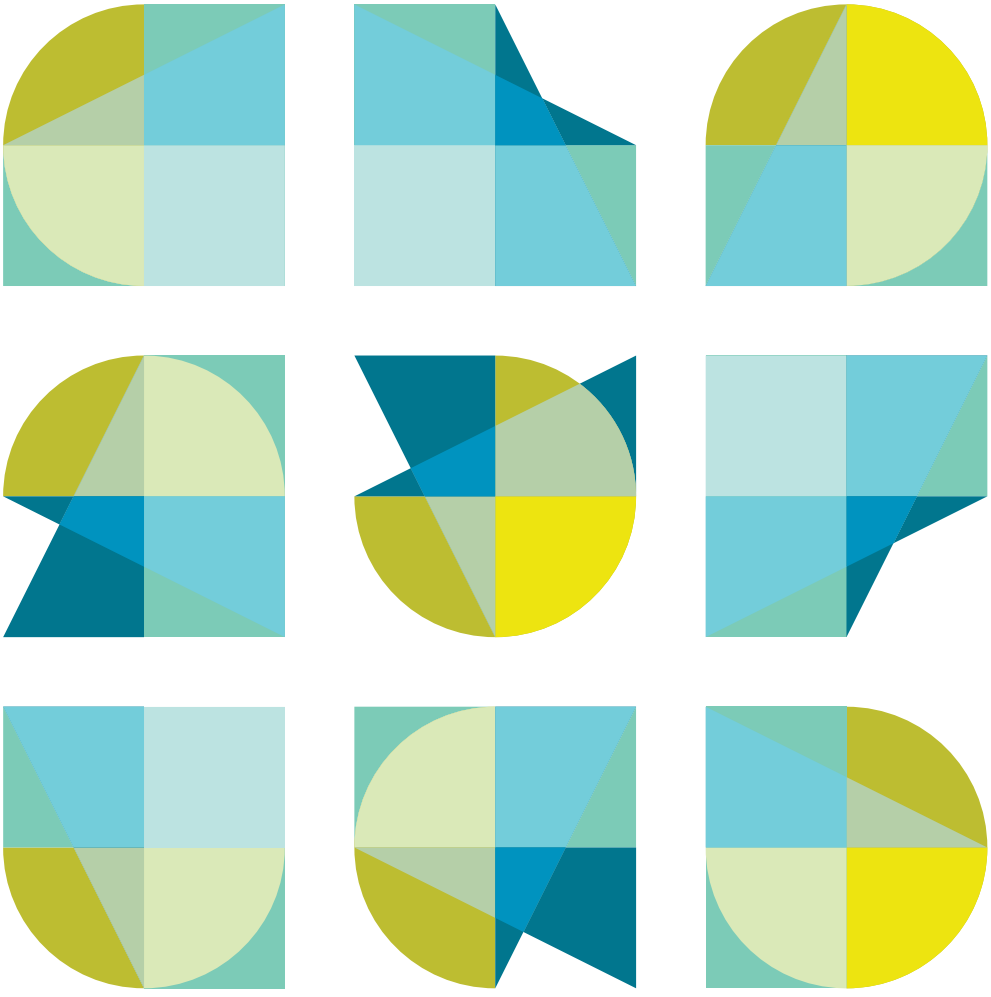
Distributed Practice and Retrieval Practice in Primary School Vocabulary Learning

Nicole Goossens



Distributed Practice and Retrieval Practice in Primary School Vocabulary Learning

Nicole Goossens

Distributed Practice and Retrieval Practice in Primary School Vocabulary Learning

Het spreiden van leermomenten en het ophalen uit het geheugen
in woordenschatonderwijs op de basisschool

Proefschrift

ter verkrijging van de graad van doctor aan de Erasmus Universiteit Rotterdam

op gezag van de rector magnificus
Prof.dr. H.A.P. Pols
en volgens besluit van het College voor Promoties.

De openbare verdediging zal plaatsvinden op
donderdag 22 januari 2015 om 15.30 uur

door

Nicole Anna Maria Catharina Goossens
geboren te Breda

Promotiecommissie

Promotor:	Prof.dr. R.A. Zwaan
Overige leden:	Prof.dr. L. Kester
	Dr. J.D. Karpicke
	Dr. D. Pecher
Copromotor:	Dr. G. Camp

ico

The research presented in this dissertation was carried out at Erasmus University Rotterdam in the context of the research school Interuniversity Center for Educational Sciences.

ISBN: 978-90-5335-997-6

Cover: Linda
Layout: Nikki Vermeulen, Ridderprint BV
Printing: Ridderprint BV

Copyright © Nicole Goossens, 2014
All rights reserved. No part of this dissertation may be reproduced or transmitted in any form, by any means, electronic or mechanical, without the prior written permission of the author, or where appropriate, of the publisher of the articles.

Contents

Chapter 1

General Introduction

Imagine you read the sentence 'In the *wusical ecters* are *minging* and *bancing*.' Probably you would not understand this sentence, because you do not know the words in italics. If you were given the meanings of these words, it would become easier to understand the sentence. In other words, if you would know that the words *wusical*, *ecters, minging,* and *bancing* stand for *musical comedy*, *actors*, *singing,* and *dancing* respectively, you would be able to understand the sentence completely.

Although the example given in the first paragraph may not seem representative for you as a reader, many people struggle with this type of problem. Especially, children who grow up in a linguistically deprived environment have problems with text comprehension. For example, if they do not understand the word *musical comedy*, it is impossible for them to understand the sentence as a whole. To help these children, it is important that researchers investigate how vocabulary learning can be enhanced. This thesis presents research about how primary school vocabulary learning can be enhanced by using effective memory strategies.

In vocabulary learning both the number of words that are known (quantity) and the level of word knowledge (quality) are important. The quality of word knowledge depends on how profound children know the words. According to Beck, McKeown, and Kucan (2002), there are four levels of word knowledge: (1) Does not know the word; (2) Has seen or heard the word before; (3) Recognizes or knows the word in a certain context; (4) Knows the word quite well, is able to explain and to use the word. The distinction between these levels of word knowledge will be explained by an example (*strawberry*). When the child has never seen or heard the word strawberry before, he will not know the word (level 1). If the mother of the child names and points at strawberries in the supermarket, the child will hear the word for the first time (level 2). If one week later, the child recognizes the strawberries in the supermarket, it is able to recognize the word in a certain context (level 3). If the child has seen strawberry fields at the horticulture, and has eaten strawberries, the child is able to explain and to use the word (level 4). Furthermore, the child is able to categorize the label *strawberry* and to build up a semantic network around the word (see e.g., Kuiken & Vermeer, 2005; Nagy & Herman, 1987; Verhoeven & Vermeer, 1996). Thus, the child knows that the strawberry is a sweet fruit (categorization), that is growing from a strawberry plant on the field or at the greenhouse, and that there are many other types of berries (network construction). At that moment, it will also become easier for the child to learn the words *agriculture, fruit bowl,* and *greenhouse*.

As said before, the quantity in word knowledge is also important. One important finding in vocabulary learning research is that there is a strong relationship between vocabulary size and reading comprehension (Anderson & Freebody, 1981). For example, it has been shown that readers need to have a vocabulary size of around 5,000 words to understand texts that are read for pleasure (Hirsch & Nation, 1992). The number of words

that have to be known for proper comprehension of a text, is also called text coverage. The estimates of how much text coverage is needed to have a proper comprehension, diverge from 83 to 95 percent (e.g., Goossens & Vermeer, 2009). Nevertheless, knowledge of a relatively small list of words can still give high percentages of text coverage. The 1,000 most frequent words can provide 70 to 80 percent text coverage for all texts (Alekseev, 1984). However, this is not enough to have a proper comprehension and this means that the difficulty lies within the other words that add the extra 10 to 20 percent to text coverage. For example, researchers have shown that if someone's vocabulary size increases from 2,000 to 11,000 words, text coverage will only grow with 9 percent (Hazenberg & Hulstijn, 1992). Thus, for adequate text coverage, relatively many additional words need to be learned. In one study with primary school children, Goossens and Vermeer (2009) showed that an optimum text coverage was reached at 88.7 percent; the children answered correctly more than half of the comprehension questions. This study also showed that the more difficult the words of the text, the less well the children understood the text and the fewer new words they learned from the text.

All in all, vocabulary knowledge is important for proper text comprehension. Also, vocabulary knowledge is important for learning new words. That is, children with normal to large vocabulary sizes have few difficulties to learn new words, because they relate new words to words they already know. However, children with smaller vocabulary sizes are not able to do that, and as a result, it will be very difficult for them to learn new words. The children with smaller vocabularies will encounter an increasing number of problems throughout the years, because they keep on falling further behind (Stahl & Nagy, 2006). In this way, the differences between children with bigger and smaller vocabulary sizes become larger over time. This is the so-called Matthew effect: the rich people become richer and the poor people become poorer (Stanovich, 1986). The question is how the Matthew effect can be reduced in vocabulary learning. Of course, this question cannot be answered very easily, but reconsideration of instructional strategies within vocabulary learning may help to answer this question.

Vocabulary Instruction Strategies

In The Netherlands, many vocabulary learning methods are based on the so-called *Viertakt* (Verhallen & Verhallen, 1994, and see for a description of this model also Verhallen, 2009). This didactical model for learning vocabulary is based on different theories of primary school vocabulary learning. For several years, researchers have tried to come to general (theoretical) guidelines for teaching vocabulary effectively and efficiently. Within these strategies are discovery strategies and consolidation strategies (e.g., Cook & Mayer, 1983; Nation, 1990; Schmitt, 1997). Discovery strategies refer to strategies that involve guessing the meaning of the word and consolidation strategies refer to strategies that

involve remembering the meaning of the word. The most general characteristics of good vocabulary instruction can be summarized as follows: (1) Vocabulary has to be learned in a language- and word-rich environment; (2) Students have to be taught to develop word-learning strategies themselves; (3) Words have to be taught by using multiple types of information and by repeated exposure (see Blachowicz, Fisher, Ogle, & Watts-Taffe, 2006). Related to the third characteristic of repetition, and more specific, in effective vocabulary it is important that children (a) are given a clear description of the meaning and the context of a word, (b) can process the word deeply by connecting new knowledge to their own knowledge, and (c) are getting more experience with the word (Stahl & Fairbanks, 1986).

There are many strategies in which these general characteristics of good vocabulary instruction are included, that can be used for repetition of the words (e.g., Schmitt, 1997). Examples of these strategies are relating strategies in which connection and grouping is important. For example, forming associations (e.g., Cohen & Aphek, 1981) and grouping the words in one meaning category (e.g., Cofer, Bruce, & Reicher, 1966; Craik & Tulving, 1975) enhance retention. Also, encoding and mental imagery strategies help to provide retrieval cues on the new word (e.g., Thompson, 1987). For example, in the keyword method the learner tries to link a new word with its meaning, by forming a mental image of the unknown word with the known word (e.g., Pressley, Levin & Miller, 1982). Another effective strategy is repeating the words in different contexts (e.g., Gipe, 1979, 1980; McKeown, 1985; Mondria, 1996). For example, children have to use the word in different contexts, or they have to discuss the meaning of the word in different sentences (Stahl & Nagy, 2006). All these strategies induce elaboration on the words, which will make the repetitions of the words richer and helps consolidation of the words. The question is how we could enhance this consolidation process. Therefore, an important goal of the research presented in this thesis is to investigate how this consolidation process can be strengthened any further. For example, in psychology many memory strategies are known that can be helpful for consolidation of words, and may be used in the classroom as well (Dunlosky, Rawson, Marsh, Nathan, & Willingham, 2013). However, these strategies are not yet used systematically and deliberately in the classroom. Thus, for many of these strategies there is a gap between research and practical use, and it is actually far from clear whether they are effective in the classroom. Therefore, in this thesis the question is whether primary school vocabulary learning can be enhanced by using memory strategies during consolidation.

The first reason for the focus on the consolidation phase in primary school vocabulary learning, is the fact that in many vocabulary learning methods the words are only repeated on the same day or in the same week, but not after a longer delay. Typically, after repetition of the words the children receive a test on the words and do not practice intentionally with the words again. However, research in cognitive psychology demonstrates that distributing

study sessions helps long-term retention (for a review see e.g., Cepeda, Pashler, Vul, Wixted, & Rohrer, 2006). The first question is therefore whether distributed practice can also enhance primary school vocabulary learning.

The second reason for the focus on the consolidation phase, is the fact that the exercises used in the classroom are relatively easy, because within these exercises there are many cues to do the exercises, without putting much effort in learning the words. The different context sentences that are used in the vocabulary exercises will enhance elaboration on the words, but from research in cognitive psychology it is known that it is better to use retrieval practice in learning sessions than to restudy the material for retention on the long term (for a review see e.g., Roediger & Karpicke, 2006). The second question is thus whether retrieval practice can also improve primary school vocabulary learning, where doing different elaborative exercises is the common practice.

Distributed Practice

The 'distributed practice effect' or 'spacing effect' refers to the phenomenon that distributed learning over time leads to better retention than massed learning (e.g., Dunlosky et al., 2013). There have been reported over 300 experiments into the effect of distributed practice (for reviews, see e.g., Cepeda, et al., 2006; Delaney, Verkoeijen & Spirgel, 2010). The effect has been found with many types of materials, as for example foreign vocabulary (e.g., Bloom & Shuell, 1981), trivia facts (e.g., Cepeda, Vul, Rohrer, Wixted, & Pashler, 2008), and texts (e.g., Rawson & Kintsch, 2005), but also with primary school children learning pictures (e.g., Toppino & DiGeorge, 1984), picture-word pairs (e.g., Cahill & Toppino, 1993) and word lists (e.g., Toppino & DeMesquita, 1984). Furthermore, the distributed practice effect has also been found in more realistic educational settings (e.g., Budé, Imbos, Van de Wiel, & Berger, 2011; Kornell, 2009; Seabrook, Brown, & Solity, 2005; Sobel, Cepeda, & Kapler, 2011). One relevant example of a study that showed a benefit of distributed practice, is a study in which eleven-year-old children had to learn eight unfamiliar English words during two learning sessions (Sobel et al., 2011). In these sessions, children learned the words by writing down the definitions and by making new sentences with the words. In the massed condition, the two learning sessions were separated by one minute, and in the spaced condition, the two learning sessions were separated by one week. On the final test after five weeks, a spacing effect was found, with children remembering 20.8% of the definitions of the distributed words and 7.5% of the massed words.

On the basis of Sobel and colleagues' (2011) study and other studies into the distributed practice effect in primary school contexts, it seems reasonable to predict that distributed practice benefits vocabulary learning in primary education. However, the conditions in many distributed practice experiments bear little resemblance to real primary school

vocabulary lessons. For example, the majority of distributed practice experiments has used massed learning intervals, spaced learning intervals and retention intervals which are considerably shorter than those in real educational settings. Also, in the vast majority of distributed practice studies the repeated learning sessions are exact copies of the first learning session. However, educators generally agree that vocabulary should be taught in an elaborated manner, thus the words should be rehearsed within different exercises (Blachowicz et al., 2006). Furthermore, in most of the distributed practice studies only word definitions were used to teach the words, while in primary schools more elaborative materials are used, as for example different context sentences, true/false questions and multiple-choice questions (e.g., Stahl & Nagy, 2006). Further, in earlier conducted studies, the words used were not from the current curriculum. To be able to generalize to normal vocabulary learning lessons, it would be better to use words that are from the regular curriculum. Thus, this thesis adds to the distributed practice literature by investigating the distributed practice effect in primary school vocabulary learning by using relevant educational material with different types of exercises, and with educationally relevant study intervals.

Retrieval Practice

The second memory strategy that will be investigated in this thesis is retrieval practice, also known as testing. The term 'testing effect' or 'retrieval practice effect' refers to the phenomenon that retrieval practice leads to better long-term retention of learning material than additional study (e.g., Dunlosky et al., 2013). Over a hundred experiments have been published about the beneficial effect of retrieval practice (for reviews, see e.g., Rawson & Dunlosky, 2011; Roediger & Butler, 2011; Roediger & Karpicke, 2006; Roediger, Putnam, & Smith, 2011). The retrieval practice effect has been investigated with word lists or word pairs (e.g., Carpenter, Pashler, & Vul, 2006; Toppino & Cohen, 2009), foreign vocabulary pairs (Carrier & Pashler 1992; Pashler, Cepeda, Wixted, & Rohrer, 2005; Pyc & Rawson, 2007), uncommon or infrequent words (e.g., Cull, 2000; Karpicke & Smith, 2012; Metcalfe, Kornell, & Son, 2007), trivia facts (e.g., Butler, Karpicke, & Roediger, 2008), facts in history (e.g., Carpenter, Pashler, & Cepeda, 2009), facts in science (e.g., McDaniel, Agarwal, Huelser, McDermott, & Roediger, 2011), locations on maps (e.g., Carpenter & Pashler, 2007; Rohrer, Taylor, & Sholar, 2010), and symbols (e.g., Coppens, Verkoeijen, & Rikers, 2011). Furthermore, the retrieval practice effect has been shown by using authentic classroom materials (e.g., Carpenter et al., 2009; Cranney, Ahn, McKinnon, Morris, & Watts, 2009; McDaniel, Anderson, Derbish, & Morrisette, 2007; Rawson & Dunlosky, 2011), and by using summative course assessments (e.g., McDaniel et al., 2011; McDaniel, Wildman, & Anderson, 2012). One relevant example of a study conducted in the field of retrieval practice is a study in which six- to thirteen-year-old

children had to learn categorical word lists that contained six critical words (Bouwmeester & Verkoeijen, 2011a). After studying one categorical word list, the children either received a free recall test to retrieve the words, or they received the word list again, to restudy the words again. After one week the children's memory was tested using a recognition test in which the children had to decide whether they had seen the word during the learning session. Overall, the children recognized more of the words learned by retrieval practice than by restudy.

So far, the retrieval practice effect has been found with foreign vocabulary and first language learning. However, in most of these studies only adults participated. The retrieval practice effect has not been investigated yet with primary school children learning new vocabulary. Also, in primary school vocabulary learning, words are not learned as word lists, but by using different materials (Stahl & Nagy, 2006) and within a meaningful context (e.g., Blachowicz et al., 2006). Thus, the control condition 'restudy' that is used in many retrieval practice effect papers is not comparable to the general learning situation in primary school vocabulary learning, because in primary school vocabulary learning there is much more elaboration. As a result, restudy is not a control condition that is ecologically valid. Also, the words used in many of the earlier conducted studies are uncommon, while in primary schools the children are learning ordinary words from the curriculum. Thus, this thesis adds to the current retrieval practice literature by investigating the retrieval practice effect in primary school children by using relevant educational material, presented within a meaningful context.

Using the Two Memory Strategies in Primary School Vocabulary Learning

In primary school vocabulary learning, it is important to focus on consolidation of the words. The aim of the studies in the present thesis is to investigate whether consolidation can be enhanced by using memory strategies such as distributed practice and retrieval practice. It is important to know whether the results of the previous studies can be generalized to a primary school vocabulary learning context, because in primary school vocabulary learning the learning situation is rather different. By investigating whether we can use distributed practice and retrieval practice in primary school vocabulary learning, the gap between theory and educational practice can be bridged. A recent review has shown that distributed practice and retrieval practice are very promising for retention of learning material, because the positive effects of both strategies have been shown in relatively many different tasks and contexts, in a number of classroom studies and in representative educational contexts, and with learners of different ages and abilities. Furthermore, the strategies are easy to implement in the classroom (Dunlosky et al., 2013). Therefore, it is predicted that both strategies will enhance primary school vocabulary learning in educationally relevant contexts. Thus, the studies in this thesis

investigate whether distributed practice and retrieval practice help primary school vocabulary learning, using current learning material in a real-life classroom situation. In these studies we will apply these strategies more systematically and more deliberately than in earlier conducted studies, to investigate whether these strategies can be utilized in primary school vocabulary learning and whether there are boundaries to the extent in which these strategies can be used in the classroom.

The Present Thesis

The present thesis has the following research questions: Could the memory strategies distributed practice and retrieval practice enhance consolidation of new words? If so, may these strategies also enhance consolidation of new words in the regular primary school vocabulary learning lessons? To answer these questions both the distributed practice effect (Chapters 2 and 5) and the retrieval practice effect (Chapters, 3, 4, and 5) are investigated with primary school children learning new vocabulary.

In Chapter 2 it was investigated whether distributed practice can enhance primary school vocabulary learning. In this study, nine-year-old children had to study fifteen words by using either a spaced study procedure and fifteen other words by using a massed study procedure. In the spaced condition the learning exercises were distributed over three sessions on three consecutive days and in the massed condition these learning exercises were combined into one session. A retention test was administered, both after one week and after five weeks, in which children had to retrieve the right word for the given definition.

In Chapter 3 it was investigated whether retrieval practice can enhance primary school vocabulary learning. In this study nine-year-old children had to study twenty words. First of all, the word-learning context in which the words were introduced, was manipulated. That is, there was a story condition in which the children first listened to a story in which the words were included, and there was a word pairs condition in which the children just heard the words with their synonyms. Then, the words were repeated by restudy or by retrieval practice. That is, ten new words and their meaning were repeated by using a retrieval practice procedure and ten other words were repeated by using a restudy procedure. After one week a cued-recall test and a multiple-choice test were administered to test children's retention of the meaning of the words.

In Chapter 4 it was also investigated whether retrieval practice can enhance primary school vocabulary learning. However, unlike in Chapter 3, the study was conducted within the classroom, and retrieval practice was not only compared to a pure restudy condition, but also to an elaborative restudy condition in which children performed exercises from the current vocabulary learning curriculum. Nine-year-old children were introduced to the words and made some exercises in which the definitions of the words were repeated

and in which the children had to focus on the word form. One day later, the children did two consolidation exercises according to three different learning conditions: pure restudy, elaborative restudy or retrieval practice. In the pure restudy condition the definitions of the words had to be partly copied, in the elaborative restudy condition the children had to connect semantically related words to the target words, and in the retrieval practice condition the children recalled the words based on their definition. After one week they received a fill-in-the-blank test in which they had to retrieve the word for the given definition, and a multiple-choice test in which they had to fill in the right word in the right context sentence.

Chapter 5 addressed both the distributed practice effect and the retrieval practice effect in seven- to thirteen-year-old children. The main goals of this study were to explore whether both effects are generalizable to different grades and whether these strategies enhance retention of vocabulary learned in the classroom. In this study the regular vocabulary learning materials were used and the regular procedure of the lessons was followed. The distributed practice effect was investigated by comparing a condition in which all words were learned in two lessons in the same week to a condition in which all words were learned in four lessons in two weeks. The retrieval practice effect was investigated by comparing a condition in which the children had to retrieve the definition of the word to a condition in which the children had to copy a part of the definition. At the end of each week the children received a cued-recall test in which they had to retrieve the definition of the word and several weeks after the experiment they received a multiple-choice test in which they had to recognize the right description of the word.

Chapter 6 summarizes the main findings of this thesis. These main findings will be discussed in the light of regular primary school vocabulary learning methods.

Chapter 2

Spreading the Words: A Spacing Effect in Vocabulary Learning

This chapter has been published as:

Goossens, N. A. M. C., Camp, G., Verkoeijen, P. P. J. L., Tabbers, H. K. & Zwaan, R. A. (2012). Spreading the words: A spacing effect in vocabulary learning. *Journal of Cognitive Psychology, 24*, 965-971. doi: 10.1080/20445911.2012.722617.

Abstract

The spacing effect refers to the frequently observed finding that distributing learning across time leads to better retention than massing it into one single study session. In the present study, we examined whether the spacing effect generalizes to primary school vocabulary learning. To this aim, children from Grade 3 were taught the meaning of 15 new words using a massed procedure and 15 other new words using a spaced procedure. The 15 words in the massed condition were divided into three sets of five words, and each set was taught three times in one of three learning sessions. In the spaced condition, learning was distributed across the three sessions: All 15 words were practiced once in each of the three learning sessions. At the retention tests after 1 week and after 5 weeks we observed that the meaning of spaced words was remembered better than the meaning of massed words.

Vocabulary size is a powerful predictor of reading comprehension. In fact, researchers agree that between 90% and 95% of the words in a text need to be known to arrive at an adequate reading comprehension of the text (Hirsch, 2003). Furthermore, reading comprehension will suffer if students' vocabulary does not grow sufficiently (e.g. Anderson & Freebody, 1981). Therefore, it is important to stimulate vocabulary development.

Blachowicz and colleagues have provided an overview of the characteristics of good vocabulary instruction (Blachowicz, Fisher, Ogle, & Watts-Taffe, 2006). The first characteristic is that words are learned in a language-rich and word-rich environment. Children should be encouraged to read, hear, and talk about new vocabulary in various ways. This means that children should not only learn vocabulary during specific vocabulary lessons, but also during other lessons, such as history and geography. The second characteristic is that words are taught intentionally during multiple exposures in which definitional and contextual information is provided. Thus, repetition of unknown words is important. The third characteristic is that children are taught to develop word-learning strategies themselves. This means that children should be encouraged to find their own strategy to learn an unknown word, by either using the context or using a part of a compound word to guess the meaning of the unknown word.

