



## The benefit of retrieval practice over elaborative restudy in primary school vocabulary learning



Nicole A.M.C. Goossens<sup>a,b,\*</sup>, Gino Camp<sup>b,a</sup>, Peter P.J.L. Verkoeijen<sup>a</sup>, Huib K. Tabbers<sup>a</sup>, Rolf A. Zwaan<sup>a</sup>

<sup>a</sup> Institute of Psychology, Erasmus University Rotterdam, Rotterdam, The Netherlands

<sup>b</sup> Welten Institute, Open University of the Netherlands, Heerlen, The Netherlands

### ARTICLE INFO

#### Article history:

Received 20 January 2014

Received in revised form 7 May 2014

Accepted 20 May 2014

Available online 10 June 2014

#### Keywords:

Elaborative exercises

Memory

Retrieval practice

Testing effect

Vocabulary learning

### ABSTRACT

The testing effect is the phenomenon that retrieval practice of learning material after studying enhances long-term retention more than restudying. We examined retrieval practice in primary school vocabulary learning in two experiments. Nine-year-old children studied word definitions and completed exercises according to three learning conditions: pure restudy, elaborative restudy or retrieval practice. Children in the pure restudy condition reread and partly copied the definitions. In the elaborative restudy condition children reread the definitions and connected semantically related words to the target words. Children in the retrieval practice condition recalled the words based on their definitions. Overall, on the fill-in-the-blank test after one week children in the retrieval practice condition outperformed children in the other conditions, but on the multiple-choice test there were no differences. Retrieval practice may be effective for primary school vocabulary learning, but there is uncertainty about the practical value and the magnitude of the retrieval practice effect.

© 2014 Society for Applied Research in Memory and Cognition. Published by Elsevier Inc. All rights reserved.

“Words are the tools we use to access our background knowledge, express ideas, and learn new concepts. The words children know will determine how well they can comprehend texts” (Stahl & Nagy, 2006, p. 4). Because words are so important, a considerable amount of time within the primary school curriculum is spent on teaching children vocabulary. A large variety of commercial vocabulary teaching programs have been developed in the last decades to support this considerable teaching endeavor, but many of these programs turned out to be unsuccessful (e.g., Blachowicz, Fisher, Ogle, & Watts-Taffe, 2006). It is therefore important to investigate whether strategies exist that can effectively augment vocabulary learning. Fundamental cognitive psychological research points at possible candidate strategies, but for many of these strategies the question is whether they generalize to classroom practice (Dunlosky, Rawson, Marsh, Nathan, & Willingham, 2013). In this article, we will investigate one strategy that holds considerable promise for classroom application, namely

the *testing effect* recently often relabeled as the *retrieval practice effect*, in the context of real-life primary school vocabulary learning.

When students engage in retrieval practice after an initial study phase, performance on a long-term memory test is better than when they study the same material twice (for a review, see Roediger & Karpicke, 2006). The *testing effect* or *retrieval practice effect* appears to be very robust. It has been observed in studies using word lists (e.g., Tulving, 1967; Wheeler, Ewers, & Buonanno, 2003), word pairs (e.g., De Jonge & Tabbers, 2013), or foreign vocabulary pairs as study material (e.g., Carpenter, Pashler, & Vul, 2006; Carpenter, Pashler, Wixted, & Vul, 2008; Carrier & Pashler, 1992; Karpicke & Roediger, 2008; Pashler, Cepeda, Wixted, & Rohrer, 2005; Toppino & Cohen, 2009). Furthermore, the retrieval practice effect has been replicated in studies in which people had to learn uncommon or infrequent words from their own language (e.g., Cull, 2000; Karpicke & Smith, 2012). Also, a few studies have reported a benefit of retrieval practice over restudy with primary school children (e.g., Bouwmeester & Verkoeijen, 2011b; Fritz, Morris, Nolan, & Singleton, 2007; Marsh, Fazio, & Goswami, 2012; Rohrer, Taylor, & Sholar, 2010).

However, to the best of our knowledge, only two studies have investigated retrieval practice in primary school vocabulary

\* Corresponding author at: Welten Institute, Open University of the Netherlands, P.O. Box 2690, 6401 DL Heerlen, The Netherlands.

E-mail address: [Nicole.Goossens@ou.nl](mailto:Nicole.Goossens@ou.nl) (N.A.M.C. Goossens).

learning. One study demonstrated that – compared to self-study – learning of definition-word pairs in sixth and seventh grade children was enhanced by using a computer program in which retrieval practice was included (Metcalf, Kornell, & Son, 2007). In another study third graders practiced twenty words and their synonyms (Goossens, Camp, Verkoeijen, & Tabbers, 2014). On the final cued-recall test after one week, word pairs learned by retrieval practice were recalled better than word pairs learned by restudy (47.0% versus 38.7%, respectively), which suggests that retrieval practice may improve vocabulary learning in children.

