

# Do Study Questions Support Long-Term Text Retrieval?

Huib Tabbers & Patrick den Hoed  
Erasmus University Rotterdam  
P.O. Box 1738, 3000 DR, Rotterdam, the Netherlands  
tabbers@fsw.eur.nl

**Abstract.** To investigate the value of the testing effect for educational practice, the long-term effects of study questions on text retrieval were investigated in a classroom setting. A geography text was studied by 107 pupils from first grade in a secondary school. Subsequently, they worked on study questions in three different formats: test question, look-up questions, and question&answer. A final test was given on both studied and non-studied questions either immediately afterwards, after 1 week or after 5 weeks. A testing effect was not found, as the results only showed a short-term benefit for look-up questions and questions&answers, and no long-term differences. The added value of study questions that lead to insufficient retrieval thus seems limited.

**Keywords:** testing effect; text; retrieval; questions; classroom.

The use of questions during study to support the learning of texts has a long history in educational psychology. Especially the use of so-called ‘adjunct questions’ to boost learning received a great deal of research attention in the 1970s and 1980s (for a review, see Hamaker, 1986). During the past 20 years however, interest had waned, partly due to inconsistent results and lack of a theoretical framework. Recently however, a ‘revival’ has been initiated by memory researchers who try to apply the so-called ‘testing effect’ to educational materials. The testing effect is the finding that an intermediate test benefits long-term retrieval more than restudying the material (for a review, see Roediger & Karpicke, 2006). Its theoretical rationale is that giving a test requires students to engage in a process of effortful retrieval, resulting in a memory representation that is less susceptible to decay than the memory representation constructed through additional study. Several researchers have applied the testing effect to text learning, and have demonstrated the benefits of asking questions after reading for long-term retrieval of the text contents (e.g., Agarwal, Karpicke, Kang, Roediger & McDermott, 2008; Chan, McDermott & Roediger, 2006; McDaniel, McDermott & Roediger, 2007). To translate these findings to educational practice, the current study investigates whether a similar testing effect can be found within the actual classroom. Moreover, the hypothesized benefit of retrieval effort is investigated in more detail by comparing retrieving answers from memory with just reading the answers and with finding the answers in the text. Also, performance on studied questions is compared with non-studied questions to see whether the benefit of testing transfers to new questions. Finally, multiple retention intervals are compared to investigate differences in forgetting rate.

## Method

### *Participants*

The sample consisted of 5 first grade classrooms from a secondary school, with a total of 107 participants (54 males and 53 females; age  $M=12.2$  years,  $SD=0.5$ ).

### *Materials*

*Text.* The text was taken from a geography book that was part of the first-year curriculum. The selected chapter dealt with some geographical aspects of the Netherlands as a country and was divided in three sections: ‘soil use’ (420 words), ‘transportation’ (389 words) and ‘holidays’ (401 words).

*Questions.* For each section of the text, 2 comparable sets of 5 questions were constructed that asked for literal text information. For example, the section on transportation contained a definition of the word ‘infrastructure’, and the accompanying test question was ‘What is meant by *infrastructure*’. Also, 3 versions were created of each question: A *question-&-answer version* that presented both the

question as well as the answer to the question; a *look-up version* that presented both the question and the relevant text paragraph containing the answer; and a *test version* that was just the question without the answer. The final test contained all 30 questions in *test version* format.

### *Design*

A 3 x 3 mixed design was used with study-question format (test, look-up, question&answer) as within-subjects factor, and retention interval (immediate, 1 week, 5 weeks) as between-subjects factor. The text sections and questions were always presented in the same order (soil use – transportation – holidays). The order of the three study-question formats, and which set of questions was used as study questions and which set was used as non-studied questions were counterbalanced across participants. Participants within a classroom were randomly assigned to retention interval (immediate:  $n=34$ ; 1 week:  $n=35$ ; 5 weeks:  $n=38$ ) and counterbalancing condition. The dependent variables were final test performance on the studied set of questions and on the non-studied set of questions.

### *Procedure*

For each class, the experiment took place during three regular lessons. During the first lesson, each text section was presented on a separate page, and participants got 3 minutes to study each page. After having read the three text sections, participants worked on the study questions. For each text section, a set of 5 questions was presented on a separate page in one of the three study-question formats, and participants got 3x3 minutes to fill in or study the answers. Subsequently, one third of the participants received the final test (the immediate condition), whereas the others worked on a Sudoku puzzle. Participants got 10 minutes to write down their answers to all questions (both studied and non-studied set). Afterwards, participants were kindly asked not to discuss texts or questions during the next weeks. After one week, another third of the participants made the final test (1 week condition), and after five weeks, the remaining participants received their test for the first time (5 weeks condition).