In this paper we will focus on the second characteristic of good vocabulary instruction; namely multiple exposures to the words. Many studies have been conducted to examine how often a word meaning has to be encountered in order to retain it. However, these studies have arrived at different estimates of the required number of exposures. For example, Nation (1990) suggested that five to 16 exposures within a context are sufficient to learn a word. However, not only the number of repetitions is important, but also how these repetitions are spread over time. For example, words can be repeated within one learning session, a procedure which we will call *massed* repetition, but it is also possible to distribute repetition across multiple learning sessions, a procedure which we will call *spaced* repetition. A robust finding that has emerged from cognitive psychological research is that spaced repetition leads to better retention than massed repetition, a phenomenon commonly referred to as the spacing effect (for a review, see Cepeda, Pashler, Vul, Wixted, & Rohrer, 2006). The question in this paper is whether spacing primary school vocabulary learning sessions will lead to a better retention of word meanings than massing these sessions.

There are several reasons to expect that spacing of vocabulary exercises will indeed benefit vocabulary learning in primary school children. First, the spacing effect is a robust finding that has been demonstrated in more than 300 published experiments with a variety of materials and with a variety of memory tests (for a review, see Cepeda et al., 2006). Second, the spacing effect has also been found with primary-school children. For instance, the spacing effect has been found in children who were learning pictures (e.g.,

Toppino & DiGeorge, 1984; Toppino, Kasserman, & Mracek, 1991), words (e.g., Toppino & DeMesquita, 1984), or a combination of pictures and words (e.g., Cahill & Toppino, 1993; Rea & Modigliani, 1987; Toppino, 1993). Third, a number of classroom studies have shown that spacing can be successfully employed in (foreign) vocabulary learning in undergraduate students (e.g. Bloom & Shuell, 1981; Kornell, 2009).

On the basis of these studies, it seems reasonable to predict that spacing should also augment vocabulary learning of primary school children in a real-world classroom/educational setting. At the same time, however, one could argue that the conditions in typical spacing experiments bear little resemblance to the conditions in primary school vocabulary lessons. For example, the majority of research on the spacing effect has used retention intervals which are considerably shorter than those in real educational settings. Also, the focus in many spacing studies is on memorizing *known* (unrelated) words. This focus is different from the vocabulary learning tasks in a real-world educational setting, which is directed towards the acquisition of new word meanings. Hence, the generalization of the findings from the spacing literature to classroom vocabulary learning might not be as straightforward as previously suggested.

It should be noted, however, that some studies have demonstrated a spacing effect in vocabulary learning in a real educational setting (e.g. Kornell, 2009; Sobel, Cepeda, & Kapler, 2011). In the study by Kornell (2009), undergraduate students studied 40 flashcards with each flashcard containing a word and its synonym (e.g., *effulgent – brilliant*). In the spaced condition, the participants studied one list of 20 word pairs during four sessions. Furthermore, within each session, the 20 word pairs were studied twice. In the massed condition, the participants studied one list of five word pairs during one of the four sessions. Within each session, the five word pairs were studied eight times. The final cued-memory test showed a spacing effect: participants in the spaced condition gave on average a correct response to 65% of the cues, whereas this percentage was merely 34% in the massed condition. Sobel et al. (2011) found similar results. In their study, fifth-graders had to learn eight unfamiliar English words (four massed, four spaced), during two identical learning sessions. These learning sessions consisted of a study-test-study-test sequence in which the children learned the words by writing down the definitions of the words and by making new sentences. In the massed condition, the second learning session was one minute after the first learning session and in the spaced condition, the second learning session was one week after the first learning session. Five weeks after the second learning session the children received a final test in which they had to write down the definitions of the words. The children recalled 20.8% of the spaced words and 7.5% of the massed words; this indicates that there was a spacing effect.

In the studies by Kornell (2009) and Sobel et al. (2011) the repeated learning sessions were exact copies of the first learning session. This might limit the external validity of

these studies, because educators generally agree that vocabulary should be taught by using different kinds of exercises during different learning sessions (Blachowicz et al., 2006). On a related note, one might even argue that the aforementioned studies would have found a much smaller spacing advantage, or perhaps no spacing advantage at all, if the researchers had varied the type of exercises between learning sessions. This argument is based on the finding that inducing encoding variability across repetitions can reduce the magnitude of the spacing effect (e.g., Dellarosa & Bourne, 1985; Gartman & Johnson, 1972). In the first experiment of Dellarosa and Bourne (1985) participants were presented with sentences that were repeated in a verbatim form (constant encoding) or a paraphrased form (encoding variability). Their second experiment was similar to the first with the only exception that sentences were repeated by the same speaker (constant encoding) or by a different speaker (encoding variability). Both experiments revealed a spacing effect in the constant-encoding condition, but not in the encoding-variability condition.

Furthermore, in the experiments of Gartman and Johnson (1972) participants had to learn homographs from lists with the same interpretation (*leg neck foot*, *arm hand foot*) or from lists with a different interpretation of the homograph (*leg neck foot*, *inch meter foot*). They found that the recall rate was higher when the context was different than when the context was the same, but that spacing did not have any influence on the recall rate in both lists. These experiments suggest that encoding variability can eliminate the spacing effect, but it is not clear whether this will also be the case in vocabulary learning.

However, other studies (e.g., Smith & Rothkopf, 1984) demonstrated that encoding variability does not affect the magnitude of the spacing effect. Hence, it is still somewhat unclear whether and, if so, under which conditions, encoding variability influences the spacing effect.

In the present experiment we investigated if there was a spacing advantage in vocabulary learning when children performed different types of exercises. We adopted the procedure of the experiments by Kornell (2009) to investigate if primary school children remember the meaning of new words better when they study them once on three consecutive days (spaced) than when they study them three times on one day (massed). In the present study, we used different exercises from current vocabulary learning material. The final tests after 1 week and after 5 weeks consisted of open-ended questions that required children to provide the correct word given its definition.

Method

Participants

We started out with 48 primary school children from Grade 3. However, only data from children who participated in all the sessions of the experiment were included in

the analysis, resulting in a final sample size of 33 participants. The mean age of these 33 children was 8.91 years (*SD* = 0.40). The children were recruited from two classes from a medium-sized primary school. This primary school was situated in an urban environment in Rotterdam. The children knew they participated in an experiment and their parents had given informed consent for participation.

Design and Materials

In this experiment, we manipulated the distribution of words within learning sessions (massed learning vs. spaced learning). Furthermore, we varied retention interval (1 week vs. 5 weeks) within subjects.

We selected 30 words and their exercises from current Grade 4 learning material. These words were presented in thematic sets of five words. Twenty-five words were nouns, four words were verbs and one word was an adverb. Twenty-two words were concrete and eight were abstract. The median word frequency based on the Dutch Measure of Lexical Richness for primary school materials (Schrooten & Vermeer, 1994) was 6, which is low. Most of the words consisted of two or three syllables: Thirteen words consisted of two syllables, thirteen words consisted of three syllables, three words consisted of four syllables and only one word consisted of one syllable. An example of a word set is *the musical comedy, the contribution*, *the platform*, *the scenery, to dine out* (Dutch: *de musical, het aandeel, het podium, het decor, dineren*).

The words were first presented with their definition and an accompanying phrase (e.g. *A musical comedy – A play in which actors sing and dance. – Every year, the children of Grade 6 perform a musical comedy.*). There were three types of exercises, taken from the learning material, which consisted of fill-in-the-blank questions (e.g., *Tonight we are going to watch a …*), true/false questions (e.g., *A musical comedy is a play in which actors are singing and dancing.*), and multiple choice questions (*What is a musical comedy? a. A wedding in which people are singing; b. A CD with music; c. A performance with songs.*). At the final test, memory for the words was tested using the definition as a cue (e.g. *A play in which actors are singing and dancing. – …*).

The 30 stimulus words were randomly split into two lists of 15 words each (List 1 and List 2). These lists were counterbalanced across the distribution conditions. Due to practical reasons (see the procedure below), the counterbalance sequence was nested within classroom. Specifically, all children in one class had to learn List 1 in the massed condition and List 2 in the spaced condition, while the children in the other class had to learn List 1 in the spaced condition and List 2 in the massed condition. At the final tests, we also balanced the list order. That is, half of the participants started the final test with List 1 and concluded the test with List 2, whereas the other half of the participants took the test in the reversed list order.

Procedure

The experiment took place in a classroom setting. At the start of the experiment, the children were told that they were going to learn some new words with the new vocabulary teacher (who in fact was the experimenter). In addition, they were told that they would take an unspecified test after a week. There were four learning sessions on 4 consecutive days and one test session 1 week after the final learning session. Four weeks after the first test session there was a second test session. In total in both conditions, three different exercises were performed for each item. The items were practiced in thematic sets of 5 items. The sequence of the words within each thematic set was different for the three exercises. For an overview of the procedure see Table 1.

Session 1 was an instruction session about the 30 vocabulary words. In this instruction session, the experimenter presented the words one by one in a PowerPoint presentation. For each word, children were first asked to provide a meaning themselves. Afterwards the experimenter gave a definition of the word and an accompanying phrase (e.g., *A musical comedy – A play in which actors sing and dance. – Every year, the children of Grade 6 perform a musical comedy*.). In general, the children failed to come up with correct word meanings, indicating that the words were new to them.

In Session 2, the children practiced 15 items (three thematic sets) in the spaced condition and five items (one thematic set) in the massed condition. In the spaced condition, children had to complete one exercise for each item. In a short break after this, the children did maths exercises and puzzles. The procedure was self-paced and the children received feedback when they all had completed their exercises. The experimenter told the children the correct answers, so that they could evaluate their own performance. In the massed condition, the children had to perform three different exercises in a row on the five items. After the first exercise, they had to do maths exercises and puzzles, after which they received feedback on their performance. This procedure was repeated for the second exercise and third exercise. Session 3 and 4 followed the same procedure as Session 2, except that in every session a different thematic set was used in the massed condition. After Session 4 the teacher did not rehearse the words, thus the words were not practiced in the classroom before Session 5.

In Session 5, 1 week after the fourth session, the children received a test consisting of open-ended questions that required them to write down the correct word in response to a given definition. During this final test, they were tested on all learned words. In Session 6, 4 weeks after Session 5, the children received the same test again.

Table 1 Procedure of the experiment

Session 1	Session 2	Session 3	Session 4	Session 5	Session 6
Presentation Items 1-30	*Items 1-15 Exercise 1*	*Items 1-15 Exercise 2*	*Items 1-15 Exercise 3*	Final Test Items 1-30	Final Test Items 1-30
	Items 16-20 Exercise 1	*Items 21-25 Exercise 1*	*Items 26-30 Exercise 1*		
	Items 16-20 Exercise 2	*Items 21-25 Exercise 2*	*Items 26-30 Exercise 2*		
	Items 16-20 Exercise 3	*Items 21-25 Exercise 3*	*Items 26-30 Exercise 3*		

Results

Two independent raters scored all responses of the first test given after 1 week. For each answer, the children received either one point or no point. One point was awarded to each answer that was either literally correct or phonetically correct. In all other cases, an answer received no point. The percentage correct answers was used as the test score. Pearson's correlation (*r*) between the two raters was .99, indicating a high interrater reliability. Because of the high interrater agreement on the first test, only one rater scored the second test after 5 weeks.

Performance on the Exercises of the Spaced and Massed Items

Before we did the analysis on the test results we assessed the performance on the exercises made during the learning sessions. We did this by scoring the responses of all exercises in the same way as we did for the final test. Afterwards, we analyzed the performance on the 90 exercises (45 massed and 45 spaced). In this analysis, three children had to be excluded because we could not trace their exercise booklets. Consequently, this analysis was based on thirty participants. The children performed well on the exercises (M = 87.07%, SD = 13.66).

Because all children learned the spaced words at the beginning of each learning session and massed words at the end, a waning concentration level might have led to an acquisition advantage of spaced words over massed words. To examine whether such an acquisition advantage had occurred, we compared performance on the exercises of the spaced and massed items. The children performed well on both massed and spaced items (massed: M = 86.81%, SD = 14.68; spaced: M = 87.33%, SD = 15.39). The difference between the massed and spaced conditions was not significant, t (29) = -0.226, p = .823, d = 0.042.

Furthermore, there was no significant difference in performance between the two participating classes, t (28) = -.891, p = .380, d = 0.166.

Comparison of Spaced and Massed Items

The mean percentage of correct recall of spaced and massed words after 1 week and after 5 weeks is shown in Table 2. We analysed the results with a 2 (learning condition) x 2 (retention interval) repeated measures ANOVA. The analysis revealed a main effect of condition, $F(1, 32) = 10.118$, $p = .003$, $\eta_p^2 = .240$: Retention in the spaced condition was better than retention in the massed condition. Also, there was an effect of retention interval, $F(1, 32) = 13.103$, $p = .001$, $\eta_p^2 = .291$. On average, children performed better on the words after 1 week than after 5 weeks. The interaction between learning condition and retention interval was not significant ($F < 1$).

Table 2 Mean percentage of correct recall of massed and spaced words (with *SD* in parentheses)

Learning Condition	Final Test After One Week	Final Test After Five Weeks
Massed	46.46% (25.85)	42.22% (23.07)
Spaced	55.96% (26.24)	49.49% (27.13)

Comparison of Spaced and Massed Items in Session 4

In the present experiment, the time between study session and final test session after 1 week differed for the spaced and massed condition. For the massed items, the time between the study session and the final test session after 1 week was 9, 8 or 7 days, while for the spaced items this was always 7 days. Only for the massed items studied in the last learning session (Session 4) was the time between the last learning session and the test the same as in the spaced condition. Thus, it could be that the larger interval between learning and test sessions for a subset of the massed items contributed to the difference in retention between spaced and massed items. Therefore, we compared retention of the massed items from Session 4 with the retention of all spaced items. For these items, the delay between learning sessions and test sessions was identical. The performance on the spaced items ($M = 55.96$, $SD = 26.24$; see also Table 2) was better than the performance on the massed items studied during Session 4 ($M = 46.06$, $SD = 29.78$), $t(32) = 2.406$, $p = .022$, $d = 0.39$. Thus, even when the delay between learning and test sessions was the same, spacing was better than massing.

Discussion

This study shows that distributing words across learning sessions is better than massing them into one learning session when children are learning vocabulary in a school setting. To our knowledge, this study was the first to investigate the spacing effect using different types of exercises during repetitions. In addition, it should be noted that the total spacing from the first to the last study session was the same in the spaced and massed condition

which makes the massed condition relatively more spaced than in other experiments (e.g. Kornell, 2009; Sobel et al., 2011). That is, both spaced items and massed items were already studied in the initial study session, so the massed items were not studied only on 1 day. This is novel, because in earlier spacing studies the massed items were learned during only one session. Taken together, our results indicate that spacing can benefit vocabulary learning in an educational context.

The present study extends the findings of Sobel et al. (2011) who found a spacing effect in vocabulary learning in fifth grade children. In their study, children did the same exercise in two learning sessions to learn word definitions. This means that there was no variability *within* the learning sessions. In our experiment there was variability within the learning sessions, because the children had to complete different exercises on the to-be-learned words. This kind of variability is common in vocabulary learning in the classroom. Hence, the results of the present study seem to be more informative to classroom practice than those of Sobel and colleagues.

However, there are some limitations to this study. First, since recall for spaced items was better after 1 week, the children had the opportunity to learn more spaced words than massed words, which may have produced a spacing effect after 5 weeks. Second, the spaced words were always studied first. Some studies found a better performance on the massed items than on the spaced items during practice (e.g. Karpicke & Roediger, 2007). In our study, we did not find this difference, which may indicate that spaced items gained some processing advantage due to being studied first. This in turn, might have resulted in a larger spacing effect at the final memory test as compared to the spacing effect in a completely balanced procedure.

A number of questions still remain. For example, there is the question if spacing could reduce the number of repetitions needed to retain a word. Nation (1990) argues that the required number of exposures to retain a word varies from five to 16, but it is not clear if spacing could reduce this required number. From the point of view that spacing helps to learn the words more effectively (i.e. spacing leads to a better retention of the words than massing, even if total exposure time is the same), we would also hypothesize that it helps to learn the words more efficiently (i.e. fewer exposures are needed in order to retain the words). Another question is at which spacing interval a maximum memory performance is obtained. Prior research (Cepeda et al., 2006) has shown that – somewhat extremely put – there is an inverted u-shape relationship between the spacing interval and memory performance. In addition, the optimal spacing increases with the length of the retention interval. Therefore, in future research directed at the optimalization of vocabulary learning it might be useful to examine the interaction between spacing and retention interval.

In short, in vocabulary learning in primary school children it is better to space the words during multiple learning sessions than to mass them during one learning session.

Acknowledgements

This research was financially supported by the Board of Public Education Rotterdam (Stichting BOOR). The authors would like to thank the director and teachers of the school, the children, and their parents for participating in this research.

Chapter 3

The Effect of Retrieval Practice in Primary School Vocabulary Learning

This chapter has been published as:

Goossens, N. A. M. C., Camp, G., Verkoeijen, P. P. J. L., & Tabbers, H. K. (2014). The effect of retrieval practice in primary school vocabulary learning. *Applied Cognitive Psychology, 28*, 135-142. doi: 10.1002/acp.2956.

Abstract

The testing effect refers to the finding that retrieval practice leads to better long-term retention than additional study of course material. In the present study, we examined whether this finding generalizes to primary school vocabulary learning. We also manipulated the word learning context. Children were introduced to 20 words by listening to a story in which novel words were embedded (story condition) or by listening to isolated words (word pairs condition). The children practiced the meaning of 10 words by retrieval practice and 10 words by restudy. After 1 week, they completed a cued-recall test and a multiple-choice test. Words learned by retrieval practice were recalled better than words learned by additional study, but there was no difference in recognition. Furthermore, the children in the word pairs condition outperformed the children in the story condition. These results show that retrieval practice may improve vocabulary learning in children.

Vocabulary learning has become one of the core components of language learning in the last 25 years (e.g., Vermeer, 2001). Already by the end of Grade 2, large differences exist in children's vocabulary size. On average, children will then have acquired around 6,000 word meanings, whereas children with a greater vocabulary knowledge have acquired around 8,000 word meanings and children with a lower vocabulary knowledge just 4,000 (Biemiller, 2005). Many studies have shown that vocabulary knowledge is a significant predictor of reading comprehension (e.g., Anderson & Freebody, 1981; Biemiller & Boote, 2006). Without sufficient knowledge of words it is difficult or even impossible to understand a text. Furthermore, a smaller vocabulary size of children has been found to be persistent throughout the school years and can harm later school success (e.g., Baker, Simmons, & Kame'enui, 1998; Hart & Risley, 1995; Nation, 2001).

The differences in vocabulary knowledge and the problems that arise from insufficient vocabulary knowledge indicate that there is a need for vocabulary instruction that is optimal for learning new words and their meaning. Some of the characteristics of good vocabulary instruction have been described by Blachowicz and colleagues (Blachowicz, Fisher, Ogle, & Watts-Taffe, 2006). A central characteristic is repetition: Words have to be practiced during multiple exposures. The central question in the present study is *how* the words have to be practiced during repeated study. In the literature two kinds of repetition have frequently been investigated, namely *restudy* and *retrieval practice* (for a review, see Delaney, Verkoeijen & Spirgel, 2010). More specifically, we wanted to investigate whether retrieval practice after initial learning can help children to learn new words more effectively than restudying. The instructional guideline to practice with retrieval is based on the so-called *testing effect*. The testing effect refers to the finding that retrieval practice enhances long-term retention more than additional study (for a review, see Roediger & Karpicke, 2006).

Most experiments on retrieval practice have been conducted using word lists or word pairs as study material (e.g., Carpenter, Pashler, & Vul, 2006; Carpenter, Pashler, Wixted, & Vul, 2008; Toppino & Cohen, 2009; Tulving, 1967; Wheeler, Ewers, & Buonanno, 2003). A classic example is an experiment by Webb (1921). In this study, participants studied 15 Hebrew words and their English equivalents for 5 minutes. For the next 3 minutes, half of the participants restudied the 15 word pairs, whereas the other participants received only the Hebrew words and had to retrieve their English equivalents (without getting feedback on their answers). In a similar cued-recall test given 1 week later, participants in the retrieval practice condition recalled twice as many English equivalents as participants in the restudy condition, hence clearly demonstrating the benefits of retrieval practice for long term recall performance.

Many studies have replicated these results in learning foreign vocabulary pairs (e.g., Carpenter et al., 2008; Carrier & Pashler, 1992; Karpicke, 2009; Karpicke & Roediger,

2008; Pashler, Cepeda, Wixted, & Rohrer, 2005; Pyc & Rawson, 2007, 2009, 2011; Toppino & Cohen, 2009) and in studies where people had to learn uncommon or infrequent words from their own language (e.g., Cull, 2000; Karpicke & Smith, 2012). For example, in the study of Karpicke and Smith (2012) participants studied 30 word pairs, each pair consisting of an uncommon English word and a one-word synonym. In the first part of the learning phase, all participants had to cycle through six study/recall periods in which they had to study and recall the word pairs during 7 seconds. When a participant correctly recalled the synonym of a word pair for the first time, the actual manipulation started. In the *drop* condition, the word pair was removed from the series of further study and recall trials. In the *repeated study* condition, the word pair was presented in two subsequent study trials. In the *repeated retrieval* condition, the word pair was presented in two subsequent retrieval trials. After 1 week, participants took a final cued-recall test in which they had to give the correct synonym in response to a provided English cue word. The results of this test showed that the participants in the repeated retrieval condition outperformed the participants in the two other conditions.

In sum, positive effects of repeated retrieval practice on long-term retention were found in both foreign vocabulary learning and first language learning. However, the described studies were conducted using only adult participants, and although the words were infrequent and uncommon, it was not clear whether participants had any prior knowledge about these words. Also, in most of these studies, words were studied without any context, whereas in primary school, new vocabulary words are always introduced within a meaningful context (e.g., Blachowicz et al., 2006). Therefore, the question remains whether the results from previous studies can be generalized to vocabulary learning in a primary school setting with young children.

A number of studies have reported positive effects of retrieval practice with primary school children. For example, Fritz, Morris, Nolan, and Singleton (2007) showed that preschoolers who learned names of toys recalled the names better after expanding retrieval practice than after expanding re-presentation or massed elaboration both after 1 minute and after 1 day. Further, in a study by Rohrer, Taylor, and Sholar (2010), fourth and fifth graders had to learn regions or cities on map locations by retrieval practice and by restudying. On the final test after 1 day, they received both an identical task and a transfer task on the learned material. On both tasks the children performed better after retrieval practice than after restudying. Marsh, Fazio, and Goswick (2012) showed that retrieval practice can benefit learning even for children of Grade 2, but only when they received feedback. Finally, Bouwmeester and Verkoeijen (2011a) compared restudying with taking an intervening free recall test with second to sixth graders learning Deese-Roediger-McDermott lists (Deese, 1959; Roediger & McDermott, 1995). One week later, the children were tested using a recognition test in which the children had to decide whether they had

seen the word in the learning session 1 week before. In this recognition test, critical lures of the Deese-Roediger-McDermott lists were also included. The results showed an overall benefit of retrieval practice over restudying. Bouwmeester and Verkoeijen (2011a) also found differences between children in how much they benefited from retrieval practice and that these differences were dependent on the amount of gist processing. Furthermore, the results showed no developmental trends in cognition; thus, the benefit of retrieval practice did not depend on the age of the children. In sum, these four studies showed benefits of retrieval practice with children, but these studies did not investigate primary school vocabulary learning in which new words have to be learned.

The central aim of the current study was to investigate whether retrieval practice is an effective strategy for improving vocabulary learning in primary school children. The first question was whether retrieval practice would indeed benefit long-term retention of new vocabulary words in primary school children. On the basis of the results from previous studies, our hypothesis was that retrieval practice after initial study would lead to better long-term retention of new vocabulary than additional study. A second question was whether presenting the words in a meaningful context would affect the hypothesized benefit of retrieval practice. In primary school vocabulary learning methods (e.g., Janssen & Van Ooijen, 2012; Van de Gein, Van de Guchte, & Kouwenberg, 2008), it is standard practice to present the words in a meaningful context. Therefore, we investigated whether retrieval practice is still effective in vocabulary learning in which a meaningful context is included. Earlier studies demonstrated that adding a (rich) context in itself can benefit vocabulary learning (for a review, see Stahl & Fairbanks, 1986). For example, in the study of Gipe (1979), four vocabulary learning methods were compared in third and fifth graders: an association method, a category method, a dictionary method and a context method. On the fill-in-the-blank test at the end of the week the children in the context condition performed better than the children in the other conditions. These results indicate that the use of context is helpful in vocabulary learning. The present study evaluated whether the retrieval practice effect has practical value in the classroom, where providing a context in vocabulary learning is standard practice.