In the studies of Metcalfe et al. (2007) and Goossens et al. (2014) children learned word pairs in isolation. Yet, this is uncommon in classroom practice, which is characterized by children learning new words and their definitions in a meaningful context (e.g., Fuchs et al., 2003; Janssen & Van Ooijen, 2012; Van de Gein, Van de Guchte, & Kouwenberg, 2008). In the present study, we addressed this problem by examining whether retrieval practice benefits primary school vocabulary learning under conditions that mimic real-life vocabulary teaching more than the conditions in the studies of Metcalfe et al. (2007) and Goossens et al. (2014). One important feature of real-life vocabulary teaching is that children get acquainted with new words through a separate introductory learning session, which helps children focus on the word forms and word meanings. In several learning sessions after this initial learning session, children will practice the new words again. A second important feature of real-life vocabulary teaching is that repeated practice through ‘pure restudy’ (i.e., the exact repetition of words and their definitions) hardly – if ever – occurs. Instead, vocabulary lessons are characterized by repeated practice with new to-be-learned words in various meaningful exercises (e.g., Blachowicz et al., 2006; Fuchs et al., 2003; Janssen & Van Ooijen, 2012; Van de Gein et al., 2008). This more elaborative form of restudy is likely to lead to richer word representations than pure restudy. Indeed, previous research on vocabulary learning has shown that repetition in different contexts led to better memory for word meanings than repetition in a single context (e.g., Anderson & Reder, 1979; Bolger, Balass, Landen, & Perfetti, 2008; Carey, 1978; Coomber, Ramstad, & Sheets, 1986). Thus, to be of added value for classroom practice, retrieval practice should be more effective than elaborative restudy. Interestingly, Karpicke and Smith (2012) recently found that adults learning foreign vocabulary under conditions of retrieval practice had better long-term retention than under conditions of imagery or verbal elaboration, but whether this finding generalizes to vocabulary learning in the classroom is still an open question.

In the present study, we examined the effect of retrieval practice in primary school vocabulary learning in two experiments that were almost direct replications of each other. In each of the experiments, we incorporated the aforementioned features of real-life-vocabulary teaching. That is, all children received an introductory lesson before practice, and retrieval practice was not only compared to ‘pure restudy’, but also to ‘elaborative restudy’ using meaningful exercises based on textbook examples. In each experiment, children took a fill-in-the-blank test (in which they had to fill in the right word for a given definition), and a multiple-choice test (in which they had to choose the right word for a given context sentence) one week after the final learning session. The fill-in-the-blank test was comparable to the final tests used in earlier studies on retrieval practice (e.g., Karpicke & Smith, 2012), and was always administered first. We added the multiple-choice test for exploratory reasons, because this type of test is used very often in classroom settings. Based on earlier findings (Goossens et al., 2014; Karpicke & Smith, 2012), we hypothesized that retrieval practice would benefit vocabulary learning in the classroom compared to pure restudy and to elaborative restudy.

## 1. Experiment 1

### 1.1. Method

#### 1.1.1. Participants and design

One hundred forty seven nine-year-old children were recruited from six different classes of two primary schools. The children were from the Dutch Grade 5, which is equivalent to US Grade 3. Nine children were not given permission by their parents to participate, twelve were not able to participate during both learning sessions of the experiment, and four indicated they had difficulties understanding the instructions and their data were therefore excluded. This resulted in a sample of 122 participants (65 boys, 57 girls) with a mean age of 9.18 years (range 7.84–10.60, SD = 0.42). The children knew they participated in an experiment and their parents had given informed consent.

In this experiment, learning condition (pure restudy, elaborative restudy, and retrieval practice) was manipulated between subjects. From the 122 children that participated, 41 children were in the pure restudy condition, 42 children in the elaborative restudy condition and 39 children in the retrieval practice condition. The dependent variables were cued recall as measured by a fill-in-the-blank test and recognition as measured by a multiple-choice test, both administered one week after the learning sessions.

#### 1.1.2. Materials

The vocabulary words were selected from existing learning materials of the Dutch Grade 6 (Fuchs et al., 2003). The original learning material consisted of two stories that contained nine and eight target words. We excluded two words to have a final selection of fifteen words. See Table 1 for the Dutch words and their English translations. The median word frequency based on the Dutch Measure of Lexical Richness for primary school materials (Schrooten & Vermeer, 1994) was 3 (range 1–81), which is rather low.

**1.1.2.1. Introduction and exercises for the first learning session.** The target words were introduced to the children by a PowerPoint presentation and a booklet with exercises that focused on the definition and the word form. In the presentation, the fifteen words were presented each with a picture and a definition (e.g., *A pile in the garden with vegetable, fruit and garden waste, is called a compost pile.*). The booklet contained a list of the words and their definitions and two exercises. In the first exercise, children were presented with three lists of five target words and five definitions, and for each list they were instructed to connect the correct definitions with the correct target words by drawing a line. In the second exercise, children received each definition with a consonants-only cue, and they

**Table 1**  
Dutch words and their English translations.

Dutch word	English translation
composthoop	compost pile
kringloop	recycling
kunstmest	artificial manure
waterdamp	water vapor
milieuvuiling	pollution
milieuvriendelijk	environment friendly
smeltwater	meltwater
aluminium	aluminum
cement	cement
centrale	power station
dynamiet	dynamite
graniet	granite
ijzererts	iron ore
rots	rock
schacht	shank

Note. The English translations can deviate from the original Dutch meaning.

had to write down the correct target word (e.g., *A pile in the garden with vegetable, fruit and garden waste, is called a c.mp.st p.l.*).

