.

Metcalfe, J., Kornell, N., & Finn, B. (2009). Delayed versus immediate feedback in children's and adults' vocabulary learning. Memory & Cognition,37(8), 1077-1087.

Pashler, H., Zarow, G., & Triplett, B. (2003). Is temporal spacing of tests helpful even when it inflates error rates?. Journal of Experimental Psychology: Learning, Memory, and Cognition, 29(6), 1051.

Pashler, H., Cepeda, N. J., Wixted, J. T., & Rohrer, D. (2005). When does feedback facilitate learning of words?. Journal of Experimental Psychology: Learning, Memory, and Cognition, 31(1), 3.

Pressley, M., Tanenbaum, R., McDaniel, M. A., & Wood, E. (1990). What happens when university students try to answer prequestions that accompany textbook material? Contemporary Educational Psychology, 15, 27–35.

Pyc, M.A. and Rawson, K.A. (2009) Testing the retrieval effort hypothesis: does greater difficulty correctly recalling information lead to higher levels of memory? J. Mem. Lang. 60, 437–447.

Richland, L. E., Kornell, N., & Kao, L. S. (2009). The pretesting effect: Do unsuccessful retrieval attempts enhance learning?. Journal of Experimental Psychology: Applied, 15(3), 243.

Roediger, H.L., III and Karpicke, J.D. (2006a) The power of testing memory: basic research and implications for educational practice. Persp. Psychol. Sci. 1, 181–210.

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

Roediger, H. L., & Butler, A. C. (2011). The critical role of retrieval practice in long-term retention. Trends in cognitive sciences, 15(1), 20-27.

Roediger, H.L., III and Karpicke, J.D. (2010) Intricacies of spaced retrieval: a resolution. In Successful Remembering and Successful Forgetting: Essays in Honor of Robert A. Bjork (Benjamin, A.S., ed.), Psychology Press, pp. 23–47.

Rothkopf, E. Z., & Bisbicos, E. E. (1967). Selective facilitative effects of interspersed questions on learning from written materials. Journal of Educational Psychology, 58, 56–61.

Skinner, B.F. (1954) The science of learning and the art of teaching. Harv. Educ. Rev. 24, 86–97.

Smith, T.A. and Kimball, D.R. (2010) Learning from feedback: spacing and the delay-retention effect. J. Exp. Psychol. Learn. Mem. Cogn. 36, 80–95.

Toppino, T. C., & Cohen, M. S. (2009). The testing effect and the retention interval. Experimental Psychology (formerly Zeitschrift für Experimentelle Psychologie), 56(4), 252-257.

Van den Broek, G. S., Segers, E., Takashima, A., & Verhoeven, L. (2014). Do testing effects change over time? Insights from immediate and delayed retrieval speed. Memory, 22(7), 803-812.

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