Context was manipulated in the current study in the following way. In the *story* condition, children were introduced to the words by reading a story that included the target words. This method is similar to the way in which vocabulary is typically introduced in the vocabulary learning method we used. In the *word pairs* condition, children were introduced to the words by reading a list of the target words without any context. Subsequently, in both conditions, half of the words were repeatedly restudied, whereas the other half of the words were repeatedly studied by retrieval practice. As dependent variables, children's long-term retention and recognition of the target words were measured after 1 week.

Method

Participants

Originally, 62 third graders of three Dutch primary schools participated. As one of the children missed the session in which the vocabulary size of the children was tested and one of the children missed the second session, the data from 60 children (28 boys, 32 girls) were used. The children were aged 8 to 11 years (M = 9.24 years, SD = 0.49). The primary schools were situated in an urban environment in Rotterdam, and the children had diverse ethnic backgrounds. The children knew they participated in an experiment, and their parents had given informed consent for participation.

The vocabulary size of the children was tested with a normed test for Dutch primary school children in Grade 3 until Grade 6 (Leeswoordenschattaak, Taaltoets Allochtone Kinderen Bovenbouw, Verhoeven & Vermeer, 1993). The test consisted of 50 sentences containing an underlined word. For each sentence, children had to select the best description of the underlined word from four options. The Cronbach's alpha of this test for third to sixth graders is sufficient to good for group-wise comparisons of mean performance (α = .81, .83, .79, and .83, respectively). The participants in the two context conditions (story or word pairs) were matched on vocabulary size, resulting in 30 participants in the word pairs condition and 30 participants in the story condition. This procedure ensured that any differences between the two groups on the dependent variable were not caused by a priori differences in vocabulary size.

Materials and Design

A prose passage introducing 10 new words and a word list containing the synonyms of these 10 difficult words were selected from a grammar book typically used in Grade 5. For the purpose of the experiment, the prose passage was adapted by deleting some irrelevant sentences and adding sentences with 10 other difficult words. In this way, the final prose passage and the final word list consisted of 20 difficult words. The synonyms of the words were not included in the prose passage. The difficulty level of the words was measured using the Measure of Lexical Richness by computing the median word frequency (Schrooten & Vermeer, 1994; Vermeer, 2000). This median word frequency is a measure of the frequency of the word in several Dutch text books for primary school children (for a description, see Schrooten & Vermeer, 1994). Nineteen of the 20 words were in the list of Schrooten and Vermeer (1994). The median word frequency based on these 19 words was 3 (range 1 to 17), which is relatively low. The words were also pre-tested before the children started studying the words. Four children (6.7%) knew the synonyms of two words, and 10 children (16.7%) knew the synonym of one word. The other 46 children (76.7%) did not know any of the words. The Dutch words and synonyms and the English translations of the words and their synonyms are presented in Table 1.

Table 1 Dutch target words and synonyms and their English translations

Dutch target word	Dutch synonym	English target word	English synonym
apart	bijzonder	unique	special
baret	muts	a beret	a hat
beduusd	verrast	perplexed	confused
chaos	rommel	chaos	a mess
deponeren	gooien	to dispose	to throw
gift	cadeau	an offering	a present
heengaan	doodgaan	to pass away	to die
heimelijk	stiekem	secretly	sneaky
kris	mes	dagger	knife
kwiek	blij	briskly	happily
meedelen	vertellen	to describe	to tell
perplex	verbaasd	speechless	amazed
pronkstuk	mooi	showpiece	something beautiful
signaal	teken	a sign	a symbol
stug	stijf	rigid	firm
vaal	grauw	colourless	pale
vermoeden	idee	asssumption	idea
verprutsen	verknoeien	to make a mess of	to fail
wenen	huilen	to weep	to cry
weerzinwekkend	lelijk	repulsive	ugly

Note. The English translations can deviate from the original Dutch meaning, making the synonyms in English seem to fit less well with the targets.

In this experiment we used a 2 (context) x 2 (learning condition) mixed design. The context (story or word pairs) was manipulated between subjects. Thirty of the children first listened to a story in which the words were introduced (story condition), whereas 30 of the children were introduced to the words by just listening to the words and their synonyms without any context (word pairs condition). In both conditions the words were presented twice. The story is included in Dutch and in English in Appendices A and B, respectively. The learning condition was manipulated within subjects. After the introduction phase, half of the words was assigned to the restudy condition, in which the word pairs were studied seven times (SSSSSSS). The other half of the words was assigned to the retrieval practice condition, in which the word pairs were studied four times (e.g., *to weep – to cry*), retrieved once through cued recall (e.g., *to weep* - …), studied once more and then again retrieved (SSSSTST). In earlier studies, it was found that providing feedback after retrieval practice further strengthens the benefits of retrieval practice (e.g., McDaniel, Howard, & Einstein, 2009). Therefore we also included a restudy phase in the retrieval practice condition, namely the final 'S' in the SSSTST sequence, as a way of providing indirect feedback to the children.

The assignment of the two lists of words to the learning conditions was counterbalanced. Also, the order in which the restudy and retrieval practice lists were alternated was counterbalanced. The order in which items were presented in the different learning phases was randomized anew for each phase. The dependent variables were long-term retention and long-term recognition, as measured by the scores on a cued-recall test and the scores on a multiple-choice test. The cued-recall test was comparable to tests that are generally used in retrieval practice effect experiments on word-pair learning (e.g., Karpicke & Smith, 2012). In this test, the children had to orally retrieve the synonym of a given word (e.g., *to weep – …).* We scored the results on the final cued-recall test by only awarding points to synonyms that were phonetically identical to the synonyms that the children had learned during the learning phases. For each answer, the children received either one point or no point. The multiple-choice test consisted of a sentence in which one of the previously learned words was presented in a bold font along with four possible synonyms of the word (e.g., *Don't* ***weep*** *so much. – A. to pose; B. to cry; C. to talk; D. to throw).* On each test trial, one of the distractors was the synonym of another word that the children learned. For each test trail, children indicated which of the presented synomyms matched the bold word best. Both tests were administered after one week, which is a common delay at which long-term retention is measured in testing effect research (e.g., Roediger & Karpicke, 2006).

Procedure

The children were tested individually in a quiet room outside the classroom. The experiment consisted of two sessions. The first session was a learning session, followed by a test session 1 week later. The experiment was developed using E-Prime 2.0 and was presented on a laptop computer. The learning material was presented to the children on the screen, and the experimenter typed in the answers given by the children.

At the beginning of the learning session, the experimenter asked the children to give the meaning of each of the 20 new words and wrote down the answers of the children. After this pretest, the learning session consisted of seven phases. The first and second learning phase, in which the context was manipulated, took about 20 minutes altogether. The third to the seventh learning phase in which the word pairs were practiced took about 15 minutes altogether.

In the first learning phase, the experimenter told the children that they would first listen to a story or to a word list depending on the context condition (story or word pairs) and that after this they would be told the meaning of the words. Thus, at first, the experimenter read aloud the story or the word list depending on the context condition (story or word pairs) without any explanation of the words and without giving the synonyms of the words.

In the second learning phase, the experimenter read the story or the words to the children again, while the story (story condition) or the word list (word pairs condition) was presented on the screen. Thus, the children could read the story or the words while the experimenter was explaining the words. The experimenter gave the synonym of each word and explained its meaning by giving the synonym and a description of the word consisting of one sentence. Thus, in the story condition, after each sentence that contained a new word, the reading of the story was interrupted by the explanation of the new word.

In the third learning phase, the children in both conditions studied all 20 words. Every word and its synonym were presented for 8 seconds on the laptop (e.g., *to weep – to cry)*, in random order. The children had to read the words and their synonyms aloud.

In the fourth learning phase, the children studied all 20 words again to practice the words sufficiently before the manipulation of learning condition (restudy or retrieval) would start. After this learning phase, there was a short break of 2 minutes in which the children had to work in a puzzle book. The puzzle book included, for example, math exercises, to divert the children from the language task.

In the fifth learning phase, the children first restudied half of the words, and then, they were tested on the other half, or the other way around, depending on the counterbalancing condition. The words in the restudy condition were studied together with their synonyms, as in the previous learning phase. The words in the retrieval practice condition were shown for 8 seconds, and the children were asked to read it aloud and then try to retrieve its synonym (e.g., *to weep - …)*. After this phase, the children again had a short break of 2 minutes in which they continued with their puzzle book.

In the sixth learning phase, the children again restudied all 20 words once, with the same procedure as in the second learning phase. After this phase, the children again had a short break of 2 minutes in which they continued with their puzzle book.

In the seventh learning phase, the procedure was identical to the fifth learning phase. The children restudied the same 10 words and were tested on the same 10 words. After the children had finished this learning phase, they were asked not to talk about the words with their classmates, and they returned to their classroom.

In the test session, 1 week after the learning session, the children were given a cued-recall test in which they had to orally retrieve the synonym of a given word. After that the children completed a multiple-choice test consisting of a list of written sentences, with each sentence containing (presented in bold font) a word they previously learned. The children were instructed to read aloud each sentence and to select out of four options the synonym that matched best with the bold word. If the children did not know the answer on this test, they had to guess which answer they considered best. These tests were administered individually in a quiet room outside the classroom. There were no time constraints. After the children had finished the tests, they were asked not to talk about the final test with their classmates, and they returned to their classroom.

Results

First of all, we checked if our experimental groups were comparable with regard to vocabulary size by an independent sample *t*-test with the independent variable of learning condition (story or word pairs) and vocabulary size as a dependent variable for all 60 children. This analysis showed that there was no significant difference between the two context conditions on the vocabulary size measure, namely $t(58) = -1.12$, $p = .266$, $d = 0.29$. This means that our matching procedure, based on the vocabulary size of the participants, was appropriate. Further analyses showed that vocabulary size correlated positively significant with the score on the restudied items (.60) and on the retrieved items (.56) of the cued-recall test and also with the score on the restudied items (.42) and on the retrieved items (.47) of the multiple-choice test. Thus, the higher the score on the vocabulary size test, the higher the score on the cued-recall test and multiple-choice test. Therefore we used the *z*-score of this vocabulary size measure as a covariate in our analyses.

Learning Session Data

During the learning session there were two phases in which the children received a practice test on half of the words. Table 2 shows the mean scores on these tests for each context group. We analyzed these scores using a 2x2 mixed analysis of covariance (ANCOVA) with context as a between-subjects factor, retrieval-practice phase as a within-subjects factor and the *z*-score of vocabulary size as a covariate.

Table 2 Proportion correct on the initial tests in the first and second retrieval practice phase (*SD* in parentheses)

Retrieval practice phase	Word pairs condition (*n* = 30)	Story condition (*n* = 30)
First phase	0.52 (0.22)	0.41 (0.22)
Second phase	0.65 (0.22)	0.52 (0.22)

The mixed design ANCOVA on the practice tests during the learning sessions showed that the covariate, vocabulary size, was significantly related to the score on the two practice tests, $F(1, 57) = 55.19$, $p < .001$, $\eta^2 = .43$. With the covariate included in the model, there was a significant effect of retrieval practice phase, $F(1,57) = 53.02$, $p < .001$, $\eta^2 = .48$. Children recalled more synonyms during the second retrieval practice phase than during the first retrieval practice phase (first: $M = 0.46$, $SD = 0.23$; versus second: $M = 0.58$, $SD = 0.23$). There was no significant interaction between retrieval practice phase and vocabulary size, $F < 1$. Also, there was a significant main effect of context, $F(1,57) = 16.20$, $p < .001$, $\eta^2 = .13$, indicating that children in the word pairs condition recalled more synonyms in the retrieval practice phases than children in the story condition (word pairs: $M = 0.58$, SD

= 0.21; versus story: M = 0.47, SD = 0.21). There was no significant interaction between retrieval practice phase and context, $F < 1$.[1]

Test Session Data

Table 3 shows the results on the final cued-recall test and Table 4 shows the results on the final multiple-choice test. We analyzed the results on both tests using a 2x2 mixed analysis of covariance (ANCOVA) with context as a between-subjects factor, learning condition as a within-subjects factor and z-score of vocabulary size as a covariate.

Table 3 Proportion correct on the cued-recall test in the word pairs condition and story condition for restudied and retrieved words (*SD* in parentheses)

Cued-Recall Test	**Word pairs condition (n = 30)**	**Story condition (n = 30)**
Restudied words	0.41 (0.22)	0.36 (0.20)
Retrieval practice words	0.53 (.23)	0.41 (0.20)

Table 4 Proportion correct on the multiple-choice test in the word pairs condition and story condition for restudied and retrieved words (*SD* in parentheses)

Multiple-Choice Test	**Word pairs condition (n = 30)**	**Story condition (n = 30)**
Restudied words	0.82 (0.15)	0.71 (0.22)
Retrieval practice words	0.83 (0.16)	0.73 (0.18)

The mixed design ANCOVA on the cued-recall test showed that the covariate, the vocabulary test, was significantly related to the score on the cued-recall test, F (1, 57) = 48.36, p <.001, η^2= .42. With the covariate included in the model, there was a significant effect of learning condition, F (1, 57) = 15.03, p < .001, η^2= .20. There was no significant interaction between learning condition and vocabulary size, $F < 1$. There was a significant main effect of context, F(1,57) = 10.25, p = .002, η^2= .09, indicating that children in the word pairs condition recalled more synonyms than children in the story condition (word pairs: M = 0.47, SD = 0.21, versus story: M = 0.39, SD = 0.19). There was a marginally significant interaction between learning condition and context after controlling for the effect of the vocabulary test, F(1,57) = 2.99, p = .089, η^2= .04. It seems that the beneficial effect of retrieval practice in the story condition is less strong than in the word pairs condition.

The mixed design ANCOVA on the multiple-choice test showed that the scores on the covariate, the vocabulary test, were significantly related to the scores on the multiple-choice test, F (1, 57) = 34.02, p <.001, η^2= .31. On the final multiple-choice test, there was no significant main effect of learning condition after controlling for the effect of vocabulary

1 The careful reader might note that the different values of η^2 sum up to more than 1. However, this occurs because effect sizes are calculated separately for the between-subjects variables and the within-subjects variables.

size, F <1. There was a significant main effect of context after controlling for the effect of vocabulary size, $F(1,57) = 18.46$, $p < .001$, $\eta^2 = .17$, indicating that children in the word pairs condition recognized more synonyms than children in the story condition (word pairs: $M = 0.83$, $SD = 0.14$, versus story: $M = 0.72$, $SD = 0.16$). There was no interaction between learning condition and context, F<1.

During scoring of the final cued-recall test, we noticed that many children had produced synonyms that were incorrect but semantically similar to the synonyms presented in the experiment. Thus, we performed an additional analysis in which we applied a more liberal scoring method, counting semantically similar synonyms as correct. Using this more liberal scoring method, we obtained the same pattern of results as with the more strict scoring method.

Furthermore, we did an additional analysis on the final cued-recall test in which we excluded the words that the children already knew at the pretest from the analysis. Again, we obtained the same pattern of results as in the scoring method in which also known words were included (Table 5).

Table 5 Proportion correct on the cued-recall test in the word pairs condition and story condition for restudied and retrieved words (*SD* in parentheses) when words already known at the pretest were excluded from the analysis

Cued-Recall Test	Word pairs condition (n = 30)	Story condition (n = 30)
Restudied words	0.40 (0.22)	0.36 (0.20)
Retrieval practice words	0.53 (0.23)	0.40 (0.20)

Discussion

Our main research question was whether retrieval practice benefits vocabulary learning in primary school children. The results of the final cued-recall test showed that there was a benefit of repeated retrieval practice on the long-term when compared with repeated study. Children recalled more word synonyms that they had retrieved during learning than word synonyms that they had restudied. To our knowledge, this is the first study that showed a benefit of retrieval practice in primary school vocabulary learning. This study extends the earlier findings from adult vocabulary learning regarding the positive effect of retrieval practice (e.g., Carpenter et al., 2008; Karpicke & Smith, 2012) to primary school vocabulary learning. The positive effect of retrieval practice we found in a classroom-based setting with current learning material is in line with recent classroom experiments on text learning (e.g., Butler & Roediger, 2007; McDaniel, Agarwal, Huelser, McDermott, & Roediger, 2011; McDaniel, Anderson, Derbish, & Morrisette, 2007; Roediger, Agarwal, McDaniel, & McDermott, 2011). In each of these studies, the to-be-learned materials were embedded in some kind of meaningful context. However, to the best of our knowledge,

none of these studied aimed at investigating whether the retrieval practice effect varies with context.

The second question was whether providing a context would affect the benefits of retrieval practice. Our results suggest that it does as we found a marginally significant interaction between learning condition and context. This interaction effect showed that the positive effect of retrieval practice in the word pairs condition was somewhat larger than in the story condition. However, children in the word pairs condition retrieved more words than the children in the story condition during the first retrieval practice sessions. This in turn might be a plausible explanation for the difference in the retrieval practice effects between the word pairs condition and the context condition. All in all, because the interaction between learning condition and context was just marginally significant, we have to be careful with interpreting the data.

One remaining question is why we did not find any differences between restudying and retrieval practice on the multiple-choice test. Although we did not expect these results, these results are in line with some other studies in which also no benefit of retrieval practice was found on recognition tasks (e.g., Hogan & Kintsch, 1971; Verkoeijen, Tabbers, & Verhage, 2011; Wenger, Thompson, & Bartling, 1980). However, our experiment was different from these studies in the sense that in our experiment, the recognition test was always preceded by the final cued-recall test, which may have confounded the effect of learning condition on recognition performance. Therefore, we think it is better to base our claims about the beneficial effect of retrieval practice on the final cued-recall test given before the multiple-choice test.

Another remaining question is why providing a context of a story did not lead to a memory benefit compared with providing word pairs. Certainly, we did not expect a better performance on both the cued-recall test and the multiple-choice test for the word pairs condition than for the story condition. It should be noted, however, that the results are consistent with other studies in which no benefit of adding contextual information was found in vocabulary learning (e.g., Jones, Levin, Levin, & Beitzel, 2000; McDaniel & Pressley, 1984, 1989). For example, the results of the study of McDaniel and Pressley (1984) with graduate students showed that a context method led to worse recall of definitions than a keyword method and a control method in which a one-word to two-word definition was given. Furthermore, Jones et al. (2000) found a benefit in recall of definitions of a mnemonic keyword strategy compared with a context strategy in sixth grade children. In contrast to the aforementioned studies (Jones et al., 2000; McDaniel & Pressley, 1984, 1989), Rodríguez and Sadoski (2000) found a benefit of adding contextual cues to the keyword mnemonic in ninth grade children who had to learn Spanish-English word pairs. The combination method was superior to the keyword method, which contrasts with our results in which adding a context harmed recall of the words.

There are several possible explanations why we did not find a benefit of context in our study. One explanation is that the context was only provided in the first learning phase but not in the other learning phases and in the test session. Thus, a possible benefit of providing a context may have been diluted over time. However, in the retrieval practice trials that immediately followed the context presentation, words in the word pairs condition were already recalled better than words in the story condition. Thus, we do not think the dilution explanation is very plausible.

Another explanation may be that the contextual information diverted the children from the meaning of the words and disrupted the learning of the new word and its synonym. The explanation of the words throughout reading the story could have harmed the context benefit in the story condition. This fits well with the finding that words in the story condition were already remembered more poorly than in the word pairs condition in the first retrieval practice session. Also, the form of the cued-recall test may have matched better with the word pairs condition than with the story condition. In the word pairs condition, the words were always shown without the story, which is identical to the presentation format in the final test. In contrast, in the story condition, the words were shown within the context of a story.

For future research, it would be interesting to address the benefit of retrieval practice in which context is also used in the other learning phases and in the final test phase, because the use of context is common in primary school vocabulary learning. In this way we could better match the different learning phases and the final test phase with each other.

Acknowledgements

The Board of Public Education Rotterdam (Stichting BOOR) provided financial support for this research. The authors would like to thank the directors and teachers of the schools for their cooperation. Furthermore, we thank the children for participating in this research. Also, we thank Priscillia Bos for collecting the data. Furthermore, we thank Paul Blaney for his help with translating the words and their synonyms into English.

Appendices

Appendix A

Story in Dutch with 20 difficult words for the story condition

Ayoena, een meisje van 10 jaar oud, gaat op bezoek bij haar grootmoeder. Wanneer ze haar huis inloopt, roept ze: 'Het is een **chaos** hier!'. 'Wat is dit?', roept ze dan **kwiek**. Ze komt terug met een **vaal** tasje in haar hand. Het is helemaal versleten en **weerzinwekkend**. Een brede klep met riempjes sluit het tasje af. 'Oma, wat is dit? Het lag bij opa's oude **baret**.' 'Vast niets', zegt oma. Misschien moeten we dat oude tasje maar in de vuilniszak **deponeren**,' zegt oma.

Ayoena wil niet dat het tasje in de vuilniszak gaat, want het is een heel **apart** tasje. Ayoena peutert **heimelijk** de **stugge** riempjes van het tasje los en doet de klep open. 'Er zit iets in!' Oma kijkt **perplex** als ze een **kris** tevoorschijn haalt.

'De kris van opa,' fluistert oma Ietje. Ik dacht dat hij hem al lang weggegeven had. Dit was echt opa's **pronkstuk**'. Ze denkt terug aan opa die **heengegaan** is en begint te **wenen**. Ayoena denkt dat oma verdriet heeft om haar. Ayoena denkt dat ze het **verprutst** heeft, maar dat is niet zo. Oma is eigenlijk ook wel blij met het **signaal** van opa.

Beduusd roept Ayoena: 'Er zit een briefje bij de kris!' Ayoena vouwt het briefje open en leest hardop: 'Ik heb het **vermoeden** dat mijn nieuwsgierige kleindochter nu mijn kris gevonden heeft. Als dat het geval is, is hij voor haar bestemd. Ik hoop dat je deze laatste **gift** aanneemt, Ayoena. Oma zal je **meedelen** hoe je met een kris moet omgaan. Dag. Opa.'

Appendix B

English translation of story in Dutch with 20 difficult words for the story condition

Ayoena, a girl of 10 years old is visiting her grandmother. When she walks into the house, she shouts 'It is total **chaos** here!' Then she shouts **briskly**, 'What is this?' She has a **colourless** bag in her hand. It is totally worn out and **repulsive**. You can close the bag with a broad flap with belts. 'Grandmother, what is this? It was lying next to the old **beret** of grandfather.' 'Surely nothing', grandmother says. Maybe we have to **dispose** the old bag in the trashcan'.

Ayoena does not want to put the bag in the trashcan, because it is a very **unique** bag. Ayoena tampers **secretly** with the **rigid** belts and opens the flap. 'There is something in it!' Grandmother looks **speechless** when she shows **a dagger**.

Grandmother whispers, 'Grandfather's dagger'. I thought he had already given it away. This really was grandfather's **showpiece**'. She thinks back to grandfather who **passed away** and starts to **weep**. Ayoena thinks grandmother is crying because of her. Ayoena thinks she has **failed**, but that is not true. Actually, grandmother is very happy with **the sign** of grandfather.

Ayoena shouts **perplexed**: 'There is a note in the bag with the dagger!' Ayoena opens the note and reads aloud: 'I have the **assumption** that my curious granddaughter has found my dagger. If this is the case, it is for her. I hope you will accept this last **offering**, Ayoena. Grandmother will **describe** to you how to use the dagger. Goodbye. Grandfather.'

Chapter 4

The Benefit of Retrieval Practice over Elaborative Restudy in Primary School Vocabulary Learning

This chapter has been published as:

Goossens, N. A. M. C., Camp, G., Verkoeijen, P. P. J. L., Tabbers, H. K., & Zwaan, R. A. 2014). The benefit of retrieval practice over elaborative restudy in primary school vocabulary learning. *Journal of Applied Research in Memory and Cognition, 3*, 177-182. doi: 10.1016/j.jarmac.2014.05.003.

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.

"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, 2011a; Fritz, Morris, Nolan, & Singleton, 2007; Marsh, Fazio, & Goswick, 2012; Rohrer, Taylor, & Sholar, 2010).

However, to the best of our knowledge, only two studies have investigated retrieval practice in primary school vocabulary 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 (Metcalfe, 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, Camp, Verkoeijen, and Tabbers (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, Camp, Verkoeijen, and Tabbers (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, Camp, Verkoeijen, & Tabbers, 2014; Karpicke & Smith, 2012), we hypothesized that retrieval practice would benefit vocabulary learning in the classroom compared to pure restudy and to elaborative restudy.

Experiment 1

Method

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.

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.

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 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.*).

Table 1 Dutch words and their English translations

Dutch word	English translation
composthoop	compost pile
kringloop	recycling
kunstmest	artificial manure
waterdamp	water vapor
milieuvervuiling	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.

Exercises for the second learning session

For each learning condition two exercises were created (for examples, see Appendix A, B, and 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…*).

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.

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 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.

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 reliabilty. Because of the high agreement, one rater scored the remaining tests.

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, $F < 1$.

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	0.75 (0.22)	0.72 (0.23)	0.82 (0.15)
Multiple-choice	0.92 (0.14)	0.91 (0.16)	0.92 (0.10)

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.

Method

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.