**1.1.2.2. Exercises for the second learning session.** For each learning condition two exercises were created (for examples, see Appendix A–C). In the pure restudy condition, both exercises required children to copy a part of the definition and the target word (e.g., *A pile in the garden with vegetable, fruit and garden waste, is called a compost pile. – A pile in the garden with vegetable, fruit and garden waste, . . .*). In the elaborative restudy condition, the definitions and the target words were always presented together with the exercises. In the first exercise a word web was presented in which the target word (e.g., *compost pile*) was surrounded by six other words (e.g., *to manure, plastic, delicious, orange-peels, mailbox, and dead leaves*). In this exercise, the three words that were related to the word in the middle had to be selected by drawing a line to the target word. In the second exercise, word-lists with three words were presented (e.g., *dead leaves, orange-peels, ground*), and children had to write down the semantically related target word. In the retrieval practice condition, the definition was given and the target word had to be retrieved based on either a three-letter cue (first exercise) or a one-letter cue (second exercise) (e.g., *A pile in the garden with vegetable, fruit and garden waste, is called a com . . .*).

**1.1.2.3. The final tests.** In the fill-in-the-blank test, a definition was given and the children had to fill in the target word (e.g., *A pile in the garden with vegetable, fruit and garden waste, is called a . . .*). The multiple-choice test consisted of fifteen sentences in which the target word was left out (e.g., *John throws his fruit waste on the . . . He can use this later to manure his garden.*). For each set of five sentences, children had to pick the correct target words from a list of ten words that included five distractor words. For an example of the final tests, see Appendix D.

### 1.1.3. Procedure

The study took place in the classroom. One week prior to the experiment, the children completed a standardized vocabulary size test for Dutch primary school children (Verhoeven & Vermeer, 1993). This test consists of 50 sentences each containing one underlined word, for which the children had to select the best description out of four options. This test was used to match the three learning conditions on mean vocabulary size.

The actual experiment consisted of two learning sessions on two consecutive days, and one test session one week after the second learning session. In the first learning session, the experimenter started with the PowerPoint presentation. Each word was first presented with the corresponding picture, and two or three children were asked to guess the meaning of the word based on the picture. If the children did not succeed, the experimenter gave a short explanation of the word. Then the written definition was presented. If the definition was not clear, the children could ask questions. After all words had been introduced this way, the children did the two different exercises from the booklet. After all children had completed the exercises, the experimenter gave feedback in a group session. For each exercise, the experimenter first asked a child in the classroom to give the answer and then the correct answer was shown. After the children had checked their answers, the experimenter told them they would get two other exercises on the next day.

In the second learning session, the children received a booklet with for each word two exercises, varying according to learning condition. Before the children started, the experimenter explained one example of each type of exercise. After having completed the first exercise the children checked their own performance by comparing their responses to an answer sheet. They subsequently completed the second exercise and again checked their

**Table 2**

Proportion correct on the fill-in-the-blank test and the multiple-choice test in the three learning conditions (SD in parentheses) in Experiment 1.

Final test	Pure restudy (n = 41)	Elaborative restudy (n = 42)	Retrieval practice (n = 39)
Fill-in-the-blank	.75 (.22)	.72 (.23)	.82 (.15)
Multiple-choice	.92 (.14)	.91 (.16)	.92 (.10)

performance using another answer sheet. Whenever they had finished the two exercises, the children continued with school work that was not related to the vocabulary lesson.

In the test session one week later, the children first took the fill-in-the-blank test. After they finished this test, they received the multiple-choice test. When they finished both tests, they continued with their school work.

### 1.1.4. Data scoring

For the fill-in-the-blank-test and the multiple-choice test the maximum total score was fifteen points (one point for each correct answer). An answer was either correct or incorrect. Answers from the fill-in-the-blank-test that were phonetically similar to the intended word were scored as correct. Thus if a child made spelling mistakes we awarded points when it was clear that the right word was intended (e.g., *dyamite, dynamit or diamite* instead of *dynamite*).

Two independent raters scored twenty percent of the final fill-in-the-blank tests. The intraclass correlation between the two raters was .98, indicating a high interrater reliability. Because of the high agreement, one rater scored the remaining tests.

## 1.2. Results

A one-way ANOVA with learning condition (pure restudy, elaborative restudy or retrieval practice) as independent variable and vocabulary size as dependent variable did not reveal any differences between learning conditions on vocabulary size,  $F < 1$ , showing that the matching procedure had been successful. Furthermore, vocabulary size scores were positively correlated with the scores on the fill-in-the-blank test,  $r = .58, p < .001$ , and with the scores on the multiple-choice test,  $r = .42, p < .001$ . Therefore, we used vocabulary size as a covariate in our analyses of the final test results.