### **Results**

Mean performance on the study questions in the *test version* was 28% correct ( $SD = 23$ ), and in the *look-up version* 66% correct ( $SD = 23$ ). Figure 1 shows the final test scores for each retention interval. Because of differences in final test performance between the different text sections, all test scores were converted into Z-scores and analysed with a 3x3 repeated measures ANOVA. For the studied questions, significant main effects were found for both study-question format,  $F(2,208)=14.74$ ,  $p<.001$ ,  $\eta^2=.12$ , and retention interval,  $F(2,104)=55.75$ ,  $p<.001$ ,  $\eta^2=.52$ . Also, a significant interaction was found between study question and retention interval,  $F(4,208)=8.84$ ,  $p<.001$ ,  $\eta^2=.15$ . For the non-studied questions, only a significant main effect of retention interval was found,  $F(2,104)=22.66$ ,  $p<.001$ ,  $\eta^2=.30$ . Post-hoc comparisons within each retention group showed that in the immediate group, performance in the question&answer and look-up conditions was significantly higher than in the test condition. In the 1-week group only performance in the question&answer condition was significantly higher, whereas in the 5-weeks group, no differences were found. A direct comparison of performance between the studied and non-studied questions in each condition showed an overall significant advantage for the studied questions, except for the test question condition in the immediate group, and the look-up condition in the 5 weeks group.

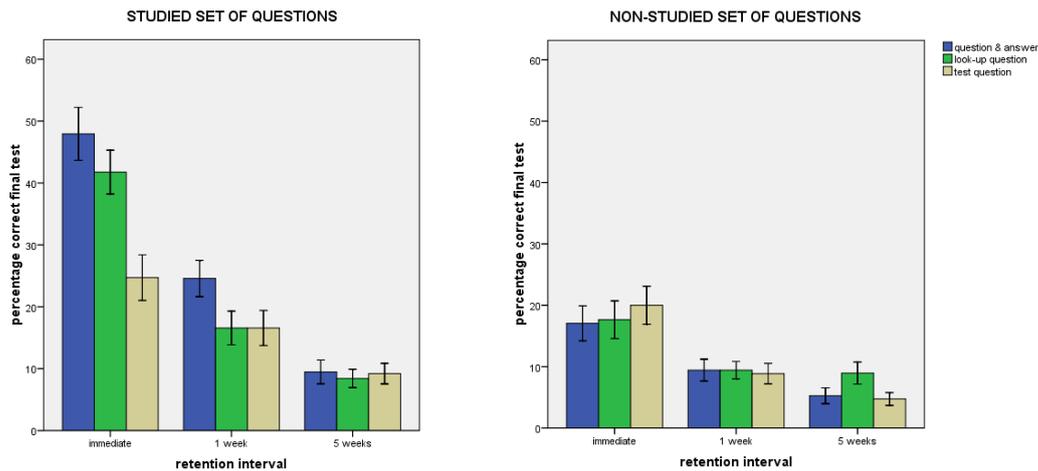


Figure 1. Mean percentage correct on final test for each condition on both studied questions (graph on the left) and non-studied questions (graph on the right). Error bars represent standard errors of the mean.

## General discussion

In sum, this study does not show a testing effect in terms of a long-term advantage for tested questions because of extra retrieval effort. Final test performance on the questions tested during study is not any higher than performance on questions for which answers were provided or for which the answers could be looked up. So on the one hand, the current study shows that the testing effect does not seem to transfer that easily to the classroom as is often implied. On the other hand, compared to prior testing effect studies, initial retrieval during study is quite low, which might have created floor effects in the test condition. Moreover, the interaction between retention interval and study question does seem to indicate different forgetting patterns. Thus increasing initial retrieval by extending study time and/or increasing learning by providing feedback might bring back the long-term benefits of testing. Based on the current study however, the practical implication is that giving students the opportunity to read or find out answers to study questions will benefit final test performance compared to study questions without feedback, although the effect is limited to a short interval and to identical test questions.

## References

- Agarwal, P. K., Karpicke, J. D., Kang, S. H. K., Roediger, H. L., & McDermott, K. B. (2008). Examining the testing effect with open- and closed-book tests. *Applied Cognitive Psychology, 22*, 861–876.
- Chan, J. C. K., McDermott, K. B., & Roediger, H. L. (2006). Retrieval-induced facilitation: Initially non-tested material can benefit from prior testing of related material. *Journal of Experimental Psychology: General, 135*, 553-571.
- Hamaker, C. (1986). The effect of adjunct questions on prose learning. *Review of Educational Research, 56*, 212–242.
- McDaniel, K. , McDermott, K. B., & Roediger, H. L. (2007). Generalizing test-enhanced learning from the laboratory to the classroom. *Psychonomic Bulletin & Review, 14*, 200-206.
- Roediger, H. L. & Karpicke, J. D. (2006). The power of testing memory. Basic research and implications for educational practice. *Perspectives on Psychological Science, 1*, 181-210.