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.

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.

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.

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, η_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.

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	0.69 (0.22)	0.70 (0.22)	0.83 (0.14)
Multiple-choice	0.91 (0.11)	0.92 (0.10)	0.91 (0.16)

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 (Figure 1), and between the retrieval practice condition and the elaborative restudy condition (Figure 2). A positive point estimate reflects an advantage of retrieval practice over either pure restudy or elaborative restudy. The CIs in Figure 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. The same applies to comparisons between retrieval practice and elaborative restudy in Figure 2. All in all, Figures 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.

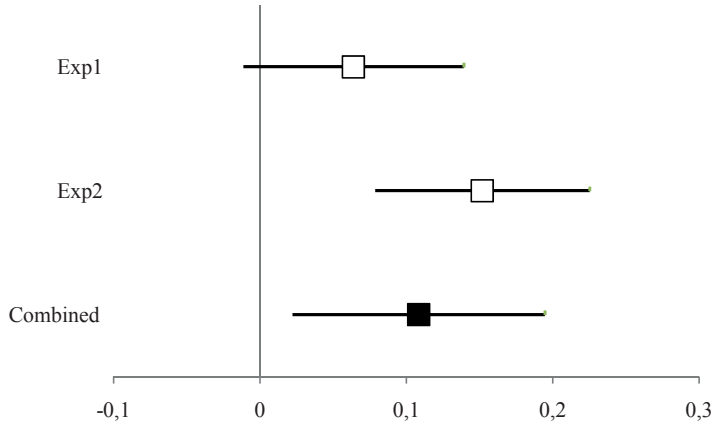


Fig. 1. 95% CIs of the adjusted mean proportion difference between the retrieval practice condition and the pure restudy condition in Experiments 1 and 2 and a combined effect. A positive point estimate indicates a benefit of retrieval practice over pure restudy.

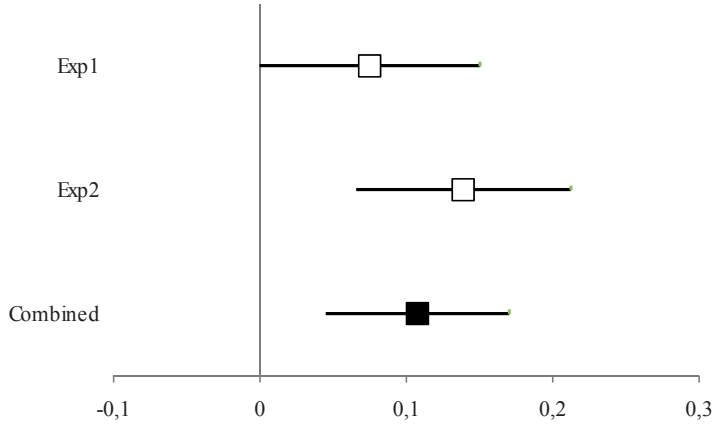


Fig. 2. 95% CIs of the adjusted mean proportion difference between the retrieval practice condition and the elaborative restudy condition in Experiments 1 and 2 and a combined effect. A positive point estimate indicates a benefit of retrieval practice over elaborative restudy.

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, Camp, Verkoeijen, & Tabbers, 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, 2011b; Morris, Bransford, & Franks, 1977; Thomas & McDaniel, 2007).

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.

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 Bonnevits 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.

Appendices

Appendix A: Pure Restudy Exercises

Exercise 1

Please read the following sentences. Write down the correct missing words on the black lines. By doing so, the first and second sentence will become the same.

A pile in the garden with vegetable, fruit and garden waste, is called **a compost pile**.

_____ the garden with vegetable, fruit and garden waste, is called _____.

Exercise 2

Please read the sentences carefully. The sentences are scrambled. Search for the right sentence and write down the right words on the black line.

A pile in the garden with vegetable, fruit and garden waste, is called **a compost pile**.
A factory in which we generate electricity, is called **a power station**.
You can use old paper to make new paper, what will become old paper again. This is what we call **a recycling**.

You can use used paper to make new paper, that will become used paper again. _____
_____.

A pile in the garden with vegetable, fruit and garden waste, _____
_____.

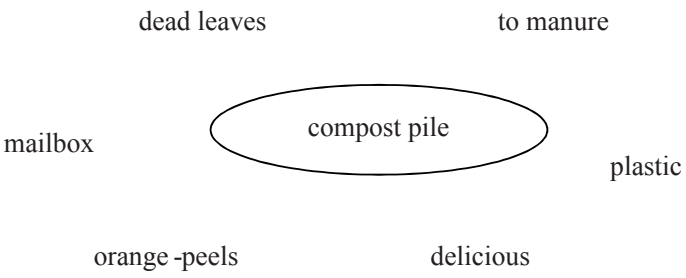
A factory in which we generate _____.

Appendix B: Elaborative Restudy Exercises

Exercise 1

Please pick three words from the web that are related to the word in the middle. Draw a line between these words and the word in the middle.

A pile in the garden with vegetable, fruit and garden waste, is called a **compost pile**.



Exercise 2

Please read the sentences carefully. Afterwards please do the exercises.
A powder that we can use to blow up rocks is called **dynamite**.
A very big stone is called **a rock**.
A pile in the garden with vegetable, fruit and garden waste, is called **a compost pile**.
Chunks of stone that contain iron, are called **iron ore**.

Choose from these words and write down the word in the right word list:
dynamite, a rock, a compost pile, iron ore

metal – iron mine – raw material – _____

dead leaves – orange-peels – ground – _____

explosive – dangerous – bomb – _____

hard – nature – beach – _____

Appendix C: Retrieval Practice Exercises

Exercise 1

Please read the sentences carefully. Write down one of the words you previously learned. To help you the first three letters of the word are already given.

A pile in the garden with vegetable, fruit and garden waste, is called a com_____.

Exercise 2

Please read the sentences carefully. Write down one of the words you previously learned. To help you the first letter of the word is already given.

A pile in the garden with vegetable, fruit and garden waste, is called a c_____.

Appendix D: Final Test Items

Fill-in-the-blank test

Write down one of the words you previously learned.

A pile in the garden with vegetable, fruit and garden waste, is called a _____.

Multiple-choice test

In the sentences there is one word missing. Please complete the sentence with the right word.

Choose from: *dynamite – flower bulbs – shell – the warming – iron ore – mailbox – a flare – recycling – rock – compost pile*

1. To extract iron ore from the ground, Jack has to blow up a rock using _____.
2. Jack is working in an iron mine, he takes _____ from the ground.
3. John is collecting used paper. Later he will turn it into new paper. The paper is part of a _____.
4. Near to the water John climbed a big _____, to have a nice view over the sea.
5. John throws his fruit waste on the _____. He can use this later to manure his garden.

Chapter 5

Distributed Practice and Retrieval Practice in Primary School Vocabulary Learning: A Multi-Classroom Study

This chapter has been submitted for publication as:
Goossens, N. A. M. C., Camp, G., Verkoeijen, P. P. J. L., Tabbers, H. K., Bouwmeester, S., & Zwaan, R. A. (submitted). Distributed practice and retrieval practice in primary school vocabulary learning: A multi-classroom study.

Abstract

We examined the effects of distributed practice and retrieval practice in seven- to thirteen-year-old children. The practice was integrated in real-life primary school vocabulary lessons during four weeks. For the distributed practice manipulation, children performed the exercises in one week (short-lag condition) or in two weeks (long-lag condition). For the retrieval practice manipulation, children copied a part of the description of the word (restudy condition) or recalled the description (retrieval practice condition). At the end of each week the children produced the descriptions of the words in a cued-recall test and after several weeks they had to recognize the correct descriptions of the words in a multiple-choice test. On the cued-recall test benefits of short-lag repetition and restudy were found, and on the multiple-choice test no differences between conditions were found. These results question the practical value of long-lag repetition and retrieval practice in real-life primary school vocabulary lessons.

Vocabulary development is important for enhancing reading comprehension (e.g., Anderson & Freebody, 1981) and also for better world knowledge in general (Stahl & Nagy, 2006). The more vocabulary knowledge one has, the more precisely and complexly one can speak and think about the world (Stahl & Nagy, 2006). For instance, a person who knows the terms *searing*, *stewing,* and *poaching* will think about cooking in a different way (and will maybe even *cook* in a different way), than a person who is limited to *baking*, *boiling,* and *roasting* (after examples of Stahl & Nagy, 2006, p. 5). Although teachers are aware of the importance of vocabulary knowledge in primary schools, some studies have shown that only a small amount of the school time is devoted to vocabulary learning (e.g., Scott, Jamieson-Noel, & Asselin, 2003). Furthermore, the time spent on vocabulary learning is not always used effectively. For example, teachers may instruct their pupils to only copy word definitions from dictionaries, even though this has been shown to be ineffective (Scott et al., 2003).

Vocabulary teaching and learning should be guided by well-supported diverse principles, such as pre-teaching the words, using dictionaries and glossaries during reading texts, and using contexts to guess the words in a text (Nation, 2004). In addition to these principles, a number of learning techniques stemming from cognitive and educational psychology have consistently shown beneficial effects on memorizing word pairs in the psychological laboratory (for an overview, see Dunlosky, Rawson, Marsh, Nathan, & Willingham, 2013). However, there is a paucity of research into these techniques in primary school vocabulary learning. The aim of the current study is therefore to investigate two robust memory strategies that emerged from the psychological laboratory in the context of primary school vocabulary learning (in Grade 2, 3, 4, and 6). These memory strategies are distributed practice (for reviews, see e.g., Cepeda, Pashler, Vul, Wixted, & Rohrer, 2006; Delaney, Verkoeijen, & Spirgel, 2010) and retrieval practice (for reviews, see e.g., Rawson & Dunlosky, 2011; Roediger & Butler, 2011; Roediger, Putnam & Smith, 2011; Roediger & Karpicke, 2006).

The first memory strategy we will investigate is distributed practice. Many studies have demonstrated that distributed practice (i.e., spacing learning over time) leads to better retention than massed practice (i.e. massing learning without any intervening items). This so-called spacing effect has been demonstrated in more than 300 published experiments (for reviews, see e.g., Cepeda et al., 2006, Delaney et al., 2010). Most of these studies were conducted in a laboratory setting with adults learning word-pairs. However, the beneficial effect of distributed practice has also been found in primary school children learning pictures (e.g., Toppino & DiGeorge, 1984; Toppino, Kasserman & Mracek, 1991), words (e.g., Toppino & DeMesquita, 1984), or a combination of pictures and words (e.g., Cahill & Toppino, 1993; Rea & Modigliani, 1987; Toppino, 1993). Furthermore, some studies have also demonstrated that distributed practice can enhance (foreign) vocabulary learning

in undergraduate students (e.g., Bahrick, Bahrick, Bahrick, & Bahrick, 1993; Bloom & Shuell, 1981). In addition, a few studies have demonstrated a benefit of distributed practice in vocabulary learning (e.g., Goossens, Camp, Verkoeijen, Tabbers, & Zwaan, 2012; Kornell, 2009; Sobel, Cepeda, & Kapler, 2011). For example in the study of Kornell (2009), undergraduate students had to study 20 flashcards each with an unknown word and its synonym (e.g., *effulgent – brilliant*) by distributed learning, namely twice in four consecutive sessions, and 20 flashcards by massed learning, namely eight times in one single session. On the final test a beneficial effect of distributed practice was found. In the study of Goossens et al. (2012), children in Grade 3 had to learn 30 words by performing three different exercises from the regular learning materials of Grade 4. These exercises were presented on the same day (massed condition) or on three consecutive days (distributed condition). On the final cued-recall tests after one week and after five weeks, the children performed better on the words learned by distributed practice than on the words learned by massed practice. Thus, distributed practice has shown to be effective for learning vocabulary.

The second memory strategy we will investigate in primary school vocabulary learning is retrieval practice, which is also commonly referred to as testing. The testing effect is the phenomenon that retrieval practice leads to better long term retention than additional study (for reviews see e.g., Rawson & Dunlosky, 2011; Roediger & Butler, 2011; Roediger, Putnam & Smith, 2011; Roediger & Karpicke, 2006). As in the studies conducted on distributed practice, most experiments on retrieval practice have been conducted using word lists or word pairs as learning materials (e.g., Bouwmeester & Verkoeijen, 2011a; Carpenter, Pashler, & Vul, 2006; Carpenter, Pashler, Wixted, & Vul, 2008; Toppino & Cohen, 2009; Tulving, 1967; Wheeler, Ewers, & Buonanno, 2003), and foreign vocabulary pairs (e.g., Carpenter et al., 2008; Carrier & Pashler, 1992; Fritz, Morris, Acton, Voelkel, & Etkind, 2007; Karpicke, 2009; Karpicke & Roediger, 2008; Pashler, Cepeda, Wixted, & Rohrer, 2005; Pyc & Rawson, 2007, 2009, 2011; Toppino & Cohen, 2009). In addition, the beneficial effect of retrieval practice was found in studies where people had to learn uncommon, infrequent, or unknown words from their own language in both adults (e.g., Cull, 2000; Karpicke & Smith, 2012) and primary school children (e.g., Goossens, Camp, Verkoeijen, & Tabbers, 2014; Goossens, Camp, Verkoeijen, Tabbers, & Zwaan, 2014; Metcalfe, Kornell, & Son, 2007). For example, in the study of Karpicke and Smith (2012), undergraduate students had to learn uncommon English words and their one-word definition by either restudy or retrieval practice. In the restudy condition, participants studied each word pair on a computer screen for seven seconds (e.g., *antiar – poison*). In the retrieval practice condition, they had seven seconds to type in the definition (e.g., *antiar - …*). On the final cued-recall test after one week, the retrieval practice condition outperformed the restudy condition. In the study of Goossens, Camp, Verkoeijen, and

Tabbers (2014), third graders practiced the meaning of ten words by reading aloud the words, thus by restudy (e.g., *to weep – to cry*), and ten other words by retrieval practice (e.g., *to weep* – …). On the final cued-recall test after one week a benefit of retrieval practice was found.

Thus, it has been shown that both distributed practice and retrieval practice can facilitate word learning in adults and in primary school children. In the current study, the research question was whether the use of distributed practice and retrieval practice is beneficial when integrated in primary school vocabulary lessons conducted in the classroom. The goal was to provide an ecologically valid test of the possible benefits of distributed practice and retrieval practice to determine the relevance for application of these learning strategies in the classroom. To ensure a high level of ecological validity, we took several measures. First, in many experiments on the distributed practice effect, a spaced condition is compared to a massed condition in which words are learned during one single learning session (see for example, Goossens et al., 2012; Sobel et al., 2011). However, in real-life vocabulary learning curricula, vocabulary exercises are not massed in one single learning session, but distributed over several learning sessions during the week (e.g., Janssen & Van Ooijen, 2012; Van de Gein, Van de Guchte, & Kouwenberg, 2008). Thus, it is more relevant for educational practice to investigate the benefit of distributed practice using different spacing intervals (i.e., to investigate the *lag* effect). In the present study, we compared a short-lag condition in which practice is already somewhat distributed (across one week) to a long-lag condition in which practice is distributed across two weeks. Second, in the current study we attempted to make repeated practice more similar to regular educational practices. In earlier experiments, participants simply repeated word pairs (see for example, Goossens, Camp, Verkoeijen, & Tabbers, 2014). Actually, in real-life vocabulary learning, words are not only restudied by simple repetition, but also practiced with different types of exercises (e.g., Janssen & Van Ooijen, 2012; Van de Gein et al., 2008). Therefore, it is relevant to investigate the benefit of distributed practice and retrieval practice in a situation in which the words are learned by doing different exercises from the regular learning material. Third, the current study will be conducted within real-life classroom vocabulary lessons using words and exercises that are part of the curriculum. Fourth, in contrast to studies done with only one age group (e.g., Goossens et al., 2012; Goossens, Camp, Verkoeijen, & Tabbers, 2014; Goossens, Camp, Verkoeijen, Tabbers, & Zwaan, 2014; Sobel et al., 2011), we conducted experiments in Grade 2, 3, 4, and 6, to be able to generalize our results to different age groups. Fifth, we used multiple-choice tests to measure long-term effects, because in educational practice, multiple-choice tests are most often used to establish vocabulary size (e.g., Van Berkel & Alberts, 2009; Van Berkel & Hilte, 2009; Verhoeven & Vermeer, 1993). Finally, the current study combines distributed practice and retrieval practice. As far as we know only a few

experiments have been done in which these strategies were investigated together (e.g., Cull, 2000). Importantly, distributed practice and retrieval practice have not yet been investigated simultaneously in primary school vocabulary learning. So, our main interest was to investigate whether the benefits of distributed practice and retrieval practice found in previous studies (e.g., Goossens, Camp, Verkoeijen, & Tabbers, 2014; Sobel et al., 2011) will generalize to actual classroom practice.

In short, in the current experiment we investigated the effects of distributed practice and retrieval practice in primary school children in Grade 2, 3, 4, and 6. This experiment was integrated in the regular vocabulary-learning curriculum. According to the literature, distributed practice and retrieval practice are strong and robust memory strategies that should generalize to an ecologically valid learning situation (e.g., Dunlosky et al., 2013). We expected to find main effects of distributed practice and retrieval practice. Thus, we expected that long-lags would benefit primary school vocabulary learning more than short-lags, and that retrieval practice would benefit primary school vocabulary learning more than restudy.

Method

Participants

Children of the Dutch Grade 4, 5, 6, and 8, this is equivalent to Grade 2, 3, 4, and 6 in the U.S., were all recruited from the same school in The Netherlands. A total number of 237 children participated, that is, 60 children from Grade 2, 55 children from Grade 3, 53 children from Grade 4 and 69 children from Grade 6. For 36 children, parents withheld permission to use the children's data. Another 69 children were not able to participate during all sessions, and the data of 3 other children could not be used due to procedural errors. This resulted in a sample of 129 participants (64 boys, 65 girls), with a mean age of 9.90 years (*SD* = 1.67). The group consisted of 33 participants in Grade 2 (15 boys) with a mean age of 7.93 years (range 7.42 to 8.50, *SD* = 0.29), 32 participants in Grade 3 (15 boys) with a mean age of 9.26 years (range 8.25 to 11.17, *SD* = 0.61), 31 participants in Grade 4 (15 boys) with a mean age of 10.14 years (range 9.42 to 11.42, *SD* = 0.51), and 33 participants in Grade 6 (19 boys) with a mean age of 12.27 years (range 11.67 to 13.42, *SD* = 0.48). Only 125 of these children took the multiple-choice test, namely 31 children of Grade 2, 32 children of Grade 3, 30 children of Grade 4, and 32 children of Grade 6. The children knew they participated in an experiment and their parents had given informed consent for using the children's data. In the sample there were no children with learning problems.

Design

In this research we used a 4x2x2 mixed design in which grade was a between-subjects factor and in which both distribution of learning (short-lag versus long-lag) and repetition of learning (restudy versus retrieval practice) was manipulated within-subjects. Distribution of learning was manipulated by distributing the exercises of the same words over one week (short-lag condition), or over two weeks (long-lag condition). Repetition of learning was manipulated by providing two additional exercises for each word that required either copying a part of the word description (restudy condition), or recalling the word description (retrieval practice condition). This resulted in four learning conditions: (1) short-lag restudy, (2) short-lag retrieval practice, (3) long-lag restudy, and (4) long-lag retrieval practice. There were two dependent variables. The first dependent variable was the performance on a cued-recall test given one, two, or three days after the final learning session, in which the children were given the vocabulary words and for each word they had to write down the correct description. The second dependent variable was the performance on a multiple-choice test in which the children had to choose for each vocabulary word the correct description out of four options. This test was intended to measure long-term retention and was given one week (for Grade 6), two weeks (for Grade 3) or eleven weeks (for Grade 2 and 4) after the last learning session.

Materials

For each grade, forty words and their exercises were selected from the regular vocabulary learning materials that were currently used in that grade within the school. These 40 words were presented in thematic lists of five words and were introduced in blocks of twenty words. Each word was introduced in the vocabulary learning book by means of an illustration, a definition and a context sentence. For each word we selected four different exercises from the textbook and the workbook. For a small number of words we found less than four exercises in the vocabulary learning material, so in those cases we constructed similar exercises ourselves.

Table 1, 2, 3 and 4 show the Dutch words, their English translations and their characteristics (their word type, amount of letters, and word frequency) used in Grade 2, 3, 4 and 6 respectively. Table 5 shows the characteristics of the four vocabulary lists of 40 words. The median word frequency is based on the Dutch Measure of Lexical Richness for primary school materials (Schrooten & Vermeer, 1994).

Table 1 All Dutch words used in Grade 2 with English translations and their characteristics

Dutch word	English translation	Word type	Letters	Frequency
oversteken	to cross over	verb	10	89
fietspad	cycle track	noun	8	18
stoep	sidewalk	noun	5	159
zebrapad	pedestrian crossing	noun	8	6
papierbak	waste-paper basket	noun	9	2
in de buurt	nearby	adverb	9	196
veilig	safe	adjective	6	136
omweg	detour	noun	5	8
plantsoen	public garden	noun	9	13
station	station	noun	7	66
de weg vragen	to ask for directions	verb	11	-
verdwalen	to get lost	verb	9	65
brug	bridge	noun	4	89
kruispunt	intersection	noun	9	12
bushalte	bus stop	noun	8	6
in de buurt blijven	to stay close	verb	16	196
openbaar vervoer	public transport	noun	15	-
verkeersdrempel	speed bump	noun	15	-
wegwijzer	sign post	noun	9	19
oprit	driveway	noun	5	2
grond	ground	noun	5	21
ijskoud	ice-cold	adjective	7	23
luchtje	odor	noun	7	23
vleugel	wing	noun	7	122
meer	lake	noun	4	152
fris	fresh	adjective	4	14
kenmerk	characteristic	noun	7	2
natuur	nature	noun	6	172
omgeving	environment	noun	8	48
een onderzoek doen	to investigate	verb	16	47
rommel	mess	noun	6	89
spons	sponge	noun	5	23
vuilniszak	garbage bag	noun	10	40
vuilnisman	garbage collector	noun	10	9
vuilniswagen	garbage truck	noun	12	19
afval	garbage	noun	5	78
een kijkje nemen	to take a look	verb	14	16
proefje	experiment	noun	7	38
temperatuur	temperature	noun	11	54
zo goed mogelijk	as good as possible	adverb	14	-

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

Table 2 All Dutch words used in Grade 3 with English translations and their characteristics

Dutch word	English translation	Word type	Letters	Frequency
fototoestel	camera	noun	11	18
fotograaf	photographer	noun	9	28
muzikant	musician	noun	8	33
kunstschilder	painter	noun	13	11
gitaar	guitar	noun	6	34
acrobaat	acrobat	noun	8	16
publiek	audience	noun	7	86
dwarsfluit	German flute	noun	10	17
applaus	applause	noun	7	20
model	model	noun	5	1
tentoonstelling	exhibition	noun	15	28
viool	violin	noun	5	78
piano	piano	noun	5	56
zanger	singer	noun	6	17
danseres	female dancer	noun	8	13
bibliotheek	library	noun	11	31
reiziger	traveler	noun	8	21
taxi	taxi	noun	4	18
optreden	performance	noun	8	1
taalprobleem	language problem	noun	12	-
kwaken	to croak	verb	6	44
in de hoek zetten	to be put in the corner of the classroom for punishment	verb	14	-
paling	eel	noun	6	12
met mes en vork eten	to eat with a knife and fork	verb	16	-
verschrikkelijk	terrible	adjective	15	85
grazen	to graze	verb	6	27
dobberen	to float	verb	8	17
rakker	rascal	noun	6	3
ruif	rack	noun	4	2
de trom slaan	to beat the drum	verb	11	-
Japanner	Japanese	noun	8	3
Chinees	Chinese	noun	7	30
Zweed	Swede	noun	5	-
Marokkaan	Moroccan	noun	9	15
Surinamer	Surinamer	noun	9	11
cello	cello	noun	5	1
contrabas	double-bass	noun	9	-
saxofoon	saxophone	noun	8	3
keyboard	keyboard	noun	8	9
harp	harp	noun	4	-

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

Table 3 All Dutch words used in Grade 4 with English translations and their characteristics

Dutch word	English translation	Word type	Letters	Frequency
slopen	to demolish	verb	6	7
tractor	tractor	noun	7	17
kiepauto	dump truck	noun	8	1
hekwerk	fencing	noun	7	-
bouwmaterialen	construction materials	noun	14	-
riolering	sewerage	noun	9	1
moker	sledgehammer	noun	5	-
het puin afvoeren	to transport rubble	verb	15	5
sloper	demolisher	noun	6	4
wals	roller	noun	4	4
dertien hoog	on the 14th floor	adverb	11	-
twee onder één kap	semi-detached house	noun	15	-
woonboot	houseboat	noun	8	6
beton	concrete	noun	5	20
balkon	balcony	noun	6	41
landhuis	country cottage	noun	8	2
villa	villa	noun	5	23
oprijlaan	entranceway	noun	9	6
kraakpand	squat	noun	9	-
bungalow	bungalow	noun	8	9
herstellen	to mend	verb	10	6
helaas	unfortunately	adverb	6	46
verstoren	to disturb	verb	9	4
verspillen	to waste	verb	10	6
afhankelijk	dependent	adjective	11	11
elektriciteitscentrale	power station	noun	22	6
grondstof	raw material	noun	9	29
aantasten	to affect	verb	9	10
recycling	recycling	noun	9	-
aardgas	natural gas	noun	7	14
stiltegebied	quiet area	noun	12	10
natuurreservaat	nature reserve	noun	15	14
wild	wildlife	noun	4	21
hei	heather	noun	3	14
moeras	swamp	noun	6	68
natuurbehoud	nature conservation	noun	12	1
fazant	pheasant	noun	6	7
natuurlandschap	nature landscape	noun	15	17
zandverstuiving	sand drifts	noun	15	5
poel	pool	noun	4	4

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

Table 4 All Dutch words used in Grade 6 with English translations and their characteristics

Dutch word	English translation	Word type	Letters	Frequency
redactie	editorial staff	noun	8	5
permanent	permanent	adjective	9	-
herinnering	memory	noun	11	24
opstellen	to draw up	verb	9	16
anoniem	anonymous	adjective	7	-
coördineren	to coordinate	verb	11	-
inspiratie	inspiration	noun	10	-
kopij	copy	noun	5	4
rode draad	theme	noun	9	-
aandenken	keepsake	noun	9	1
tenzij	unless	adverb	6	7
toekomen aan iets	to get round to something	verb	15	1
principe	principle	noun	8	1
vermenigvuldigen	to multiply	verb	16	1
tijdelijk	temporary	adjective	9	6
in je achterhoofd houden	to keep in mind	verb	21	-
weemoed	melancholy	noun	7	2
deadline	deadline	noun	8	-
overzichtelijk	well-ordered	adjective	14	4
uitgave	publication	noun	7	3
faalangst	fear of failure	noun	9	-
planning	schedule	noun	8	-
voortzetten	to continue	verb	11	8
administratie	administration	noun	13	1
kaften	to cover a book	verb	6	1
spieken	to peek	verb	7	4
pressen	to press	verb	7	-
tussenuur	odd hour	noun	9	-
voortijdig	premature	adjective	10	-
raadplegen	to consult	verb	10	2
stranden	to get bogged down	verb	8	4
havo	higher general secondary education	noun	4	20
vmbo	prepatory secondary vocational education	noun	4	-
vwo	pre-university education	noun	3	5
mentor	tutor	noun	6	-
afhaken	to drop out	verb	7	-
falen	to fail	verb	5	1
registreren	to register	verb	11	-
vakkenpakket	curriculum	noun	12	-
benjamin	Benjamin	noun	8	-

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

Table 5 Characteristics of the vocabulary lists used in Grade 2, 3, 4 and 6: The number of nouns, verbs, adjectives and adverbs, the average number of letters in the words (SD in parentheses), the number of words that were in the Measure of Lexical Richness list (out of 40), and the median frequency of the words.