We analyzed the scores on the fill-in-the-blank-test and multiple-choice test separately by using a one-way ANCOVA with learning condition (pure restudy, elaborative restudy, retrieval practice) as independent variable, vocabulary size as covariate and score on the final test as dependent variable. For the unadjusted mean scores on both final tests see Table 2.

The one-way ANCOVA on the fill-in-the-blank test showed that the vocabulary size scores were significantly related to the scores on the fill-in-the-blank test,  $F(1, 118) = 57.63, p < .001, \eta_p^2 = .33$ . Table 2 shows that the children in the retrieval practice condition outperformed the children in the pure restudy and elaborative restudy condition on the fill-in-the-blank-test. However, this difference was not significant,  $F(2, 118) = 2.29, p = .106, \eta_p^2 = .04$ . The one-way ANCOVA on the multiple-choice test demonstrated that the vocabulary size scores were significantly related to the scores on the multiple-choice test,  $F(1, 118) = 24.89, p < .001, \eta_p^2 = .17$ . On the multiple-choice test, the children in all three conditions performed relatively well, and again, there were no significant differences,  $F < 1$ .

## 2. Experiment 2

Experiment 2 was conducted at the same time as Experiment 1. The only procedural difference between the experiments was in the way the words were introduced and practiced during the first

learning session. In Experiment 2 we did not introduce the words through a PowerPoint Presentation, but instead we introduced the words in the context of the two stories from the original materials. From a theoretical perspective there was no reason to assume that the relationship between final test scores and learning condition should differ for the two types of introduction. Therefore, we had the same hypotheses as in Experiment 1.

## 2.1. Method

### 2.1.1. Participants and design

In this study 131 nine-year-old children from five different classes of three primary schools participated. As in Experiment 1, the children were from the Dutch Grade 5, which is equivalent to US Grade 3. Nine of them were not able to participate during both learning sessions of the experiment. This resulted in a sample of 122 participants (44 boys, 78 girls) with a mean age of 9.10 years (range 8.01–10.36,  $SD = 0.45$ ). The children knew they participated in an experiment and their parents had given informed consent. The design of the experiment and the matching procedure was the same as in Experiment 1. From the 122 children that participated, 40 were assigned to the pure restudy condition, 40 to the elaborative restudy condition and 42 to the retrieval practice condition.

### 2.1.2. Materials and procedure

The same fifteen words were used as in the first experiment. In the first learning session, the words were introduced with two stories and two posters taken from the original learning materials. The posters were presented digitally on an interactive whiteboard, and illustrated the target words from the stories. When a poster was presented, the children were asked to explain what was shown on the poster. Then, the experimenter read aloud the story and explained the definitions of the words within the context of the story and by pointing at the illustrations. Afterwards, the children had to answer questions that were asked by the experimenter about the content of the story. When both stories had been presented and all questions had been answered, the children had to make exercises from a booklet. During these exercises the words were presented on the PowerPoint Presentation. For seven target words there was an exercise in which the syllables of the word were presented within the definition in the wrong order, and these syllables had to be rewritten in the right order (e.g., *water – melt* instead of *melt – water*). For the other eight target words the letters of the words were presented in a scrambled fashion within the definition in a circle, and had to be rewritten. The materials and procedure from the second learning session and the test session were identical to the materials and procedure in Experiment 1.

### 2.1.3. Data scoring

The scoring of both final tests was the same as in Experiment 1. Two independent raters scored all responses of the final fill-in-the-blank test. The intra-class correlation between the two raters was .99, indicating a high interrater reliability.

## 2.2. Results

A one-way ANOVA with learning condition (pure restudy, elaborative restudy or retrieval practice) as independent variable and vocabulary size as dependent variable showed that there was no significant difference between conditions on the mean vocabulary-size score,  $F < 1$ . This implies that our matching procedure was successful. Further analyses showed that vocabulary size scores were positively correlated with the scores on the fill-in-the-blank test,  $r = .49, p < .001$ , and with the scores on the multiple-choice test,  $r = .46, p < .001$ . Therefore, we used vocabulary size as a covariate for subsequent analyses.

**Table 3**

Proportion correct on the fill-in-the-blank test and the multiple-choice test in the three learning conditions (SD in parentheses) in Experiment 2.