Grade	Noun	Verb	Adj.	Adv.	Letters	MLR Words	Frequency*
2	29	6	3	2	8.55 (3.43)	36	35.5 (range 2 to 196)
3	33	6	1	0	8.25 (3.17)	33	17 (range 1 to 86)
4	31	6	1	2	8.98 (4.02)	33	7 (range 1 to 68)
6	20	14	5	1	8.93 (3.48)	22	4 (range 1 to 24)

* Frequency based on the Dutch Measure of Lexical Richness for primary school materials (Schrooten & Vermeer, 1994)

The words were first described in one or two sentences (e.g., *A pedestrian crossing – A place to cross the street safely. There are white stripes on the road.*). There were different types of textbook exercises, like for example questions in which the word had to be connected with the right picture, fill-in-the-blank questions (e.g., *We cross the street on …*), or questions in which the words had to be related to (a part of) their description (e.g., *White stripes on the road*).

We also constructed additional exercises in which repetition of learning was manipulated. In the restudy exercises children had to copy a part of the description (e.g., *A pedestrian crossing is a place to cross the street safely. There are white stripes on the road. - A pedestrian crossing is a place to cross the street safely. There are white …*) In the retrieval practice exercises the children had to retrieve the description and to write it down (e.g., *A pedestrian crossing – …*).

The cued-recall test consisted of questions in which the children had to give the description of the word (e.g., *A pedestrian crossing - …*) and the multiple-choice test consisted of questions in which the children had to choose the correct description of the word, that was somewhat different than the description from the textbook (e.g., *Which words are most related to the meaning of* ***a pedestrian crossing?*** *A.car – bike – pedestrian; B. road – to walk – zebra; C. white stripes – to walk – pedestrian; D. zebra – white stripes – to cycle)*.

For each grade, the 40 stimulus words were split into four lists of ten words each, in which five easy and five difficult words were included (the qualification of the words in 'easy' and 'difficult', was provided by the authors of the textbook). These four lists were assigned to the four learning conditions (short-lag restudy, short-lag retrieval practice, long-lag restudy and long-lag retrieval practice). For counterbalancing purposes we rotated the four lists over the conditions, with a constant order of conditions (short-lag

restudy, short-lag retrieval practice, long-lag restudy, and long-lag retrieval practice). For practical purposes, we used only four different counterbalancing conditions. At the final tests we used another list order than the order in which the words were presented in the textbook. At the multiple-choice test in which all 40 words were tested, the lists were mixed, but we used six different versions with different questions in order to prevent copying.

Procedure

The experiment took place within the classroom and was the same for every grade. The learning phase lasted four weeks and was divided into two blocks of two weeks that were procedurally identical. In every week of the learning phase, there was a learning session on Monday, another learning session on Tuesday, Wednesday or Thursday and a test session on Friday[1]. For an overview of the complete procedure, see Table 6.

Procedure in the first week

For the procedure in the first week see the first row in Table 6. In the first learning session of the first week, the experimenter (the first author) showed an illustration from the textbook in which twenty words were presented (five from each condition). Next to the illustration, four lists of five words and their definitions were given. At first, the children looked at the illustration and talked with the experimenter about what was shown in the pictures. Then, the experimenter explained each word that was illustrated, by giving a definition, a context sentence of the word and some additional information about the meaning of the word. After all words were explained to the children they had to do two sets of two exercises. In the first set the children had to do exercises for ten words (Exercise 1a for five words from the short-lag condition and Exercise 1a for five words from the long-lag condition). After finishing the first set, the children checked their performance with an answer sheet, and corrected wrong answers. Then, the children continued with the second set of exercises for five words from the first set (Exercise 2a for the short-lag condition), and for five new words (Exercise 1b for the long-lag condition). Again, the children checked their performance with an answer sheet.

1 For Grade 2 and Grade 4 and for some children of Grade 6 in one week the first learning session was on Tuesday instead of on Monday, and for Grade 2 and 4 in one week the test session was on Thursday instead of on Friday.

Table 6 Procedure of the experiments. Every row represents a week. Every column represents a learning session in which short-lag and long-lag items were learned.

Week No.	Session 1 Short-lag items	Session 1 Long-lag items	Session 2 Short-lag items	Session 2 Long-lag items	Session 3
1	*Introduction*	*Introduction*	*Repetition Exercises*	*Repetition Exercises*	*Test*
	1-10	11-20	Restudy (A): 1-2-3	Restudy (A): 11–12	1-5
			Retrieval Practice (B): 4-5	Retrieval Practice (B): 13-14-15	
			Restudy (C): 1-2-3	Restudy (C): 16-17-18	
			Retrieval Practice (D): 4-5	Retrieval Practice (D): 19-20	
	Textbook Exercises	*Textbook Exercises*	*Textbook Exercises*	*Textbook Exercises*	
	1a: 1-5	1a: 11-15	3a: 1-5	2a: 11-15	
	2a: 1-5	1b: 16-20	4a: 1-5	2b: 16-20	
2			*Repetition Exercises*	*Repetition Exercises*	*Test*
			Restudy (A): 6–7	Restudy (A): 11–12	6-20
			Retrieval Practice (B): 8–9–10	Retrieval Practice (B): 13–14–15	
			Restudy (C): 6–7	Restudy (C):16–17-18	
			Retrieval Practice (D): 8–9–10	Retrieval Practice (D): 19–20	
	Textbook Exercises	*Textbook Exercises*	*Textbook Exercises*	*Textbook Exercises*	
	1b: 6–10	3a: 11–15	3b: 6–10	4a: 11–15	
	2b: 6–10	3b: 16–20	4b: 6–10	4b: 16–20	
3	*Introduction*	*Introduction*	*Repetition Exercises*	*Repetition Exercises*	*Test*
	21-30	31-40	Restudy (A): 21–22	Restudy (A): 31–32–33	21-25
			Retrieval Practice (B): 23–24–25	Retrieval Practice (B): 34–35	
			Restudy (C): 21–22	Restudy (C): 36–37	
			Retrieval Practice (D): 23–24–25	Retrieval Practice (D): 38–39–40	
	Textbook Exercises	*Textbook Exercises*	*Textbook Exercises*	*Textbook Exercises*	
	1c: 21–25	1c: 31–35	3c: 21–25	2c: 31–35	
	2c: 21–25	1d: 36–40	4c: 21–25	2d: 36–40	
4			*Repetition Exercises*	*Repetition Exercises*	*Test*
			Restudy (A): 26–27–28	Restudy (A): 31–32–33	26 – 40
			Retrieval Practice (B): 29–30	Retrieval Practice (B): 34–35	
			Restudy (C): 26–27-28	Restudy (C): 36–37	
			Retrieval Practice (D): 29–30	Retrieval Practice (D): 38–39–40	
	Textbook Exercises	*Textbook Exercises*	*Textbook Exercises*	*Textbook Exercises*	
	1d: 26–30	3c: 31–35	3d: 26–30	4c: 31–35	
	2d: 26–30	3d: 36–40	4d: 26–30	4d: 36–40	
5 – 15					*MC Test*
					1–40

In the second learning session, the children started with four repetition exercises (A, B, C, and D) on the words already practiced in the first learning session. In Exercise A, five words were practiced by restudy (two or three words from the short-lag condition and two or three words from the long-lag condition). In Exercise B, five words were practiced by retrieval practice (two or three short-lag words and two or three long-lag words). Exercise C and D were similar to Exercise A and B, except that the same five short-lag words and the other five long-lag words were practiced.

After these repetition exercises, the children again performed two sets of regular textbook exercises. As in the first learning session, the children practiced five words with two different exercises (Exercise 3a and Exercise 4a for the short-lag condition) and ten words with one exercise (Exercise 2a and Exercise 2b for the long-lag condition). Again, the children used an answer sheet to check their performance.

In the test session on Friday, the children were tested on the five words of the short-lag condition that had been practiced during the week. After the children finished the test, they could continue with their own school work.

Procedure after the first week

For the procedure in the second, third, and fourth week, see respectively the second, the third, and the fourth row in Table 6. The procedure of the second week was the same as the first week, with a few exceptions. First, there was no introduction during the first learning session; the children immediately started with the exercises. Second, the children practiced with five new short-lag words and with the ten long-lag words already practiced during the first week. Third, in the test session the children were tested on both the five short-lag words, as well as on the ten long-lag words. The procedure of the third and fourth week was the same as the procedure in the first and second week. In these weeks twenty other words were practiced.

One week (Grade 6), two weeks (Grade 3) or eleven weeks (Grade 2 and 4) after the last test session of the learning phase, the children were tested again on all 40 words, with the multiple-choice test. For the children of Grade 3 and Grade 6 the multiple-choice test session was relatively short after the learning phase, due to the fact that the learning phase for these grades was at the end of the school year, whereas for the children of Grade 2 and 4 the learning phase was earlier in the school year.

Data Analysis

Both the cued-recall test and the multiple-choice test consisted of 40 questions. The 40 items were equally distributed across the four conditions, thus 10 items for each condition. For the final cued-recall questions the children could receive a maximum of two points per question. An answer was wrong (0 points), sufficient (1 point), or completely right (2

points). For the final multiple-choice questions the children could receive maximally one point per question. Hence within each condition, for the cued-recall test, the score range was 0-20 points and for the multiple-choice test, this was 0-10 points.

An independent rater scored all cued-recall tests, and another independent rater scored 20% of the cued-recall tests, by using the same scoring form, to check the interrater reliability. Intraclass Correlation (*r*) between the two raters was .86, indicating a high interrater reliability.

Results

Learning Phase Data Retrieval Practice Conditions

Before we report the main analyses, we will first explore the data from the retrieval practice conditions on the retrieval practice exercises B and D during the learning phase. In exploring this data we had to consider that because of the design of the experiment, there were differences for the short-lag and the long-lag condition in the number of textbook exercises the children had done before they did the two retrieval practice exercises. For the short-lag words, the children completed two textbook exercises before they received the first and second retrieval practice exercise. For the long-lag words, the children completed only one textbook exercise before they received the first retrieval practice exercise, and they completed in total three textbook exercises before they received the second retrieval practice exercise. The mean scores on these retrieval practice exercises are presented in Table 7.

Table 7 Proportion correct on the words in the first and second retrieval practice exercise for the four grades and the children of the four grades together (*SD* in parentheses).

Grade	Short-lag first exercise	Short-lag second exercise	Long-lag first exercise	Long-lag second exercise
2	0.52 (0.18)	0.48 (0.18)	0.48 (0.19)	0.52 (0.20)
3	0.65 (0.19)	0.64 (0.18)	0.63 (0.16)	0.65 (0.13)
4	0.62 (0.19)	0.63 (0.18)	0.59 (0.16)	0.61 (0.18)
6	0.52 (0.15)	0.54 (0.15)	0.52 (0.14)	0.57 (0.14)
All	0.57 (0.19)	0.57 (0.18)	0.55 (0.17)	0.58 (0.17)

The children had an overall performance on the retrieval exercises of 0.57 (*SD* = 0.16). For Grade 2 the overall performance was 0.50 (*SD* = 0.17), for Grade 3 it was 0.64 (*SD* = 0.15), for Grade 4 it was 0.61 (*SD* = 0.15), and for Grade 6 it was 0.53 (*SD* = 0.13). Furthermore, the mean proportion correct on the short-lag retrieval exercises was 0.57 (*SD* = 0.18), and on the long-lag retrieval exercises it was 0.57 (*SD* = 0.16). The mean proportions correct in the different retrieval phases are as follows: in the first phase 0.56 (*SD* = 0.17), and in the second phase 0.58 (*SD* = 0.16).

Descriptive Statistics of the Final Test Data

We calculated the unadjusted mean proportion correct scores and the standard deviations for the four conditions (i.e., short-lag restudy; short-lag retrieval practice; long-lag restudy; long-lag retrieval practice) on the cued-recall test (see Table 8) and on the multiple-choice test (see Table 9).

Table 8 shows that the overall mean proportion score on the cued-recall test is 0.59 correct (*SD* = 0.15). Furthermore, the overall mean proportion correct for the short-lag condition was 0.61 (*SD* = 0.15), and for the long-lag condition it was 0.57 (*SD* = 0.16). The overall mean proportion score for the restudy condition was 0.61 (*SD* = 0.16), and for the retrieval practice condition it was 0.57 (*SD* = 0.15). Overall, the average proportion correct on the cued-recall test is quite similar in the different grades, but the standard deviation in Grade 2 is larger than in the other grades.

Table 9 shows that the overall mean proportion correct on the multiple-choice test is 0.73 correct (*SD* = 0.14). Furthermore, the overall mean proportion score for the short-lag condition is 0.73 (*SD* = 0.16), and for the long-lag condition it is 0.74 (*SD* = 0.16). Further, the overall mean score is 0.74 (*SD* = 0.15), and for the retrieval practice condition it is 0.73 (*SD* = 0.16).

Overall, the average proportion score on the multiple-choice test is quite similar in the different grades. However, the standard deviations in Grade 2 and 4 are larger than in Grade 3 and 6. Probably this is because the multiple-choice test for these grades was given eleven weeks after the learning lessons, instead of only one or two weeks after the learning lessons.

Table 8 Proportion correct on the cued-recall test in the four learning conditions for the four grades and the children of the four grades together (*SD* in parentheses).

Grade	Short-lag restudy	Short-lag retrieval practice	Long-lag restudy	Long-lag retrieval practice
2	0.63 (0.20)	0.57 (0.21)	0.57 (0.24)	0.52 (0.21)
3	0.67 (0.15)	0.62 (0.14)	0.66 (0.15)	0.61 (0.13)
4	0.65 (0.17)	0.65 (0.15)	0.60 (0.17)	0.52 (0.15)
6	0.57 (0.15)	0.53 (0.16)	0.56 (0.16)	0.53 (0.15)
All	0.63 (0.17)	0.59 (0.17)	0.59 (0.19)	0.55 (0.17)

Table 9 Proportion correct on the multiple-choice test in the four learning conditions for the four grades and the children of the four grades together (*SD* in parentheses).

Grade	Short-lag restudy	Short-lag retrieval practice	Long-lag restudy	Long-lag retrieval practice
2	0.71 (0.22)	0.72 (0.22)	0.73 (0.19)	0.73 (0.19)
3	0.80 (0.12)	0.76 (0.14)	0.80 (0.13)	0.75 (0.17)
4	0.76 (0.20)	0.76 (0.23)	0.74 (0.20)	0.72 (0.19)
6	0.70 (0.16)	0.66 (0.19)	0.69 (0.18)	0.73 (0.18)
All	0.74 (0.18)	0.72 (0.20)	0.74 (0.18)	0.73 (0.18)

Analysis of the Scores on the Cued-Recall Test

We analyzed the scores on the cued-recall test using a 4x2x2 mixed ANOVA with grade as between-subject factor and distribution of learning (lag) and repetition of learning (repetition) as within-subject factors. The mixed design ANOVA on the cued-recall test showed that there was no effect of grade, F (3, 125) = 2.33, p = .078, η_p^2= .05.

There was a significant medium to large effect of lag, F (1, 125) = 25.58, p < .001, η_p^2= .17, indicating that there was a benefit of short-lag (M = 0.61, SD = 0.15, 95% CI [.586, .637]) over long-lag (M = 0.57, SD = 0.16, 95% CI [.545, .599]). There was also a significant interaction between lag and grade, F (3, 125) = 7.13, p < .001, η_p^2= .15. Further analyses on the effect of lag for each grade separately, were as follows. For Grade 2, the effect of lag was significant and large, F (1, 32) = 13.01, p = .001, η_p^2= .29, indicating that there was a benefit of short-lag (M = 0.60, SD = 0.19, 95% CI [.535, .668]) over long-lag (M = 0.55, SD = 0.20, 95% CI [.473, .618]). For Grade 3, there was no significant effect of lag (F < 1). For Grade 4, the effect of lag was significant and large, F (1, 30) = 30.81, p < .001, η_p^2= .51, indicating that there was a benefit of short-lag (M = 0.65, SD = 0.14, 95% CI [.599, .703]) over long-lag (M = 0.56, SD = 0.15, 95% CI [.505, .614]). For Grade 6, there was no significant effect of lag (F < 1). Thus, there was only a benefit of short-lag in Grade 2 and 4.

There was a significant small to medium effect of repetition, F (1, 125) = 15.44, p < .001, η_p^2=.11, indicating that there was a benefit of restudy (M = 0.61, SD = 0.16, 95% CI [.584, .641]) over retrieval practice (M = 0.57, SD = 0.15, 95% CI [.544, .598]). There was no significant interaction effect between repetition and grade (F < 1). Also, there was no significant interaction effect between lag and repetition (F < 1). Further, there was no significant interaction effect between lag, repetition and grade, F (3, 125) = 1.57, p = .200, η_p^2= .04.

Analysis of the Scores on the Multiple-Choice Test

We analyzed the scores on the multiple-choice test using a 4x2x2 mixed ANOVA with grade as between-subject factor and distribution of learning (lag) and repetition of learning (repetition) as within-subject factors. The mixed design ANOVA on the multiple-choice test showed that there was no significant effect of grade, F (3, 121) = 1.97, p = .122, η_p^2= .05. Also, there was no significant effect of lag (F < 1), indicating that there were no differences between short-lag (M = 0.73, SD = 0.16) and long-lag (M = 0.74, SD = 0.16). Also, there was no significant interaction between lag and grade (F <1). Furthermore, there was no significant effect of repetition, F (1, 121) = 1.77, p = .186, η_p^2= .01, indicating that there were no differences between restudy (M = 0.74, SD = 0.15) and retrieval practice (M = 0.73, SD = 0.16). Also, there was no significant interaction effect between repetition and grade, F (3, 121) = 1.18, p = .320, η_p^2= .03, no significant interaction effect between lag and repetition (F < 1), and no significant interaction effect between lag, repetition and grade, F (3, 121) = 1.04, p = .379, η_p^2= .03.

Discussion

The aim of our experiments was to investigate whether distributed practice and retrieval practice can enhance primary school vocabulary learning. In contrast to earlier conducted studies (e.g., Goossens et al., 2012; Goossens, Camp, Verkoeijen, & Tabbers 2014; Goossens, Camp, Verkoeijen, Tabbers, & Zwaan, 2014; Sobel et al., 2011), the experiment was fully integrated in the vocabulary learning curriculum. Furthermore, we did not investigate these memory strategies in only one grade, but in four different grades (Grade 2, 3, 4, and 6). The effect of distributed practice was investigated by comparing a short-lag condition in which words were learned during two sessions in one week, to a long-lag condition in which words were learned during four sessions in two weeks. The effect of retrieval practice was investigated by comparing a restudy condition in which children had to copy parts of the descriptions of the words, to a retrieval practice condition in which children had to recall the descriptions of the words. To establish the effects of both strategies, we administered two final tests: a cued-recall test on the short term in which children had to give descriptions of the words, and a multiple-choice test on the long term in which children had to choose the right descriptions of the words. In contrast to our expectations, these experiments showed benefits of short-lag practice (at least in two grades) and restudy practice on the cued-recall test, and no effects on the multiple-choice test. Thus, contrary to earlier studies into distributed practice and retrieval practice in primary school vocabulary learning (e.g., Goossens et al., 2012; Goossens, Camp, Verkoeijen, & Tabbers, 2014; Goossens, Camp, Verkoeijen, Tabbers, & Zwaan, 2014; Sobel et al., 2011), we did not find benefits of these memory strategies.

So how can we explain our findings? We will first describe possible procedural reasons for these results, and then we will describe the practical implications of the current study. One procedural difference between the distributed practice manipulation of the current study and of other studies (e.g., Goossens et al., 2012; Kornell, 2009; Sobel et al., 2011), is that even in our 'massed' condition, practice was distributed over several days, instead of on the same day. As a result, our study was more about the optimal study 'lag' rather than about the effect of spaced versus massed practice. Perhaps, the retention interval of the cued-recall test (1 to 3 days) suited the short-lag condition better (1 to 3 days), than the long-lag condition (1 to 6 days). It has been shown that the optimal lag between learning sessions is dependent on the retention interval (e.g., Cepeda et al., 2006; Cepeda, Vul, Rohrer, Wixted, & Pashler, 2008; Küpper-Tetzel, Kapler, & Wiseheart, 2014). When the retention interval increases, the optimal lag between learning sessions increases as well (i.e., to a certain extent). Thus for determining the optimal distribution of learning, the retention interval has to be taken into account. For practical use this means that for the planning of the lags between repetitions, we have to know when the final test will be given.

One procedural difference between the retrieval practice manipulation of the current study and of other studies (e.g., Goossens, Camp, Verkoeijen, & Tabbers, 2014; Karpicke & Smith, 2012), is that apart from the retrieval and restudy exercises, the participants performed textbook exercises. Thus, there was no 'pure' comparison between restudy and retrieval practice. Although in one study additional vocabulary exercises were done, these exercises were only done in the first learning session (Goossens, Camp, Verkoeijen, Tabbers, & Zwaan, 2014). In the current study, retrieval practice was implemented in only a small part of the exercises, which can explain why we did not find a benefit of retrieval practice. However, this cannot explain why there was a benefit of restudy on the cued-recall test. An explanation for this benefit might be that children put more effort in the restudy exercises than in the retrieval practice exercises. In contrast to the restudy exercises, the children could skip the retrieval practice exercises, without even trying to retrieve the words. Although the experimenter checked whether the children were doing their exercises seriously during the learning sessions, the children could still pretend that they had tried to retrieve the words, while they had not. As a result, time on task might have been longer in the restudy condition than in the retrieval practice condition, leading to better performance in the restudy condition.

Another procedural difference between earlier studies (e.g., Goossens et al., 2012; Goossens, Camp, Verkoeijen, & Tabbers, 2014; Goossens, Camp, Verkoeijen, Tabbers, & Zwaan, 2014; Sobel et al., 2011) and the current study conducted in regular lessons, is the fact that there was less experimental control on external factors, as time on task and as working silently. Although it might be argued that these factors influence only overall performance and that they therefore cannot account for the lack of benefits of distributed practice and retrieval practice, we think probably they can. For example, it is imaginable that concentration problems are more harmful in the long-lag condition and in the retrieval practice condition, than in the short-lag condition and in the restudy condition respectively, because these conditions ask more effort. That is, the longer the lag, the more difficult study-phase retrieval will be (e.g., Cepeda et al., 2006), and retrieving a description will be more difficult than rewriting parts of a description.

One possible explanation for not finding benefits of distributed practice and retrieval practice might be that the children did not learn many additional words between learning sessions. In particular, children recalled 56% of the words in the first retrieval practice exercise, 58% in the second retrieval practice exercise, and 59% in the final test. However, if the children really did not learn any additional words between learning sessions, we would not have found benefits of short-lag practice and restudy practice as well. Also, we do not know how many words the children knew at the start of the learning sessions and immediately after the last learning session. Therefore, in further research it is recommended to give at least a pretest.

Although there are procedural differences between our design and designs used in other studies into distributed practice and retrieval practice in primary school vocabulary learning (e.g., Goossens et al., 2012; Goossens, Camp, Verkoeijen, & Tabbers, 2014, Goossens, Camp, Verkoeijen, Tabbers, & Zwaan, 2014; Sobel et al., 2011), our design helps us to generalize to the real-life classroom setting. Therefore, our results have a potential high practical relevance for primary school vocabulary learning. One important point is the fact that in the current study we found differences of 4% between the short-lag and the long-lag condition and also of 4% between the retrieval practice and the restudy condition on the cued-recall test. Although these percentages do not seem to be very high, for the short-lag condition this effect was medium to large, and for the restudy condition this effect was small to medium. Considering that children practice 10 words a week during 40 weeks a school year, the benefits of the short-lag condition and the restudy condition may be estimated around 16 words a year. Again, this number of words does not seem to be very high. Nevertheless, this number of words learned intentionally, will have an influence on the number of words learned incidentally as well. Children with larger vocabulary sizes will have less difficulties to learn additional words than children with smaller vocabulary sizes. In this way, the differences between children will become larger over time, which is also called the Matthew effect (Stanovich, 1986).

Although it is suggested that distributed practice and retrieval practice are promising memory strategies for the classroom (e.g., Dunlosky et al., 2013), the results of this study show that these strategies cannot be applied without thought to primary school vocabulary learning. Before we can give practical guidelines about distributed practice and retrieval practice, we have to know more about the relation between lag and retention interval (e.g., Cepeda et al, 2006; Cepeda et al., 2008; Küpper-Tetzel, et al., 2014), and about the number of retrieval practice exercises needed. Furthermore, it is not known why the results on the cued-recall test and on the multiple-choice test were dissimilar. Maybe the results on the multiple-choice test were caused by the fact that the multiple-choice test was given on a relatively long term, or by the fact that the children had received a final cued-recall test on the words already. Both explanations seem appropriate, but they both suggest that finding benefits of distributed practice and retrieval practice on a long-term multiple-choice test are conditional on factors that are very important in the classroom. In particular, in education it is important that positive effects of learning strategies are still visible after a longer delay, and also after (several) intervening tests, because this is inherent to a real-life classroom situation. Therefore, it is important to investigate why we did find benefits of short-lag practice and restudy practice on the short-term cued-recall test, and why we did not find benefits of distributed practice and retrieval practice on the long term multiple-choice test. All in all, the results on both the final cued-recall test and the multiple-choice test give some food for thought about the practical value of

distributed practice and retrieval practice as instructional strategies in primary school vocabulary learning.

Conclusion

We investigated distributed practice and retrieval practice in real-life primary school vocabulary learning settings in four different grades. In contrast to our expectations, and in contrast to earlier experiments into primary school vocabulary learning (e.g., Goossens et al., 2012; Goossens, Camp, Verkoeijen, & Tabbers, 2014; Goossens, Camp, Verkoeijen, Tabbers, & Zwaan, 2014; Sobel et al., 2011) we found benefits of the short-lag condition and the restudy condition on the cued-recall test, and no differences between conditions on the multiple-choice test. These results show that we cannot simply apply memory strategies from the laboratory to a regular classroom situation, and that we cannot simply provide guidelines to teachers to use distributed practice and retrieval practice in primary school vocabulary lessons. Additional research into these two memory strategies is needed. By doing this, we could explore further whether and to what extent distributed practice and retrieval practice can be implemented in real-life primary school vocabulary lessons for different age groups.

Acknowledgements

The Board of Public Education Rotterdam (Stichting BOOR) provided financial support for this research. The authors would like to thank the director and teachers of the school for their cooperation. Furthermore, we thank the children for participating in this research. Also, we thank Pim Renkema for scoring the final cued-recall tests.

Chapter 6

Summary and General Discussion

The aim of this thesis was to investigate whether particular memory strategies stemming from cognitive and educational psychology, enhance primary school vocabulary learning. The memory strategies investigated in this thesis were distributed practice and retrieval practice. The *distributed practice effect* refers to the phenomenon that distributed practice is better for long term retention than massed practice. In other words, if someone has to learn new words, it is better for him to learn the words across multiple learning sessions, than to repeat them within one single learning session. The distributed practice effect has been shown in more than 300 experiments (for reviews, see e.g., Cepeda, Pashler, Vul, Wixted, & Rohrer, 2006; Delaney, Verkoeijen, & Spirgel, 2010). The *retrieval practice effect* refers to the phenomenon that retrieval practice is better for long-term retention than restudy. Put differently, if someone has to learn new words, it is better for him to try to retrieve the words than to repeat them by additional study. The retrieval practice effect has been demonstrated in more than 100 experiments (for reviews, see e.g., Rawson & Dunlosky, 2011; Roediger & Butler, 2011; Roediger & Karpicke, 2006; Roediger, Putnam, & Smith, 2011). Both distributed practice and retrieval practice have been investigated many times in the laboratory, but not many times in the classroom.

The main research questions of this thesis were as follows: (1) Is distributed practice a beneficial memory strategy for primary school vocabulary learning? (2) Is retrieval practice a beneficial memory strategy for primary school vocabulary learning? (3) Can these memory strategies be successfully integrated in current primary school vocabulary lessons? To answer these questions the distributed practice effect has been investigated in a simulated classroom setting (Chapter 2) and in a real-life classroom setting (Chapter 5). The retrieval practice effect has been investigated in a laboratory setting (Chapter 3), in a simulated classroom setting (Chapter 4), and in a real-life classroom setting (Chapter 5).

Summary of the Main Results

In Chapter 2 the research question was whether distributed practice benefits primary school vocabulary learning in a simulated classroom setting in Grade 3. In this study, a comparison was made between a spaced condition in which the words were repeated once in three different learning sessions, and a massed condition in which the words were repeated three times in one single learning session. Because the words were repeated in a different session than the session in which the words were introduced, the massed condition was relatively 'less massed' than in other experiments (e.g., Sobel, Cepeda, & Kapler, 2011). Furthermore, in this study different types of exercises from the regular vocabulary learning material of Grade 4 were used. This departs from earlier research into the distributed practice effect in (primary school) vocabulary learning, because the repeated exercises in these studies were identical (e.g., Kornell, 2009; Sobel et al.,

2011). The external validity of these studies was therefore limited, because in vocabulary learning different types of exercises are used (e.g., Blachowicz, Fisher, Ogle, & Watts-Taffe, 2006). Also, earlier experiments have shown that encoding variability can eliminate the distributed practice effect (e.g., Dellarosa & Bourne, 1985; Gartman & Johnson, 1972). Thus, it was not clear whether using different exercises would have an influence on the distributed practice effect. In the final tests after one week and after five weeks the children had to provide the correct word to its given definition. On both tests a benefit of distributed practice was found.

In Chapter 3 the research question was whether we would find a benefit of retrieval practice in primary school target-synonym word-pair learning in a laboratory setting in Grade 3. In this study a comparison was made between a restudy condition in which children had to read target-synonym word-pairs aloud, and a retrieval practice condition in which children had to try to retrieve the synonyms when the target was given. The second question was whether introducing the words in a meaningful context would affect the hypothesized benefit of retrieval practice. Therefore, the words were introduced in two different conditions. In the story condition, the words were introduced by a story that was told by the experimenter, and in the word-pairs condition the words were introduced without any contextual information, by presenting only the target words and their synonyms. This second question was formulated because in most primary school vocabulary learning methods, words are presented in a meaningful context (e.g., Janssen & Van Ooijen, 2012; Van de Gein, Van de Guchte, & Kouwenberg, 2008). Furthermore, presenting words in a meaningful context has been demonstrated to benefit vocabulary learning (for a review, see Stahl & Fairbanks, 1986). Thus, the question whether there would still be an advantage of retrieval practice after presenting the words in a meaningful context has practical relevance. After one week the children had to do a cued-recall test and a multiple-choice test, in which the children were asked to retrieve and to recognize the synonyms of the words. On the multiple-choice test there was no difference between the restudy and retrieval practice conditions, but there was a benefit of the word-pairs condition compared to the story condition. Most importantly, the results of the cued-recall test showed a benefit of retrieval practice over restudy. This indicates that retrieval practice may indeed benefit vocabulary learning. Secondly, a benefit of the word-pairs condition over the story condition was found. Finally, there was a marginally significant interaction between learning condition and context, showing that the benefit of retrieval practice was somewhat larger in the word-pairs condition than in the story condition. These results have to be interpreted with caution, because this difference was marginally significant and because the children in the word-pairs condition already retrieved more words during the first retrieval practice session than the children in the story condition. However, the benefit of the word-pairs condition was not in line with our expectations,

but consistent with other studies in which no benefit of context was found in vocabulary learning (e.g., Jones, Levin, Levin, & Beitzel, 2000; McDaniel & Pressley, 1984, 1989). Possibly, the contextual information diverted the children from learning the word pairs. Also, the form of the cued-recall test matched better with the word-pairs condition than with the story condition. Thus, better performance in the word-pairs condition may be due to transfer-appropriate processing (e.g., Blaxton, 1989; Morris, Bransford, & Franks, 1977; Thomas & McDaniel, 2007).

In Chapter 4 the research question was whether we would find a benefit of retrieval practice in primary school vocabulary learning in a simulated classroom setting in Grade 3. Because retrieval practice was compared to both pure restudy and elaborative restudy, this study was more ecologically valid than the study described in Chapter 3. Pure restudy is hardly ever used in real-life vocabulary lessons and thus any differences in effectiveness of pure restudy versus retrieval practice do not inform educational practice. In real-life vocabulary lessons words are repeated by elaborative exercises (e.g., Janssen & Van Ooijen, 2012; Van de Gein et al., 2008). Furthermore, previous research on vocabulary learning has shown that the addition of contextual cues aids vocabulary learning (e.g., Anderson & Reder, 1979; Blachowicz et al., 2006; Bolger, Balass, Landen, & Perfetti, 2008; Carey, 1978, Coomber, Ramstad, & Sheets, 1986). Therefore, the question was whether retrieval practice would add something to vocabulary learning lessons in which elaborative exercises were included. The study consisted of two experiments that only differed in the first learning session. In the first learning session of the study, the children were introduced to the words by pictures and definitions (Experiment 1) or by a story (Experiment 2), followed by a number of exercises that were focused on the word form. In the second learning session, the children had to do two consolidation exercises according to their learning conditions. In the pure restudy condition the children had to copy a part of the definition, in the elaborative restudy condition the children had to do exercises with semantically related words, and in the retrieval practice condition the children had to recall the words based on their definition. After one week the children had to do a fill-in-the-blank test in which they had to retrieve the word based on the given definition, and a multiple-choice test in which they had to choose the right word for the right context sentence. A combined small-scale random-effects meta-analysis on the results of the two experiments showed a retrieval practice effect on the fill-in-the-blank test. There was a small difference between retrieval practice and pure restudy and also between retrieval practice and elaborative restudy. No difference was found between pure restudy and elaborative restudy. The combined CIs showed that there was much uncertainty about the magnitude of the retrieval practice effect in the population. To obtain a more precise parameter estimate – and hence to inform educational practice better – more research is needed. The multiple-choice test showed no differences between conditions, probably due

to ceiling effects. Importantly, the lack of a difference between pure restudy and elaborative restudy was unexpected, but consistent with the results in Chapter 3 and with the results of other studies (e.g., Jones et al., 2000; McDaniel & Pressley, 1984, 1989). Probably, this finding was due to the fact that the tasks in the elaborative restudy condition were less focused on the word definition than in the pure restudy condition.

In Chapter 5 the research question was whether we would find benefits of distributed practice and retrieval practice in primary school vocabulary learning in a real-life classroom setting in different grades of primary education (Grade 2, 3, 4, and 6). The main research questions were the same as in Chapter 2, 3, and 4, but because the study was conducted in the regular vocabulary learning curriculum, this study was more ecologically valid than the other studies. The ecological validity of the study was enhanced by comparing a long-lag condition to a short-lag condition instead of to a 'pure' massed condition. In the short-lag condition, lags were used that were comparable to the lags used between repetitions in the regular vocabulary learning lessons. That is, the short-lag condition in which all words were learned during two lessons in the same week, was compared to a long-lag condition in which all words were learned during four lessons in two weeks. In addition, the children practiced the words not only in restudy or retrieval practice exercises, but also in regular exercises of their curriculum. As is the case in many vocabulary-learning methods (e.g., Janssen & Van Ooijen, 2012; Van de Gein et al., 2008) the words were practiced with different types of exercises. In half of the learning sessions the children started the session with repetition of the words by restudy exercises, in which they had to copy parts of descriptions of the words, and by retrieval practice exercises, in which they had to retrieve descriptions of the words. These exercises were followed by exercises from the regular vocabulary learning curriculum. At the end of each week the children had to do a cued-recall test in which they wrote down the definitions of the words that had been repeated six times, namely four times in the regular vocabulary learning exercises and twice in the repetition exercises (either restudy or retrieval practice). Several weeks after the experiment (one week for Grade 6, two weeks for Grade 3, and eleven weeks for Grade 2 and 4) the children had to do a multiple-choice test in which they had to choose the right description of the word out of four options. In contrast to our expectations, we found beneficial effects of both short-lag practice and restudy on the final cued-recall test. However, the positive effects of short-lag practice were only significant for Grade 2 and Grade 4. Furthermore, on the multiple-choice test we found no difference between short-lag and long-lag practice and we also found no difference between restudy and retrieval practice. Because the multiple-choice test was comparable to tests used in the classroom, there are some concerns about the practical value of distributed practice and retrieval practice.

In sum, we found a benefit of distributed practice in a simulated classroom setting (Chapter 2), but not in a real-life classroom setting (Chapter 5). Furthermore, we found a benefit of retrieval practice in a laboratory setting (Chapter 3), and in a simulated classroom setting (Chapter 4), but not in a real-life classroom setting (Chapter 5).

General Discussion

The studies presented in this thesis were the first to investigate the effect of distributed practice and retrieval practice in the context of primary school vocabulary learning. First, I will discuss the findings of the studies separately for each memory strategy. Then, I will discuss the theoretical and practical implications of the findings for both memory strategies together.

Discussion of the Results on Distributed Practice

A notable finding in this thesis was that there was a benefit of distributed practice in a simulated classroom setting (Chapter 2), but not in a real-life classroom setting (Chapter 5). The fact that there was a medium to large benefit of distributed practice in Chapter 2 was very promising for application in an educational context, because in contrast to other studies (e.g., Kornell, 2009; Sobel et al., 2011), in this study (1) the massed condition was relatively less massed than in other experiments; (2) words and exercises from vocabulary learning material (of one grade higher) were used; (3) the children performed different types of exercises. In contrast, the results of Chapter 5 are less promising for application of distributed practice in education. In Chapter 5, on the cued-recall test a medium to large benefit was found of the short-lag condition and on the multiple-choice test no difference was found between the short-lag condition and the long-lag condition. This study was more ecologically valid than the study in Chapter 2, because (1) compared to the lags of the massed condition in Chapter 2, the lags in the short-lag condition were longer; (2) words and exercises from the regular vocabulary learning curriculum were used.

One possible explanation of the differences between Chapter 2 and Chapter 5 could be the fact that the distributed practice manipulation was different for these two studies. The study in Chapter 5 focused more on the optimal study lag than on the effect of spaced versus massed practice. If the lag between study sessions increases, performance will increase as well. However, if the lag increases too much, performance will decrease slightly, as represented by an inverted U-curve (e.g., Toppino & Bloom, 2002; Verkoeijen, Rikers, & Schmidt, 2005). Even so, when the lag increases too much, there will be no study-phase retrieval of the information from an earlier presentation anymore. As a result, the memory trace will be not strengthened (e.g., Delaney et al., 2010). Perhaps in Chapter 5 the lags of the long-lag condition were too large, resulting in worse study-phase retrieval

than in the short-lag condition. For practical use these results are quite complex, because they raise the question at what point exactly the lag becomes too large.

Another possible explanation could be the fact that the optimal lag between learning sessions is dependent on the retention interval (e.g., Cepeda et al., 2006; Cepeda, Vul, Rohrer, Wixted, & Pashler, 2008; Küpper-Tetzel, Kapler & Wiseheart, 2014). That is, the larger the retention interval, the larger the optimal lag. The retention intervals of the two studies were different: In Chapter 2 the retention intervals were 7 and 35 days, and in Chapter 5 the retention intervals were 1 to 3 days and 14 to 77 days. Thus, in Chapter 5 the first test was given after a relatively short retention interval. Perhaps the retention interval of this test (1 to 3 days) suited the short-lag condition better (1 to 3 days) than the long-lag condition (1 to 6 days). That is, with shorter retention intervals probably shorter lags are needed to have optimal retention.

On the whole, the results of this thesis imply that distributed practice is beneficial for primary school vocabulary learning. Even so, there are indications that the lags can become too large as well, resulting in worse performance on a final test. Thus, for practical use it is important to take into account the optimal lag for a certain retention interval. Thus, the longer the retention interval, the longer the lag has to be, and the shorter the retention interval, the shorter the lag has to be. Even though the shorter lags led to better test performance than the longer lags (Chapter 5), this thesis has shown that distributing vocabulary exercises over different learning sessions is better for retention than massing them in one learning session (Chapter 2). This means that although in some vocabulary learning methods small sets of words are practiced in one single learning session, it is better to practice these words in exercises that are distributed over different learning sessions. Further, Cepeda et al. (2008) have shown that using lags that are longer than optimal will be less harmful for retention than lags that are shorter than optimal. That is, if the lag increases, learning performance will first increase steeply and then decline more gradually.

Discussion of the Results on Retrieval Practice

A notable finding regarding the effect of retrieval practice in vocabulary learning was that there were benefits of retrieval practice in a laboratory setting and in a simulated classroom setting (Chapters 3 and 4), but not in a real-life classroom setting (Chapter 5). The fact that there was a medium to large benefit of retrieval practice on the cued-recall test in Chapter 3 was promising for primary school vocabulary learning, because in contrast to other studies (1) primary school children were learning word pairs; (2) uncommon Dutch words were used; (3) for half of the children the words were introduced within a meaningful context. In addition, the small to medium benefit of retrieval practice on the fill-in-the-blank test in Chapter 4 was even more promising, because in this study

(1) the retrieval practice condition was compared to a more elaborative restudy condition; (2) words and exercises from current vocabulary learning material (of one grade higher) were used; (3) the words were repeated in two different learning sessions. However, the results of Chapter 5 were not that promising, because there was a small to medium benefit of restudy on the cued-recall test and because there was no difference between restudy and retrieval practice on the multiple-choice test. This study was more ecologically valid than the other studies, because (1) the retrieval practice exercises were additional to the regular exercises; (2) words and exercises from the regular vocabulary learning curriculum were used.

One possible explanation of the different results could be the fact that the retrieval practice manipulation was different in the three chapters. In Chapter 5 there was more variation within the restudy and retrieval practice conditions than in Chapters 3 and 4, because of the added textbook exercises from the standard vocabulary-learning curriculum. As a result, in Chapter 5 restudy and retrieval practice were implemented in only a small part of the exercises, which may have diluted the positive effect of retrieval practice.

Another possible explanation of the different results could be that in Chapter 5 relatively more effort was put into the restudy exercises than into the retrieval practice exercises, because, in contrast to the restudy exercises, the retrieval practice exercises could easily be skipped. That is, in contrast to in Chapters 3 and 4, in Chapter 5 the children had to retrieve the whole description of the word, instead of only the synonym or the target word itself. As a result, in Chapter 5 it was relatively easy for the children to pretend that they had tried to retrieve the description of the words, while they in fact had not. It is quite difficult to retrieve a whole word description and therefore it could be expected that the children would encounter problems in retrieving the word descriptions. In contrast, in Chapter 3 the children could not pretend falsely that they had tried to retrieve the word descriptions while they in fact had not, because the children learned the words individually with the experimenter. Furthermore, although in Chapter 4 it was also possible to skip the words, it is less likely that children skipped the exercises, because in this study they only had to write down the target word in the retrieval practice condition, which requires less effort than writing down the whole description of the word. As a result, in Chapter 5 the children could have invested relatively more time in the restudy condition than in the retrieval practice condition.

On the whole, the results of this thesis imply that retrieval practice can help primary school vocabulary learning, but that benefits of retrieval practice are not always found. The contrasting results in Chapters 3, 4, and 5 suggest that the benefit of retrieval practice depends on the way in which retrieval practice is implemented and is contrasted with control conditions. Furthermore, in Chapter 4, the combined confidence intervals of the

two experiments showed that the benefit of retrieval practice was small, and that there was much uncertainty about the estimate of the magnitude of the retrieval practice effect in the population.

Theoretical Implications

The present thesis provided evidence that distributed practice and retrieval practice may benefit primary school vocabulary learning in several learning settings, but not in all. The use of an elaborative context in Chapter 3, and the use of elaborative exercises in Chapters 2, 4, and 5, can inform us about the effects of elaboration on the distributed practice effect and the retrieval practice effect and also about the effects of elaboration on vocabulary learning in general.

The encoding variability hypothesis or contextual variability hypothesis has been proposed as an account of the distributed practice effect (e.g., Melton, 1967). According to this hypothesis, distributed practice is better for retention than massed practice, because items learned by distributed practice occur in multiple contexts, instead of in only one context, as is the case in massed practice. Because of these multiple contexts, people have more retrieval routes to access the learning material in a final test for the distributed items than for the massed items. Related to that, one might argue that the benefit of distributed practice will disappear when elaboration is added to learning, because then the learning material will be already encoded variable in different exercises. In some experiments the distributed practice effect disappeared when encoding variability was enhanced (e.g., Dellarosa & Bourne, 1985; Gartman & Johnson, 1972). Even so, the results of Chapter 2 are not in line with this hypothesis, because in this study we used different elaborative vocabulary exercises. Similarly, in a study by Smith and Rothkopf (1984) encoding variability did not affect the magnitude of the distributed practice effect. Hence, it does not seem likely that elaborative exercises reduce benefits of distributed practice. All the same, because we did not use a condition without elaborative exercises, we cannot exclude the possibility that elaborative vocabulary exercises reduce benefits of distributed practice.

The elaborative retrieval hypothesis has been proposed as an account of the retrieval practice effect (see for example Carpenter, 2009, 2011). According to this hypothesis, people elaborate on learning material when they are trying to retrieve it, but not or less so, when they are restudying it. As a result, it is possible that the addition of elaborative exercises in a restudy condition will diminish the benefit of retrieval practice, because differences between restudy and retrieval practice will become smaller. However, the results of Chapters 3 and 4 are not in line with this hypothesis. In Chapter 3 a retrieval practice effect was found when the words were introduced in an elaborative context and in Chapter 4 the retrieval practice condition outperformed the elaborative restudy condition. Further, the results of Chapters 3 and 4 are in line with the results of other

studies (e.g., Blunt & Karpicke, 2014; Karpicke, Lehman, & Aue, 2014; Karpicke & Smith, 2012; Lehman, Smith, & Karpicke, 2014) in which the elaborative retrieval hypothesis was not supported.

In general, in vocabulary learning it is assumed that elaborative repetition exercises are needed to gain deeper word knowledge (e.g., Blachowicz et al., 2006). Moreover, it has been found that a combination of definitions with active processing, such as adding contextual information, is more effective than definition instruction alone (for a review, see for example Blachowicz & Fisher, 2000). Even so, in Chapters 3 and 4 no benefits were found of elaborative (contextual) cues. For instance, in Chapter 3 the word-pairs condition outperformed the story condition, and in Chapter 4 the elaborative restudy condition performed similar to the pure restudy condition. Adding elaborative information in vocabulary learning was not beneficial for vocabulary learning in other studies as well (e.g., Jones et al., 2000; McDaniel & Pressley, 1984, 1989) and active processing was not beneficial for definition learning (e.g., Graves, 2006; Nation, 2001). All in all, rote repetition may be effective as well (Nation, 1982), and knowing the synonym of the word is often enough for understanding the word in the context (Stahl & Nagy, 2006). Therefore, it is questionable whether elaborative encoding is beneficial in every learning situation. It might be argued that elaboration can even harm retention of words and their meaning, because more (irrelevant) information about the words will be added that is not needed to learn the word definition. The question is whether elaboration on the words is always meaningful. To answer this question, first the term 'elaboration' has to be defined more precisely, because elaboration has been quite vaguely defined in terms of 'adding information to a memory trace, between memory traces, or to memory cues' (see Lehman et al., 2014).