Final test	Pure restudy ( $n = 40$ )	Elaborative restudy ( $n = 40$ )	Retrieval practice ( $n = 42$ )
Fill-in-the-blank	.69 (.22)	.70 (.22)	.83 (.14)
Multiple-choice	.91 (.11)	.92 (.10)	.91 (.16)

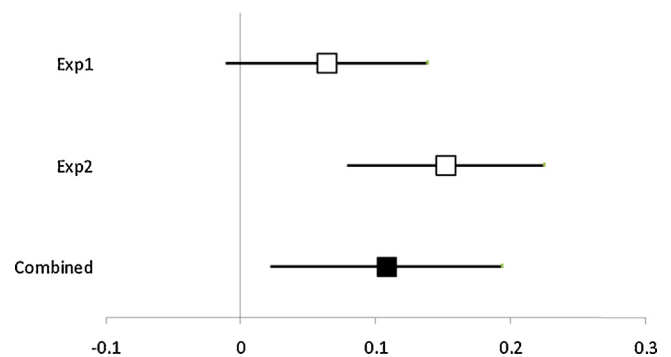
We analyzed the scores on the fill-in-the-blank-test and multiple-choice test separately by using a one-way ANCOVA with learning condition as independent variable, vocabulary size as covariate and the final test score as the dependent variable. For the unadjusted mean scores on both tests see Table 3.

The one-way ANCOVA on the fill-in-the-blank-test confirmed that the vocabulary size scores were significantly related to the fill-in-the-blank test,  $F(1, 118) = 45.89, p < .001, \eta_p^2 = .28$ . Furthermore, this time there was a significant effect of learning condition after controlling for vocabulary size,  $F(2, 118) = 10.11, p < .001, \eta_p^2 = .15$ . Planned simple contrasts revealed that the children in the retrieval practice condition outperformed the children in the pure restudy condition,  $t(80) = 4.02, p < .001, d = 0.88$ , and the children in the elaborative restudy condition,  $t(80) = 3.72, p < .001, d = 0.82$ . Thus, we found a benefit of retrieval practice on the fill-in-the-blank test.

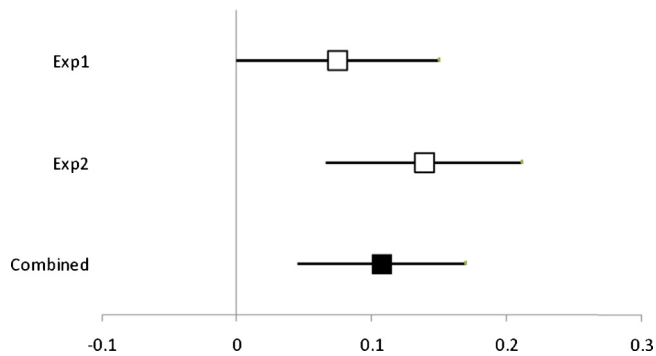
The one-way ANCOVA on the multiple-choice test confirmed that the covariate, the vocabulary test, was significantly related to the multiple-choice test,  $F(1, 118) = 31.54, p < .001, \eta_p^2 = .21$ . On the multiple-choice test, there was no significant effect of learning condition after controlling for the effect of the vocabulary test,  $F < 1$ . As the table shows the children in all conditions performed very well on this test.

## 3. Combined analysis of Experiment 1 and 2

In both experiments, the retrieval practice group outperformed the elaborative restudy group and the pure restudy group on the fill-in-the-blank test. However, this difference was statistically significant in Experiment 2, but not in Experiment 1. Following Cumming (2012, and see also Cumming, 2014) we calculated the 95% confidence intervals (CIs) of the adjusted mean difference between the retrieval practice condition and the pure restudy condition (Fig. 1), and between the retrieval practice condition and the elaborative restudy condition (Fig. 2). A positive point estimate reflects an advantage of retrieval practice over either pure restudy or elaborative restudy. The CIs in Fig. 1 show that (1) each of the experiments yields a rather imprecise estimate of the parameter of interest, (2) the point estimates of the two experiments are in the same direction, and (3) there is much overlap between the CIs.



**Fig. 1.** 95% CIs of the adjusted mean proportion difference between the retrieval practice condition and the pure restudy condition in Experiments 1 and 2 and a combined effect. A positive point estimate indicates a benefit of retrieval practice over pure restudy.



**Fig. 2.** 95% CIs of the adjusted mean proportion difference between the retrieval practice condition and the elaborative restudy condition in Experiments 1 and 2 and a combined effect. A positive point estimate indicates a benefit of retrieval practice over elaborative restudy.

The same applies to comparisons between retrieval practice and elaborative restudy in Fig. 2. All in all, Figs. 1 and 2 clearly indicate that the results of both experiments seem to reinforce rather than oppose each other.

To put this conclusion to the test, we conducted two small-scale random-effects analyses. For the difference between retrieval practice and pure restudy the combined parameter estimate of the adjusted mean difference is .108, 95% CI [.022, .195], which indicates a significant (in case of a two-tailed alpha level of .05) benefit of retrieval practice over pure restudy. In addition, for the difference between retrieval practice and elaborative restudy, the combined parameter estimate of the adjusted mean difference is .107, 95% CI [.045, .170], which indicates a significant advantage of retrieval practice over elaborative restudy. Thus combining the results from both experiments convincingly shows a benefit of retrieval practice over both restudy conditions. It should however be noted that the combined CIs are still rather wide. Consequently, although the effects are positive and deviate from zero, there is still much uncertainty about the magnitude of the retrieval practice effects in the population.