Practical Implications

This thesis has shown both positive and negative effects of distributed practice and retrieval practice in different primary school situations by using different vocabulary learning materials. I think there are some limitations in the extent to which these memory strategies are beneficial to primary school vocabulary learning.

First, the benefit of distributed practice has been shown by comparing a spaced condition in which children did one exercise during three learning lessons to a massed condition in which children did three exercises during one learning lesson (Chapter 2). For practical use this means that in contrast to cramming all word exercises in one day, as is the case in some vocabulary learning methods, it is better to distribute the word exercises over a week. However, in this thesis it also has been shown that distributing exercises over two days in one week (short-lag condition), instead of over four days in two weeks (long-lag condition), leads to better performance on a final test (Chapter 5).

Possibly in Chapter 5 there was a better match between the retention interval and the interstudy interval of the short-lag condition, than between the retention interval and the interstudy interval of the long-lag condition. For practical use these results show that distributed practice can enhance performance on a final test, but only to a certain extent. All in all, it is better to repeat words distributed over different days, than to do all word exercises during one day. Furthermore, it may be helpful to repeat the words more than once on the days on which the words are repeated, as was done in the short-lag condition in Chapter 5.

Second, in this thesis it has been shown that retrieval practice is beneficial for vocabulary learning when a retrieval practice condition is compared to a (pure) restudy condition (Chapters 3 and 4) and to an elaborative restudy condition (Chapter 4). In practical use this means that for children it is better to try to retrieve the learning material themselves, than to copy it or to do regular elaborative exercises. However, in this thesis it also has been shown that restudy practice is more beneficial for vocabulary learning than retrieval practice when these repetition exercises are only a small subset of all exercises (Chapter 5). A possible explanation for these contrasting results is that the benefit of retrieval practice depends on the number of retrieval practice exercises in proportion to the number of regular learning exercises. However, this would mean that retrieval practice may not always be helpful in primary school vocabulary learning, because in general many regular vocabulary exercises are used. Perhaps replacing a larger subset of the regular exercises by retrieval practice exercises will lead to a benefit of retrieval practice instead.

Third, the positive effects of distributed practice and retrieval practice were not found on all kinds of final tests. For instance, in Chapter 5 there were no benefits of distributed practice and retrieval practice on the description test, and in Chapters 3, 4, and 5, there were no benefits of distributed practice and retrieval practice on the multiple-choice tests. However, it may be that the multiple-choice tests were influenced by the preceding tests. For example in Chapter 4 ceiling effects were found, that were possibly caused by the preceding fill-in-the-blank test. There were only benefits of distributed practice and retrieval practice on fill-in-the-blank tests (Chapters 2 and 4) and on cued-recall tests (Chapter 3). From a practical view, the question is whether cued-recall and fill-in-the-blank tests are generalizable to tests typically used in the classroom. In general, in the regular vocabulary learning methods mostly multiple-choice tests are used that are comparable to the multiple-choice tests used in Chapter 4 (e.g., Janssen & Van Ooijen, 2012; Van de Gein et al., 2008). Furthermore, the vocabulary tests to measure overall word knowledge are more comparable to the multiple-choice test used in Chapter 5 (e.g., Van Berkel & Alberts, 2009; Van Berkel & Hilte, 2009). It could be argued that retention of words and their definitions is not enough for deep and meaningful vocabulary learning, and that therefore the cued-recall and fill-in-the-blank tests are not ecologically valid enough, because they

only measure *memory* of the words. However, I think that in vocabulary development, retention of a short description of the words is an important first step for developing deeper vocabulary knowledge. For instance, there is a strong correlation between breadth and depth of vocabulary knowledge (Vermeer, 2001). Thus, I expect that although in the current thesis only effects on retention tasks were found, comparable results may be found on tasks that require deeper vocabulary knowledge.

Fourth, the results of this thesis suggest that primary school vocabulary learning is not enhanced by elaboration, even though it is assumed that elaborative exercises help vocabulary learning (e.g., Blachowicz et al., 2006; Blachowicz & Fisher, 2000). For example, in Chapter 3 the word-pairs condition outperformed the story condition, and in Chapter 4 the performance was similar for the elaborative restudy condition and the pure restudy condition. Furthermore, the results of these studies are consistent with other studies in which no benefits of elaboration were found in vocabulary learning (e.g., Graves, 2006; Jones et al., 2000; McDaniel & Pressley, 1984, 1989; Nation, 2001). Thus, it is questionable whether elaborative encoding is beneficial in every learning situation, and whether rote repetition can lead to sufficient recall of synonyms or definitions of words.

Conclusion

The studies described in this thesis are the first studies into the distributed practice effect and the retrieval practice effect in primary school vocabulary learning. In three of these studies beneficial effects of distributed practice and retrieval practice were found. However, in one study no benefits of distributed practice and retrieval practice were found. These studies together show that there are some boundaries in the extent to which distributed practice and retrieval practice may be beneficial in primary school vocabulary learning. All in all, by conducting these studies a start has been made in bridging the gap between well-established memory strategies on the one hand, and the educational practice of primary school vocabulary learning on the other hand. These studies give us new insights in to what extent distributed practice and retrieval practice are beneficial in primary school vocabulary learning.

Suggestions for Future Research

There are some remaining questions, which may be answered by future research. One question is to what extent the benefits of distributed practice and retrieval practice depend on the characteristics of children, thus the individual differences beyond prior knowledge and ability to learn new words (e.g., Dunlosky, Rawson, Marsh, Nathan, & Willingham, 2013). For instance, one study in which primary school children had to learn word lists showed differences in how much the children benefited from retrieval practice

(Bouwmeester & Verkoeijen, 2011a). In this study three groups were distinguished: one group that showed no benefit of retrieval practice, a group that showed a small benefit of retrieval practice, and another group that showed a greater benefit of retrieval practice. Because prior knowledge and learning ability influence vocabulary learning, it has to be investigated whether, and in to what extent, children with low vocabulary sizes and/or learning problems, will benefit from distributed practice and retrieval practice.

Another question is at which lag a maximum memory performance is obtained using distributed practice. It has been shown that the optimal lag increases with the length of the retention interval, to a certain extent, as shown in an inverted u-shape relationship between lag and memory performance (Cepeda et al., 2006). In addition, distributed schedules of retrieval are better for retention than massed schedules of retrieval (e.g., Cull, 2000; Cull, Shaughnessy, & Zechmeister, 1996). It is assumed that greater lags between learning sessions increase retrieval effort, and therefore promote better retention (e.g., Jacoby, 1978; Karpicke & Roediger, 2007; Modigliani, 1976), as long as the learner can access the material and is able to recall it on the test (e.g., Spitzer, 1939). The fact that in Chapter 5 there was a benefit of the short-lag condition on the short-term test, and no difference between the conditions on the long-term test, seems to contradict the idea that lags have to be larger for optimal retention. However, because the words in primary school vocabulary learning have to be retained for a whole life, in future research directed at distributed practice and retrieval practice in primary school vocabulary learning, the optimal interaction between lag and retention interval has to be examined.

Another remaining question is whether, and if so, to what extent, the retrieval practice effect is affected by the presence of other regular exercises. In contrast to in Chapters 3 and 4, in Chapter 5 the retrieval practice manipulation was additional to the regular exercises, thus there was a relatively small number of retrieval practice exercises, and possibly this was the reason that there was not found a retrieval practice effect. Therefore, in future research directed at retrieval practice in primary school vocabulary learning, it might be useful to examine whether, and if so, to what extent, the retrieval practice effect is conditional on the number of retrieval practice exercises in proportion to the number of regular exercises.

Finally, future research into distributed practice and retrieval practice should be directed at the applicability of these strategies in real-life primary school vocabulary lessons. It is important to investigate the magnitude and homogeneity of these effects in a real-life learning setting. In Chapter 4 we found that there was much uncertainty about the magnitude of the retrieval practice effect. Furthermore, in Chapter 5 we found that the short-lag effect was quite heterogeneous over different grades, namely the reverse distributed practice effect was only found in Grade 2 and 4, while in Grade 3 and 6 there was no effect of distributed practice. There is also uncertainty about the practical value

of the memory strategies because there were benefits of distributed practice and retrieval practice on cued-recall and fill-in-the-blank tests, but no benefits on multiple-choice tests. Therefore, it is important to do more research on the effect of these memory strategies on different final tests, that are used in the classroom (for a review in which the importance of the use of different criterion tasks is emphasized, see Dunlosky et al., 2013). The results of this thesis give us some first insights in the effects of distributed practice and retrieval practice in primary school vocabulary learning, and are therefore very informative for future research.

Dit proefschrift gaat over de vraag hoe nieuwe woorden effectief aan basisschoolkinderen kunnen worden geleerd. Woordenschatontwikkeling is belangrijk voor goed tekstbegrip (Anderson & Freebody, 1981). Hoe groter je woordenschat, hoe preciezer en complexer je kunt spreken en denken over de wereld (Stahl & Nagy, 2006). Zo zal iemand die de woorden *grillen*, *stoven* en *pocheren* kent, anders over koken denken (en misschien zelfs anders koken), dan iemand die enkel de woorden *bakken*, *koken* en *braden* kent (naar voorbeelden van Stahl & Nagy, 2006, pagina 5). In dit proefschrift is onderzoek gedaan naar de vraag of twee geheugenstrategieën, voortkomend uit onderzoek in de cognitieve psychologie, effectief zijn bij het leren van nieuwe woorden.

De geheugenstrategieën die ik heb onderzocht zijn *distributed practice* en *retrieval practice*, ook wel bekend als het spreiden van leermomenten en het ophalen van informatie uit het geheugen. Het *distributed practice effect* is het fenomeen dat het spreiden van leren beter is voor het langetermijngeheugen dan het zogenoemde stampen. Met andere woorden, als iemand nieuwe woorden moet leren, is het beter om de woorden verspreid over meerdere leersessies te leren, dan deze steeds opnieuw te herhalen binnen een enkele leersessie. Het *distributed practice effect* is aangetoond in meer dan 300 experimenten (zie voor overzichtsartikelen bijvoorbeeld Cepeda, Pashler, Vul, Wixted & Rohrer, 2006; Delaney, Verkoeijen & Spirgel, 2010). Het *retrieval practice effect* is het fenomeen dat het ophalen van informatie beter is voor het langetermijngeheugen dan het opnieuw bestuderen van informatie. Anders gezegd, als iemand nieuwe woorden moet leren, is het beter om de betekenis van de woorden proberen op te halen tijdens het leren, dan deze woorden en hun betekenis steeds maar opnieuw te bestuderen. Het *retrieval practice effect* is aangetoond in meer dan 100 experimenten (zie voor overzichtsartikelen bijvoorbeeld Rawson & Dunlosky, 2011; Roediger & Butler, 2011; Roediger & Karpicke, 2006; Roediger, Putnam & Smith, 2011). Zowel gespreid leren als het ophalen van informatie zijn al vaak onderzocht in gecontroleerde contexten, maar nog niet vaak in de klas.

De hoofdvragen van dit proefschrift waren als volgt: (1) Helpt het spreiden van leren bij woordenschatverwerving van basisschoolkinderen? (2) Helpt het ophalen van informatie bij woordenschatverwerving van basisschoolkinderen? (3) Kunnen deze geheugenstrategieën succesvol worden geïntegreerd in huidige woordenschatlessen? Om deze vragen te beantwoorden is het *distributed practice effect* onderzocht in een gesimuleerde klassencontext (Hoofdstuk 2) en in een ware klassencontext (Hoofdstuk 5). Het *retrieval practice effect* is onderzocht in een meer gecontroleerde context (Hoofdstuk 3), in een gesimuleerde klassencontext (Hoofdstuk 4) en in een ware klassencontext (Hoofdstuk 5).

Samenvatting van de Resultaten

In Hoofdstuk 2 was de onderzoeksvraag of het spreiden van leren helpt bij het leren van woordenschat in een gesimuleerde klassencontext in groep 5. Na een sessie waarin dertig nieuwe woorden werden geïntroduceerd, werd het gespreid oefenen vergeleken met het gestampt oefenen. In de gespreid leren conditie werd de helft van de woorden gedurende drie verschillende oefensessies één keer geoefend. In de gestampt leren conditie werd de andere helft van de woorden gedurende één oefensessie drie keer geoefend. Omdat de woorden in andere sessies werden geoefend dan in de introductiesessie, werd in deze studie tijdens het gestampt oefenen, minder gestampt dan in andere studies (bijvoorbeeld, Sobel, Cepeda & Kapler, 2011). Ook werden er in deze studie verschillende oefeningen van regulier woordenschatmateriaal uit groep 6 gebruikt, in tegenstelling tot eerder onderzoek waarin de herhalingsoefeningen identiek waren (bijvoorbeeld, Kornell, 2009; Sobel et al., 2011). In deze eerdere studies was de ecologische validiteit niet zo hoog, omdat in woordenschatlessen juist verschillende soorten oefeningen worden gebruikt (bijvoorbeeld, Blachowicz, Fisher, Ogle & Watts-Taffe, 2006). Ook hebben eerdere onderzoeken laten zien dat variatie in het verwerken van informatie, ook wel *encoding variability* genoemd, het voordeel van spreiden van leren wegneemt (bijvoorbeeld, Dellarosa & Bourne, 1985; Gartman & Johnson, 1972). Het was dus nog niet duidelijk of het gebruiken van verschillende oefeningen invloed zou hebben op het verwachte voordeel van het spreiden van leren. In de eindtoetsen na één week en na vijf weken moesten de kinderen het goede woord invullen bij de gegeven definities. Op beide testen werd een voordeel van het spreiden van leermomenten gevonden.

In Hoofdstuk 3 was de onderzoeksvraag of het ophalen van woorden uit het geheugen helpt om moeilijke woorden en hun synoniemen te leren in een meer gecontroleerde context in groep 5. In deze studie werden twee condities met elkaar vergeleken, namelijk een conditie waarin gestudeerd werd en een conditie waarin getoetst werd. De kinderen moesten tien woordparen hardop lezen (studeren) en zij moesten van tien woordparen het synoniem van de moeilijke woorden proberen op te halen (toetsen). In deze studie was er nog een tweede vraag, namelijk of het introduceren van de woorden in een betekenisvolle context invloed zou hebben op het verwachte voordeel van testen. Om dit te onderzoeken werden de twintig woorden in twee verschillende condities geïntroduceerd. Voor de kinderen in de verhaalconditie werden de woorden geïntroduceerd in een samenhangend verhaal en voor de kinderen in de woordparenconditie werden de woorden geïntroduceerd zonder enige contextuele informatie. We hebben deze twee verschillende introducties toegevoegd, omdat in de meeste woordenschatmethodes de woorden in een betekenisvolle context worden gepresenteerd (bijvoorbeeld, Janssen & Van Ooijen, 2012; Van de Gein, Van de Guchte & Kouwenberg, 2008). Omdat aangetoond is dat het toevoegen van een context helpt bij het leren van woorden (voor een overzichtsartikel, zie

Stahl & Fairbanks, 1986) was het de vraag of er nog steeds een voordeel van oefentoetsen zou zijn, nadat de woorden in een betekenisvolle context werden gepresenteerd. Na een week moesten de kinderen twee toetsen maken, eerst een open vragen toets waarin ze de synoniemen van het woord moesten ophalen en daarna een meerkeuzetoets waarin ze de synoniemen van het woord moesten herkennen uit vier opties. In de meerkeuzetoets werd geen verschil gevonden tussen studeren en toetsen, maar er was wel een voordeel van de woordparenconditie ten opzichte van de verhaalconditie. In de open vragen toets werd een voordeel gevonden van het maken van toetsen ten opzichte van het studeren. Het maken van oefentoetsen helpt dus bij het leren van woordenschat. Ook was er een voordeel van de woordparenconditie ten opzichte van de verhaalconditie. Daarnaast was er een marginaal significante interactie tussen leerconditie en context, het voordeel van toetsen was namelijk iets groter in de woordparenconditie dan in de verhaalconditie. Deze laatste resultaten moeten echter voorzichtig geïnterpreteerd worden, omdat dit verschil maar marginaal significant was en omdat te zien was dat de kinderen in de woordparenconditie al tijdens de oefentoetsen meer woorden ophaalden dan de kinderen in de verhaalconditie. Hoe dan ook, het voordeel van de woordparenconditie kwam niet overeen met onze verwachtingen, maar de resultaten kwamen wel overeen met andere woordenschatstudies waarin geen voordelen van de toevoeging van contextuele informatie werden gevonden (bijvoorbeeld, Jones, Levin, Levin, & Beitzel, 2000; McDaniel & Pressley, 1984, 1989). Mogelijk leidde de contextuele informatie in de verhaalconditie de kinderen af van de betekenis van de woorden. Ook kwam de vorm van de open vragen toets beter overeen met de woordparenconditie dan met de verhalenconditie. Dus mogelijk was er in de woordparenconditie een gelijksoortige verwerking van de woorden tijdens de oefentoetsen en tijdens de eindtoets, ook wel *transfer-appropriate processing* genoemd (bijvoorbeeld, Blaxton, 1989; Morris, Bransford, & Franks, 1977; Thomas & McDaniel, 2007).

In Hoofdstuk 4 was de onderzoeksvraag of er een voordeel van oefentoetsen was bij basisschoolkinderen uit groep 5 die woordenschat leerden in een gesimuleerde klassencontext. Deze studie was meer ecologisch valide dan de studie beschreven in Hoofdstuk 3, omdat het maken van oefentoetsen niet alleen met 'puur studeren' werd vergeleken, maar ook met 'elaboratief studeren'. Het puur bestuderen van rijtjes woorden komt bijna nooit voor in ware woordenschatlessen, dus het enkel vinden van verschillen tussen puur studeren en het maken van oefentoetsen is niet erg informatief voor de onderwijspraktijk. In ware woordenschatlessen worden elaboratieoefeningen gedaan om de woordkennis te verdiepen (bijvoorbeeld, Janssen & Van Ooijen, 2012; Van de Gein et al., 2008). Verder heeft eerder onderzoek aangetoond dat het toevoegen van contextuele informatie helpt bij het leren van woordenschat (zie bijvoorbeeld, Anderson & Reder, 1979; Blachowicz et al., 2006; Bolger, Balass, Landen & Perfetti, 2008; Carey, 1978,

Coomber, Ramstad & Sheets, 1986). Daarom was het de vraag of oefentoetsen iets zouden toevoegen aan woordenschatlessen waarin elaboratieoefeningen voorkwamen. In deze studie werden twee vergelijkbare experimenten gedaan, alleen de eerste leersessie was verschillend. In de eerste leersessie werden de woorden geïntroduceerd met plaatjes en definities (Experiment 1) of met een verhaal (Experiment 2), gevolgd door twee oefeningen waarin gefocust werd op de woordvorm. In de tweede leersessie moesten de kinderen twee consolidatieoefeningen doen, afhankelijk van de leerconditie. In de puur studeren conditie moesten de kinderen een deel van de definitie overschrijven, in de elaboratieconditie moesten de kinderen reguliere oefeningen doen met semantisch gerelateerde woorden en in de oefentoetsenconditie moesten de kinderen het goede woord in de definitie invullen. Na een week maakten de kinderen een invultoets waarin ze het woord moesten ophalen op basis van de gegeven definitie. Ook maakten zij een meerkeuzetoets waarin ze het goede woord in de goede contextzin moesten invullen. Een gecombineerde *small-scale random-effects* meta-analyse over de resultaten van de twee experimenten toonde een voordeel van het maken van oefentoetsen op de invultoets. Er waren kleine verschillen tussen het maken van oefentoetsen en puur studeren en tussen het maken van oefentoetsen en elaboratieoefeningen. Er was geen verschil tussen puur studeren en het maken van elaboratieoefeningen. De gecombineerde betrouwbaarheidsintervallen toonden aan dat er veel onzekerheid was over de grootte van het voordeel van testen en dat een accuratere schatting van de parameter nodig was. Op de meerkeuzetoets waren geen verschillen tussen de condities, waarschijnlijk door plafondeffecten. Het feit dat we geen verschil vonden tussen puur studeren en het maken van elaboratieoefeningen was onverwacht, maar kwam overeen met de resultaten in Hoofdstuk 3 en met de resultaten van andere studies (bijvoorbeeld, Jones et al., 2000; McDaniel & Pressley, 1984, 1989). Mogelijk waren de taken in de elaboratieconditie minder gefocust op de woorddefinitie dan in de puur studeren conditie.

De onderzoeksvraag in Hoofdstuk 5 was of we voordelen van het spreiden van leermomenten en het ophalen uit het geheugen zouden vinden bij het leren van woordenschat in een ware klassencontext in verschillende basisschooljaren (Groepen 4, 5, 6 en 8). De hoofdvragen in dit hoofdstuk waren hetzelfde als de hoofdvragen in Hoofdstukken 2, 3 en 4, maar deze studie was meer ecologisch valide, omdat deze studie werd uitgevoerd in het reguliere woordenschatcurriculum. De ecologische validiteit van de studie werd versterkt door het vergelijken van een *long-lag* conditie met een *short-lag* conditie, in plaats van een 'puur stampen' conditie. In de *short-lag* conditie werd een spreiding tussen de oefenmomenten gebruikt die vergelijkbaar was met de spreiding die wordt gebruikt tussen herhalingen in de reguliere woordenschatlessen. In de *short-lag* conditie werden alle woorden geleerd gedurende twee lessen in dezelfde week, terwijl in de *long-lag* conditie alle woorden werden geleerd gedurende vier lessen in twee weken.

Bovendien oefenden de kinderen de woorden voornamelijk door reguliere oefeningen uit het curriculum te maken. Zoals in veel woordenschatmethoden, werden de woorden dus geoefend op verschillende manieren (zie hiervoor, Janssen & Van Ooijen, 2012; Van de Gein et al., 2008). Naast de reguliere oefeningen startten de kinderen in de helft van de leersessies met een herhaling van de woorden door studeer- en testoefeningen. In de studeeroefeningen moesten zij delen van de beschrijvingen van de woorden overschrijven en in de testoefeningen moesten zij de beschrijvingen van de woorden ophalen. Na deze oefeningen moesten de kinderen de oefeningen uit het reguliere woordenschatcurriculum maken. Aan het eind van elke week moesten de kinderen een open vragen toets maken waarin ze de beschrijvingen van een aantal woorden moesten geven. Een aantal weken na het experiment (één week voor groep 8, twee weken voor groep 5 en elf weken voor de groepen 4 en 6) moesten de kinderen een meerkeuzetoets maken waarin ze de goede beschrijvingen moesten kiezen uit vier keuzemogelijkheden. In tegenstelling tot onze verwachtingen vonden we voordelen van de *short-lag* conditie en van de studeerconditie op de open vragen toets. Het positieve *short-lag* effect was echter alleen significant voor de groepen 4 en 6. Verder waren er geen verschillen tussen de condities op de meerkeuzetoets. Omdat de meerkeuzetoets veel leek op testen die in de klas worden gebruikt, zijn er wat zorgen over de praktische waarde van het spreiden van leermomenten en van het ophalen van informatie uit het geheugen.

Opsommend, vonden we dus een voordeel van het spreiden van leermomenten in een gesimuleerde klassencontext (Hoofdstuk 2), maar niet in een ware klassencontext (Hoofdstuk 5). Ook vonden we een voordeel van het ophalen uit het geheugen door het maken van oefentoetsen in een gecontroleerde context (Hoofdstuk 3) en in een gesimuleerde klassencontext (Hoofdstuk 4), maar niet in een ware klassencontext (Hoofdstuk 5).

Praktische Implicaties

In dit proefschrift zijn de eerste studies uitgevoerd naar twee geheugenstrategieën in het woordenschatonderwijs, namelijk het spreiden van leermomenten en het ophalen uit het geheugen door het maken van oefentoetsen. De studies in dit proefschrift zijn een belangrijke eerste stap en tonen aan dat er zowel positieve als negatieve effecten van *distributed practice* en *retrieval practice* gevonden zijn in het woordenschatonderwijs. Waarschijnlijk zijn er beperkingen in de mate waarin deze geheugenstrategieën voordelig kunnen zijn voor kinderen die nieuwe woordenschat leren.