#### 4. Discussion

The aim of our experiments was to investigate whether retrieval practice can enhance primary school vocabulary learning in a relevant educational context. Although the benefit of retrieval practice on the final fill-in-the-blank test is only statistically significant in Experiment 2, our combined analysis indicates that the results of the two experiments reinforce each other and that it is therefore reasonable to consider the results in combination. Taken together, our experiments show a small advantage of retrieval practice over pure restudy and over elaborative restudy. So the results on the final fill-in-the-blank test seem to suggest that the positive effect of retrieval practice on vocabulary learning as shown in more controlled settings (e.g., Carpenter et al., 2008; Cull, 2000; Goossens et al., 2014; Karpicke & Smith, 2012; Pashler et al., 2005) indeed generalizes to a classroom setting. Additionally, our findings are consistent with other classroom experiments on text learning showing a clear benefit of retrieval-practice (e.g., Butler & Roediger, 2007; McDaniel, Agarwal, Huelser, McDermott, & Roediger, 2011; McDaniel, Anderson, Derbish, & Morrisette, 2007; Roediger, Agarwal, McDaniel, & McDermott, 2011).

Nevertheless, there are two points of concern. First, the relatively wide confidence intervals show that there is still much uncertainty about the magnitude of the benefit of retrieval practice in the population. To get a more accurate estimate of this magnitude, further research is warranted. Second, we did not find a benefit of retrieval practice on the multiple-choice test. As this is a

very common type of test in educational practice, it limits the practical usability of retrieval practice as a learning strategy. However, in our experiments, children performed very well on the multiple-choice test, so the effect may have been clouded by ceiling effects. Furthermore, the multiple-choice test was always preceded by the fill-in-the-blank test, so we should be careful in interpreting the findings on these tests.

Interestingly, in the present study elaborative study did not lead to a better performance on the vocabulary tests than pure restudy. Recently, Karpicke and Smith (2012) also failed to find a memory advantage of elaborative restudy over pure restudy using paired associates as stimulus materials. Earlier studies on vocabulary learning showed mixed results with some studies demonstrating positive effects of elaboration (typically in the form of adding contextual information to words in the learning phase) over identical repetition (e.g., Anderson & Reder, 1979; Bolger et al., 2008; Carey, 1978; Coomber et al., 1986) and other studies failing to observe positive effects of elaboration (e.g., Jones, Levin, Levin, & Beitzel, 2000; McDaniel & Pressley, 1984, 1989). One reason why we did not find an advantage of elaborative restudy over pure restudy may be that the tasks in the elaborative restudy condition were more dissimilar to the final test than the tasks in the other conditions. Possibly in our study it was more helpful in terms of final test performance to process the target word and its definition through pure restudy, rather than to elaborate on the target word. Hence, it may be that the elaborative restudy condition failed to outperform the pure restudy condition due to a lack of transfer-appropriate processing (e.g., Blaxton, 1989; Bouwmeester & Verkoeijen, 2011a; Morris, Bransford, & Franks, 1977; Thomas & McDaniel, 2007).

#### 5. Practical applications

The experiments in the present study have a relatively high ecological validity as the procedure in both experiments resembles primary school classroom practice in a number of important ways. Hence, our findings suggest that retrieval practice may be useful to primary school vocabulary teaching at least when the final test consists of providing the correct word to a given definition (i.e., the fill-in-the-blank test). Retrieval practice even resulted in a better performance than elaborative restudy. This may be relevant for educational practice, because often teachers try to improve vocabulary learning by using different elaborative exercises in their lessons. The present study suggests this may not be the most effective learning strategy.

However, our fill-in-the-blank results also showed there is considerable uncertainty about the magnitude of the retrieval practice effect in the population. Furthermore, we did not find any benefits of retrieval practice over restudy on the multiple-choice test, a type of test commonly used in primary schools to assess vocabulary knowledge. Yet, the latter finding may be due to ceiling effects. Thus, although our results suggest that retrieval practice may aid primary school vocabulary learning, additional research is needed to further explore the retrieval practice effect in primary school vocabulary learning.

#### Conflict of interest statement

The authors declare that they have no conflict of interest.

#### Acknowledgements

The Board of Public Education Rotterdam (Stichting BOOR) provided financial support for this research. The authors thank the directors and teachers of the schools for their cooperation. Furthermore, we thank the children for participating in this research.

Also, we thank Nathan Bonnevis and Rinske Nouwen for collecting the data. Additionally, we thank the educational and cognitive psychologists from the Institute of Psychology's publication group for their useful comments on an earlier draft of this paper.

## Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jarmac.2014.05.003.

## References

- Anderson, J. R., & Reder, L. M. (1979). An elaborative processing explanation of depth in processing. In L. S. Cermak, & F. I. M. Craik (Eds.), *Levels of processing in human memory* (pp. 385–403). Hillsdale, NJ: Erlbaum.
- Blachowicz, C. L., Fisher, P. J., Ogle, D., & Watts-Taffe, S. (2006). Vocabulary: Questions from the classroom. *Reading Research Quarterly*, 41, 524–530. <http://dx.doi.org/10.1598/RRQ.41.4.5>
- Blaxton, T. A. (1989). Investigating dissociations among memory measures: Support for a transfer-appropriate processing framework. *Journal of Experimental Psychology: Learning, Memory & Cognition*, 15, 657–668. <http://dx.doi.org/10.1037/0278-7393.15.4.657>
- Bolger, D. J., Balass, M., Landen, E., & Perfetti, C. A. (2008). Context variation and definitions in learning the meaning of words: An instance-based learning approach. *Discourse Processes*, 45, 122–159. <http://dx.doi.org/10.1080/01638530701792826>
- Bouwmeester, S., & Verkoijen, P. P. J. L. (2011a). The effect of instruction method and relearning on Dutch spelling performance of third-through fifth-graders. *European Journal of Psychology of Education*, 26, 61–74. <http://dx.doi.org/10.1007/s10212-010-0036-3>
- Bouwmeester, S., & Verkoijen, P. P. J. L. (2011b). Why do some children benefit more from testing than others? Gist trace processing to explain the testing effect. *Journal of Memory and Language*, 65, 32–41. <http://dx.doi.org/10.1016/j.jml.2011.02.005>
- Butler, A. C., & Roediger, H. L. (2007). Testing improves long-term retention in a simulated classroom setting. *European Journal of Cognitive Psychology*, 19, 514–527. <http://dx.doi.org/10.1080/09541440701326097>
- Carey, S. (1978). The child as word learner. In M. Halle, J. Bresnan, & G. A. Miller (Eds.), *Linguistic theory and psychological reality*. Cambridge, MA: The MIT Press.
- Carpenter, S. K., Pashler, H., & Vul, E. (2006). What types of learning are enhanced by a cued recall test? *Psychonomic Bulletin & Review*, 13, 826–830. <http://dx.doi.org/10.3758/BF03194004>
- Carpenter, S. K., Pashler, H., Wixted, J. T., & Vul, E. (2008). The effects of tests on learning and forgetting. *Memory & Cognition*, 36, 438–448. <http://dx.doi.org/10.3758/MC.36.2.438>
- Carrier, M., & Pashler, H. (1992). The influence of retrieval on retention. *Memory & Cognition*, 20, 633–642. <http://dx.doi.org/10.3758/BF03202713>
- Coomber, J. E., Ramstad, D. A., & Sheets, D. R. (1986). Elaboration in vocabulary learning: A comparison of three rehearsal methods. *Research in the Teaching of English*, 20, 281–293.
- Cull, W. L. (2000). Untangling the benefits of multiple study opportunities and repeated testing for cued recall. *Applied Cognitive Psychology*, 14, 215–235. [http://dx.doi.org/10.1002/\(SICI\)1099-0720\(200005\)06<14:3<215::AID-ACP640>3.0.CO;2-1](http://dx.doi.org/10.1002/(SICI)1099-0720(200005)06<14:3<215::AID-ACP640>3.0.CO;2-1)
- Cumming, G. (2012). *Understanding the new statistics effect sizes, confidence intervals, and meta-analysis*. New York: Taylor & Francis Group.
- Cumming, G. (2014). The new statistics: Why and how. *Psychological Science*, 25, 7–29. <http://dx.doi.org/10.1177/0956797613504966>
- De Jonge, M., & Tabbers, H. K. (2013). Repeated testing, item selection, and relearning: The benefits of testing outweigh the costs. *Experimental Psychology*, 60, 206–212. <http://dx.doi.org/10.1027/1618-3169/a000189>
- Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving students' learning with effective learning techniques: Promising directions from cognitive and educational psychology. *Psychological Science in the Public Interest*, 14, 4–58. <http://dx.doi.org/10.1177/1529100612453266>
- Fritz, C. O., Morris, P. E., Nolan, D., & Singleton, J. (2007). Expanding retrieval practice: An effective aid to preschool children's learning. *The Quarterly Journal of Experimental Psychology*, 60, 991–1004. <http://dx.doi.org/10.1080/17470210600823595>
- Fuchs, H., De Goei, B., Van den Heuvel, F., Geerlink, B., De Jong, H., & Kalkhoven, C. (2003). *Taal Actief Groep 6*. Den Bosch: Malmberg.
- Goossens, N. A. M. C., Camp, G., Verkoijen, P. P. J. L., & Tabbers, H. K. (2014). The effect of retrieval practice in primary school vocabulary learning. *Applied Cognitive Psychology*, 28, 135–142. <http://dx.doi.org/10.1002/acp.2956>
- Janssen, M., & Van Ooijen, M. (Eds.). (2012). *Taal Actief, Groep 4, 5, 6, 7, 8*. Den Bosch: Malmberg.
- Jones, M. S., Levin, M. E., Levin, J. R., & Beitzel, B. D. (2000). Can vocabulary-learning strategies and pair-learning formats be profitably combined? *Journal of Educational Psychology*, 92, 256–262. <http://dx.doi.org/10.1037/0022-0663.92.2.256>
- Karpicke, J. D., & Roediger, H. L. (2008). The critical importance of retrieval for learning. *Science*, 319, 966–968. <http://dx.doi.org/10.1126/science.1152408>
- Karpicke, J. D., & Smith, M. A. (2012). Separate mnemonic effects of retrieval practice and elaborative encoding. *Journal of Memory and Language*, 67, 17–29. <http://dx.doi.org/10.1016/j.jml.2012.02.004>
- Marsh, E. J., Fazio, L. K., & Goswick, A. E. (2012). Memorial consequences of testing school-aged children. *Memory*, 20, 899–906. <http://dx.doi.org/10.1080/09658211.2012.708757>
- McDaniel, M. A., Agarwal, P. K., Huelser, B. J., McDermott, K. B., & Roediger, H. L. (2011). Test-enhanced learning in a middle school science classroom: The effects of quiz frequency and placement. *Journal of Educational Psychology*, 103, 399–414. <http://dx.doi.org/10.1037/a0021782>
- McDaniel, M. A., Anderson, J. L., Derbish, M. H., & Morrisette, N. (2007). Testing the testing effect in the classroom. *European Journal of Cognitive Psychology*, 19, 494–513. <http://dx.doi.org/10.1080/09541440701326154>
- McDaniel, M. A., & Pressley, M. (1984). Putting the keyword method in context. *Journal of Educational Psychology*, 76, 598–609. <http://dx.doi.org/10.1037/0022-0663.76.4.598>
- McDaniel, M. A., & Pressley, M. (1989). Keyword and context instruction of new vocabulary meanings: Effects on text comprehension and memory. *Journal of Educational Psychology*, 81, 204–213. <http://dx.doi.org/10.1037/0022-0663.81.2.204>
- Metcalfe, J., Kornell, N., & Son, L. K. (2007). A cognitive-science based programme to enhance study efficacy in a high and low risk setting. *European Journal of Cognitive Psychology*, 19, 743–768. <http://dx.doi.org/10.1080/09541440701326063>
- Morris, C. D., Bransford, J. D., & Franks, J. J. (1977). Levels of processing versus transfer appropriate processing. *Journal of Verbal Learning & Verbal Behavior*, 16, 519–533. [http://dx.doi.org/10.1016/S0022-5371\(77\)80016-9](http://dx.doi.org/10.1016/S0022-5371(77)80016-9)
- 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, 3–8. <http://dx.doi.org/10.1037/0278-7393.31.1.3>
- Roediger, H. L., Agarwal, P. K., McDaniel, M. A., & McDermott, K. B. (2011). Test enhanced learning in the classroom: Long-term improvements from quizzing. *Journal of Experimental Psychology: Applied*, 17, 382–395. <http://dx.doi.org/10.1037/a0026252>
- 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. <http://dx.doi.org/10.1111/j.1745-6916.2006.00012.x>
- Rohrer, D., Taylor, K., & Sholar, B. (2010). Tests enhance the transfer of learning. *Journal of Experimental Psychology: Learning, Memory & Cognition*, 36, 233–239. <http://dx.doi.org/10.1037/a0017678>
- Schrooten, W., & Vermeer, A. (1994). *(Words in primary education. 15,000 words offered to pupils) Woorden in het basisonderwijs. 15.000 woorden aangeboden aan leerlingen*. Tilburg: Tilburg University Press.
- Stahl, S. A., & Nagy, W. E. (2006). *Teaching word meanings*. Mahwah, NJ: Lawrence, Erlbaum.
- Thomas, A. K., & McDaniel, M. A. (2007). The negative cascade of incongruent generative study-test processing in memory and metacomprehension. *Memory & Cognition*, 35, 668–678. <http://dx.doi.org/10.3758/BF03193305>
- Toppino, T. C., & Cohen, M. S. (2009). The testing effect and the retention interval: Questions and answers. *Experimental Psychology*, 56, 252–257. <http://dx.doi.org/10.1027/1618-3169.56.4.252>
- Tulving, E. (1967). The effects of presentation and recall of material in free-recall learning. *Journal of Verbal Learning and Verbal Behavior*, 6, 175–185. [http://dx.doi.org/10.1016/S0022-5371\(67\)80092-6](http://dx.doi.org/10.1016/S0022-5371(67)80092-6)
- Van de Gein, J., Van de Guchte, C., & Kouwenberg, B. (2008). *Zin in Taal Nieuw, A, B, C, D, E*. Tilburg: Zwijzen.
- Verhoeven, L., & Vermeer, A. (1993). *(Language proficiency test for ethnic minority children in grades 3 to 6) Taaltoets Allochtone Kinderen Bovenbouw*. Tilburg: Zwijzen.
- Wheeler, M. A., Ewers, M., & Buonanno, J. F. (2003). Different rates of forgetting following study versus test trials. *Memory*, 11, 571–580. <http://dx.doi.org/10.1080/09658210244000414>