Ten eerste is in dit proefschrift aangetoond dat het spreiden van leren een positief effect heeft op het leren van nieuwe woorden, als een conditie waarin drie dagen lang één keer wordt geoefend, wordt vergeleken met een conditie waarin drie keer op één dag wordt geoefend (Hoofdstuk 2). Voor de praktijk betekent dit dat het beter is om

de oefeningen meer te spreiden over de week, dan om deze oefeningen allemaal op één dag te doen. In dit proefschrift is echter ook aangetoond dat het spreiden van leren over twee dagen in één week (*short-lag* conditie) in plaats van het spreiden van leren over vier dagen in twee weken (*long-lag* conditie) beter is voor de eindtoetsscore (Hoofdstuk 5). Het is niet helemaal duidelijk waarom het niet altijd beter is om méér te spreiden. Een mogelijke verklaring is dat het retentie-interval relatief kort was en dat er daarom een betere overeenkomst was tussen het retentie-interval en het spreidingsinterval van de *short-lag* conditie, dan tussen het retentie-interval en het spreidingsinterval van de *long-lag* conditie. Dit zou betekenen dat als het langetermijngeheugen van woorden moet worden getest, dat het niet alleen beter is om de woorden pas later te testen, maar dat het ook beter is om het spreidingsinterval juist groter te maken en aan te passen op dit retentie-interval. Een andere mogelijke verklaring is dat de spreiding te lang was in de *long-lag* conditie. Zo is in eerder onderzoek gevonden dat als de spreiding groter wordt, de score op een toets ook hoger wordt, maar dat als de spreiding te groot wordt, de score juist weer lager wordt (bijvoorbeeld, Toppino & Bloom, 2002; Verkoeijen, Rikers, & Schmidt, 2005). Voor de praktijk zou dit betekenen dat het beter is om het oefenen te spreiden over verschillende dagen, dan om alle oefeningen op één dag te doen, maar om ook weer niet te veel te spreiden. Een praktisch voorbeeld is om de woorden niet alleen met een spreiding van één of meer dagen te herhalen, maar daarnaast op die herhalingsdagen de woorden vaker dan één keer te herhalen, zoals werd gedaan in de *short-lag* conditie in Hoofdstuk 5.

Ten tweede is in dit proefschrift aangetoond dat het maken van oefentoetsen een positief effect heeft op woordenschatverwerving, als een conditie waarin woorden worden opgehaald, wordt vergeleken met een conditie waarin de woorden worden opgelezen of overgeschreven (Hoofdstukken 3 en 4) en ook als deze conditie wordt vergeleken met een conditie waarin reguliere woordenschatoefeningen worden gemaakt (Hoofdstuk 4). Dit betekent dus voor de praktijk dat het beter is om leerlingen zelf de definitie van het woord te laten ophalen, dan om deze definitie te laten overschrijven, of om elaboratieoefeningen met het woord te doen. In dit proefschrift is echter ook aangetoond dat het overschrijven beter is dan het ophalen van de beschrijvingen van de woorden, wanneer deze herhalingsoefeningen maar een klein onderdeel uitmaken van alle oefeningen, dus als er ook veel reguliere woordenschatoefeningen worden gedaan (Hoofdstuk 5). Een mogelijke verklaring voor deze resultaten is dat het voordeel van het maken van oefentoetsen afhangt van het aantal oefentoetsen in verhouding tot het aantal reguliere oefeningen. Dit zou echter betekenen dat het maken van oefentoetsen niet altijd zo goed helpt, omdat er in het algemeen veel reguliere woordenschatoefeningen worden gedaan. Voor de praktijk zou dit betekenen dat het wellicht beter is om een aantal van de reguliere oefeningen te vervangen door oefentoetsen. Mogelijk wordt er dan wel een voordeel gevonden van het maken van oefentoetsen.

Ten derde werden de positieve effecten van *distributed practice* en *retrieval practice* niet op alle soorten eindtoetsen gevonden. Bijvoorbeeld in Hoofdstuk 5 werden er geen voordelen gevonden van beide strategieën op een toets waarin de kinderen een beschrijving van een woord moesten geven en in Hoofdstukken 3, 4 en 5 werden er geen voordelen gevonden van beide strategieën op de meerkeuzetoetsen. Het is echter niet zeker of de meerkeuzetoetsen werden beïnvloed door de voorgaande toetsen waarin de kinderen de woorden moesten ophalen. Zo werden er in Hoofdstuk 4 plafondeffecten gevonden op de meerkeuzetoets. Uiteindelijk hebben we alleen maar voordelen gevonden van het spreiden van leermomenten en van het ophalen van informatie uit het geheugen op invultoetsen waarin het nieuwe woord moest worden ingevuld of waarin het synoniem van het woord moest worden opgehaald. Vanuit een praktisch oogpunt is het de vraag of invultoetsen veel worden gebruikt in de klas. In de klas worden met name meerkeuzetoetsen gebruikt die vergelijkbaar zijn met de meerkeuzetoetsen uit Hoofdstuk 4 (bijvoorbeeld, Janssen & Van Ooijen, 2012; Van de Gein et al., 2008). Verder lijken de woordenschattoetsen om algemene woordkennis te meten meer op de meerkeuzetoetsen die in Hoofdstuk 5 werden gebruikt (bijvoorbeeld, Van Berkel & Alberts, 2009; Van Berkel & Hilte, 2009). Er zou kunnen worden gesteld dat het ophalen van woorden en hun synoniemen of definities niet genoeg is voor diepe en betekenisvolle woordenschatverwerking en dat daarom de invultoetsen niet genoeg ecologisch valide zijn, omdat ze enkel het *geheugen* van de woorden meten. Ik denk echter dat in de ontwikkeling van woordenschat het herinneren van woorddefinities een belangrijke eerste stap is voor diepere woordkennis. Zo is er een sterke correlatie tussen de breedte en de diepte van de woordenschatkennis (Vermeer, 2001). Daarom verwacht ik dat, ondanks dat in dit proefschrift alleen effecten op invultoetsen zijn gevonden, vergelijkbare resultaten gevonden zullen worden op taken die diepere woordenschatkennis vereisen.

Ten vierde, in tegenstelling tot de literatuur waarin wordt verondersteld dat elaboratieoefeningen de woordenschatverwerving versterken (bijvoorbeeld, Blachowicz et al., 2006; Blachowicz & Fisher, 2000) wekken de resultaten van dit proefschrift de suggestie dat elaboratieoefeningen de woordenschat niet vergroten. Zo was er in Hoofdstuk 3 een voordeel van de woordparenconditie ten opzichte van de verhaalconditie en in Hoofdstuk 4 waren er geen verschillen tussen de puur studeren conditie en de elaboratieconditie. De resultaten van de studies in Hoofdstukken 3 en 4 kwamen overeen met andere studies waarin geen voordeel van elaboratie werd gevonden in woordenschatverwerving (bijvoorbeeld, Graves, 2006; Jones et al., 2000; McDaniel & Pressley, 1984, 1989; Nation, 2001). We kunnen ons daarom afvragen of elaboratieve verwerking voordelig is in elke leersituatie en of letterlijke herhaling van de woorden en hun definities wellicht ook al zou kunnen leiden tot een voldoende herinnering van synoniemen of woorddefinities.

Conclusie

De studies in dit proefschrift zijn de eerste studies die gedaan zijn naar het *distributed practice effect* en het *retrieval practice effect* in het woordenschatonderwijs. In een aantal van deze studies zijn voordelen gevonden van het spreiden van leermomenten en van het ophalen uit het geheugen. Deze studies tonen aan dat deze geheugenstrategieën gebruikt kunnen worden om effectief nieuwe woorden aan te leren. In één studie zijn echter geen voordelen van het spreiden van leermomenten en van het ophalen uit het geheugen gevonden. Deze studies samen tonen dus aan dat er grenzen zijn aan de mate waarin het spreiden van leren en het maken van oefentoetsen leervoordelen kunnen geven in woordenschatverwerving. Al met al, door het uitvoeren van deze studies is een eerste belangrijke stap gezet om geheugenstrategieën, die tot robuuste effecten leiden in een laboratoriumsetting, te verbinden met de educatieve praktijk van woordenschatverwerving. De studies in dit proefschrift geven ons inzicht in de mate waarin het spreiden van leermomenten en het ophalen uit het geheugen voordelig kunnen zijn bij het leren van nieuwe woorden.

Alekseev, P. M. (1984). *Statistische Lexikographie.* Bochum: Brockmeyer.

Anderson, R.C., & Freebody, P. (1981). Vocabulary knowledge. In J. T. Guthrie (ed.), *Comprehension and teaching: Research reviews* (pp. 71-117). Newark, Del.: International Reading Association.

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.

Bahrick, H.P., Bahrick, L.E., Bahrick, A.S., & Bahrick, P.E. (1993). Maintenance of foreign language vocabulary and the spacing effect. *Psychological Science, 4*, 316-321. doi: 10.1111/j.1467-9280.1993.tb00571.x.

Baker, S. K., Simmons, D. C., & Kame'enui, E. J. (1998). Vocabulary acquisition: Research bases. In D. C. Simmons, & E. J. Kame'enui (Eds.). *What reading research tells us about children with diverse learning needs* (pp. 183-218). Mahwah, NJ: Erlbaum.

Beck, I. L., McKeown, M. G., & Kucan, L. (2002). *Bringing words to life: Robust vocabulary instruction*. New York: Guilford Press.

Biemiller, A. (2005). Size and sequence in vocabulary development: Implications for choosings words for primary grade vocabulary instruction. In A. Hiebert & M. Kamil (eds.), *Teaching and learning vocabulary: Bringing research to practice* (pp. 223-242). Mahwah, NJ: Erlbaum.

Biemiller, A., & Boote, C. (2006). An effective method for building vocabulary in primary grades. *Journal of Educational Psychology, 98,* 44-62. doi: 10.1037/0022-0663.98.1.44.

Blachowicz, C. L. Z., & Fisher, P. J. L. (2000). Vocabulary instruction. In R. Barr, M. L. Kamil, P. B. Mosenthal, & P. D. Pearson (Eds)., *Handbook of reading research* (Vol. 3) (pp. 503-523). Mahwah, NJ: Erlbaum.

Blachowicz, C. L.Z., Fisher, P. J. L., Ogle, D., & Watts-Taffe, S. (2006). Vocabulary: Questions from the classroom. *Reading Research Quarterly, 41*, 524-530. doi: 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, and Cognition, 15,* 657-668. doi:10.1037/0278-7393.15.4.657.

Bloom, K.C., & Shuell, T.J. (1981). Effects of massed and distributed practice on the learning and retention of second-language vocabulary. *Journal of Educational Research, 74*, 245-248.

Blunt, J. R., & Karpicke, J. D. (2014). Learning with retrieval-based concept mapping. *Journal of Educational Psychology*, *106*, 58-68. doi: 10.1037/a0035934.

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. doi:10.1080/01638530701792826.

Bouwmeester, S. & Verkoeijen, P. P. J. L. (2011a). 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. doi:10.1016/j.jml.2011.02.005.

Bouwmeester, S. & Verkoeijen, P. P. J. L. (2011b). 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. doi:10.1007/s10212-010-0036-3.

Budé, L., Imbos, T., Van de Wiel, M. W., & Berger, M. P. (2011). The effect of distributed practice on students' conceptual understanding of statistics. *Higher Education, 62,* 69- 79. doi: 10.1007/s10734-010-9366-y.

Butler, A. C., Karpicke, J. D., & Roediger, H. L. (2008). Correcting a metacognitive error: Feedback increases retention of low-confidence correct responses. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*, 918-928. doi: 10.1037/0278-7393.34.4.918.

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. doi:10.1080/09541440701326097.

Cahill, A., & Toppino, T. C. (1993). Young children's recognition as a function of the spacing of repetitions and the type of study and test stimuli. *Bulletin of the Psychonomic Society, 31*, 481-484. doi: 10.3758/BF03334969.

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. (2009). Cue strength as a moderator of the testing effect: The benefits of elaborative retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35,* 1563-1569. doi: 10.1037/a0017021.

Carpenter, S. K. (2011). Semantic information activated during retrieval contributes to later retention: Support for the mediator effectiveness hypothesis of the testing effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*, 1547-1552. doi:10.1037/a0024140.

Carpenter, S. K., & Pashler, H. (2007). Testing beyond words: Using tests to enhance visuospatial map learning. *Psychonomic Bulletin & Review, 14*, 474-478. doi: 10.3758/BF03194092.

Carpenter, S. K., Pashler, H., & Cepeda, N. J. (2009). Using tests to enhance 8th grade students' retention of U.S. history facts. *Applied Cognitive Psychology, 23,* 760-771. doi: 10.1002/acp.1507.

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. doi: 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. doi: 10.3758/MC.36.2.438.

Carrier, M., & Pashler, H. (1992). The influence of retrieval on retention. *Memory & Cognition, 20*, 633-642. doi: 10.3758/BF03202713.

Cepeda, N. J., Pashler, H., Vul, E., Wixted, J.T., & Rohrer, D. (2006). Distributed practice in verbal recall tasks: A review and quantitative synthesis. *Psychological Bulletin*, *132*, 354-380. doi: 10.1037/0033-2909.132.3.354.

Cepeda, N. J., Vul, E., Rohrer, D., Wixted, J. T., & Pashler, H. (2008). Spacing effects in learning. A temporal ridgeline of optimal retention. *Psychological Science, 19*, 1095-1102. doi: 10.1111/j.1467-9280.2008.02209.x.

Cofer, C. N. , Bruce, D. R. & Reicher, G. M. (1966). Clustering in free recall as a function of certain methodological variations. *Journal of Experimental Psychology, 71,* 858-866. doi: 10.1037/h0023217.

Cohen, A. D., & Aphek, E. (1981). Easifying second language learning. *Studies in Second Language Acquisition, 3*, 221-236. doi: 10.1017/S0272263100004198.

Cook, L. K. & Mayer, R. E. (1983). Reading strategies training for meaningful learning from prose. In M. Pressley and J. Levin (Eds.). *Cognitive Strategy Research* (pp. 87-131). New York: Springer Verlag. doi: 10.1007/978-1-4612-5519-2_4.

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.

Coppens, L. C., Verkoeijen, P. P. J. L., & Rikers, R. M. J. P. (2011). Learning Adinkra symbols: The effect of testing. *Journal of Cognitive Psychology, 23,* 351-357. doi: 10.1080/20445911.2011.507188.

Craik, F. I. M. & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General, 104*, 268-294. doi: 10.1037/0096-3445.104.3.268.

Cranney, J., Ahn, M., McKinnon, R., Morris, S., & Watts, K. (2009). The testing effect, collaborative learning, and retrieval-induced facilitation in a classroom setting. *European Journal of Cognitive Psychology, 21*, 919-940. doi: 10.1080/09541440802413505.

Cull, W. L. (2000). Untangling the benefits of multiple study opportunities and repeated testing for cued recall. *Applied Cognitive Psychology, 14,* 215-235. doi: 10.1002/(SICI)1099-0720(200005/06)14:3<215::AID-ACP640>3.0.CO;2-1.

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. doi: 10.1037/1076-898X.2.4.365.

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. doi: 10.1177/0956797613504966.

Deese, J. (1959). On the prediction of occurrence of particular verbal intrusions in immediate recall. *Journal of Experimental Psychology, 58*, 17-22. doi: 10.1037/h0046671.

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. doi: 10.1027/1618-3169/a000189.

Delaney, P.F., Verkoeijen, P.P.J.L., & Spirgel, 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. doi: 10.1016/S0079-7421(10)53003-2.

Dellarosa, D., & Bourne, L. E. (1985). Surface form and the spacing effect. *Memory & Cognition, 13,* 529-537. doi: 10.3758/BF03198324.

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. doi: 10.1177/1529100612453266.

Fritz, C.O., Morris, P.E., Acton, M., Voelkel, A.R., Etkind, R. (2007). Comparing and combining retrieval practice and the keyword mnemonic for foreign vocabulary learning. *Applied Cognitive Psychology, 21,* 499-526. doi: 10.1002/acp.1287.

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. doi: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.

Gartman, L. M. , & Johnson, N. F. (1972). Massed versus distributed repetition of homographs: A test of the differential-encoding hypothesis. *Journal of Verbal Learning and Verbal Behavior*, *11*, 801-808. doi: 10.1016/S0022-5371(72)80016-1.

Gipe, J. P. (1979). Investigating techniques for teaching word meanings. *Reading Research Quarterly, 14*, 624–644. doi: 10.2307/747265.

Gipe, J. P. (1980). Use of a Relevant Context Helps Kids Learn New Word Meanings. *The Reading Teacher, 33*, 398-402.

Goossens, N.A.M.C., Camp, G., Verkoeijen, P.P.J.L., & Tabbers, H. K. (2014). The effect of retrieval practice in primary school vocabulary learning. *Applied Cognitive Psychology, 28*, 135-142. doi: 10.1002/acp.2956.

Goossens, N.A.M.C., Camp, G., Verkoeijen, P.P.J.L., Tabbers, H.K., & Zwaan, R.A. (2012). Spreading the words: A spacing effect in vocabulary learning. *Journal of Cognitive Psychology, 24*, 965-971. doi: 10.1080/20445911.2012.722617.

Goossens, N.A.M.C., Camp, G., Verkoeijen, P.P.J.L., Tabbers, H. K., & Zwaan, R.A. (2014). The benefit of retrieval practice over elaborative restudy in primary school vocabulary learning. *Journal of Applied Research in Memory and Cognition*, *3,* 177-182. doi: 10.1016/j.jarmac.2014.05.003.

Goossens, N. & Vermeer, A. (2009). Wat is een optimale tekstdekking? Woordkennis en tekstbegrip in groep 6. *Toegepaste Taalwetenschap in Artikelen*, *82*, 81-92. doi: 10.1075/ttwia.82.08goo.

Graves, M. F. (2006). *The vocabulary book. Learning and instruction*. New York: Teachers College Press.

Hart, B., & Risley, T. (1995). *Meaningful differences in the everyday lives of young American children*. Baltimore: Paul H. Brookes.

Hazenberg, S., & Hulstijn, J. H. (1992). Woorden op zicht: Woordselectie ten behoeve van het NT2-onderwijs. *Levende Talen*, *467*, 2-7.

Hirsch, E. D. (2003). Reading comprehension requires knowledge – of words and the world: Scientific insights into the fourth-grade slump and the nation's stagnant comprehension scores. *American Educator, 27 (1)*, 10-29.

Hirsch, D., & Nation, P. (1992). What vocabulary size is needed to read unsimplified texts for pleasure? *Reading in a Foreign Language, 8*, 689-696.

Hogan, R. M., & Kintsch, W. (1971). Differential effects of study and test trials on long-term recognition and recall. *Journal of Verbal Learning and Verbal Behavior, 10,* 562–567. doi: 10.1016/S0022-5371(71)80029-4.

Jacoby, L.L. (1978). On interpreting the effects of repetition: Solving a problem versus remembering a solution. *Journal of Verbal Learning and Verbal Behavior, 17*, 649-667. doi: 10.1016/S0022-5371(78)90393-6.

Janssen, M. & Van Ooijen, M. (eds.) (2012). *Taal Actief, 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. doi:10.1037/0022-0663.92.2.256.

Karpicke, J. D. (2009). Metacognitive control and strategy selection: Deciding to practice retrieval during learning. *Journal of Experimental Psychology: General, 138,* 469-486. doi: 10.1037/a0017341.

Karpicke, J. D., Lehman, M., & Aue, W. R. (2014). Retrieval-based learning: An episodic context account. In B. H. Ross (Ed.), *Psychology of Learning and Motivation* (Vol. 61) (pp. 237-284). San Diego, CA: Elsevier Academic Press. doi: 10.1016/B978-0-12-800283-4.00007-1.

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. doi: 10.1037/0278-7393.33.4.704.

Karpicke, J. D., & Roediger, H. L. (2008). The critical importance of retrieval for learning. *Science, 319*, 966-968. doi: 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. doi: 10.1016/j.jml.2012.02.004.

Kornell, N. (2009). Optimizing learning using flash-cards: Spacing is more effective than cramming. *Applied Cognitive Psychology*, *23*, 1297-1317. doi: 10.1002/acp.1537.

Kuiken, F. & Vermeer, A. (Eds.) (2005). *Nederlands als tweede taal in het basisonderwijs*. Utrecht: Thieme/ Meulenhoff.

Küpper-Tetzel, C. E., Kapler, I. V., & Wiseheart, M. (2014). Contracting, equal, and expanding learning schedules: The optimal distribution of learning sessions depends on retention interval. *Memory & Cognition, 42,* 729-741. doi: 10.3758/s13421-014-0394-1.

Lehman, M., Smith, M. A., & Karpicke, J. D. (2014). Toward an episodic context account of retrieval-based learning: Dissociating retrieval practice and elaboration. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*, 1787-1794. doi: 10.1037/xlm0000012.

Marsh, E. J., Fazio, L. K., & Goswick, A. E. (2012). Memorial consequences of testing school-aged children. *Memory, 20,* 899-906. doi: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. doi: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. doi: 10.1080/09541440701326154.

McDaniel, M. A., Howard, D. C., & Einstein, G. O. (2009). The read-recite-review study strategy: Effective and portable. *Psychological Science, 20*, 516-522. doi: 10.1111/j.1467-9280.2009.02325.x.

McDaniel, M. A., & Pressley, M. (1984). Putting the keyword method in context. *Journal of Educational Psychology, 76*, 598-609. doi: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. doi:10.1037/0022-0663.81.2.204.

McDaniel, M. A., Wildman, K. M., & Anderson, J. L. (2012). Using quizzes to enhance summative-assessment performance in a web-based class: An experimental study. *Journal of Applied Research in Memory and Cognition, 1,* 18-26. doi: 10.1016/j.jarmac.2011.10.001.

McKeown, M. G. (1985). The acquisition of word meaning from context by children of high and low ability. *Reading Research Quarterly, 20,* 482-496. doi: 10.2307/747855.

Melton, A. W. (1967). Repetition and retrieval from memory. *Science, 158,* 532. doi: 10.1126/science.158.3800.532-b.

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. doi:10.1080/09541440701326063.

Modigliani, V. (1976). Effects on a later recall by delaying initial recall. *Journal of Experimental Psychology: Human Learning and Memory, 2*, 609–622. doi: 10.1037/0278-7393.2.5.609.

Mondria, J.-A. (1996). *Vocabulaireverwerving in het vreemde-talenonderwijs: de effecten van context en raden op de retentie* (Doctoral dissertation). Groningen: Universiteitsdrukkerij Groningen.

Morris, C. D., Bransford, J. D., & Franks, J. J. (1977). Levels of processing versus transfer appropriate processing. *Journal of Verbal Learning and Verbal Behavior, 16*, 519-533. doi:10.1016/S0022-5371(77)80016-9.

Nagy, W. & Herman, P. (1987). Breadth and depth of vocabulary knowledge: Implications for acquisition and instruction. In M. McKeown & M. Curtis (Eds.), *The nature of vocabulary acquisition* (pp. 19-36). Hillsdale, NJ: LEA.

Nation, I. S. P. (1982). Beginning to learn foreign vocabulary: A review of the research. *RELC Journal, 13 (1),* 14-36. doi: 10.1177/003368282013001O2.

Nation, I. S. P. (1990). *Teaching and learning vocabulary*. New York: Newbury House.

Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.

Nation, I. S. P. (2004). Vocabulary learning and intensive reading. *EA Journal*, *21 (2)*, 20–29.

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. doi: 10.1037/0278-7393.31.1.3.

Pressley, M., Levin, J. R.., & Miller, G. E. (1982). The keyword method compared to alternative vocabulary-learning strategies. *Contemporary Educational Psychology, 7*, 50-60. doi: 10.1016/0361-476X(82)90007-8.

Pyc, M. A., & Rawson, K. A. (2007). Examining the efficiency of schedules of distributed retrieval practice. *Memory & Cognition, 35*, 1917-1927. doi: 10.3758/BF03192925.

Pyc, M. A., & Rawson, K. A. (2009). Testing the retrieval effort hypothesis: Does greater difficulty correctly recalling information lead to higher levels of memory? *Journal of Memory and Language, 60*, 437-447. doi: 10.1016/j.jml.2009.01.004.

Pyc, M. A., & Rawson, K. A. (2011). Costs and benefits of dropout schedules of test-restudy practice: Implications for student learning. *Applied Cognitive Psychology, 25*, 87-95. doi: 10.1002/acp.1646.

Rawson, K. A., & Dunlosky, J. (2011). Optimizing schedules of retrieval practice for durable and efficient learning: How much is enough? *Journal of Experimental Psychology: General, 140*, 283-302. doi: 10.1037/a0023956.

Rawson, K. A., & Kintsch, W. (2005). Rereading effects depend on time of test. *Journal of Educational Psychology, 97*, 70-80. doi: 10.1037/0022-0663.97.1.70.

Rea, C.P., & Modigliani, V. (1987). The spacing effect in 4- to 9-year-old children. *Memory & Cognition, 15*, 436-443. doi: 10.3758/BF03197733.

Rodríguez, M. & Sadoski, M. (2000). Effects of rote, context, keyword, and context/keyword methods on retention of vocabulary in EFL classrooms. *Language Learning, 50,* 385- 412. doi: 10.1111/0023-8333.00121.

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. doi:10.1037/a0026252.

Roediger, H. L. & Butler, A. C. (2011). The critical role of retrieval practice in long-term retention. *Trends in Cognitive Sciences, 15*, 20-27. doi: 10.1016/j.tics.2010.09.003.

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. doi: 10.1111/j.1745-6916.2006.00012.x.

Roediger, H. L., & McDermott, K. B. (1995). Creating false memories: Remembering words not presented on lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*, 803-814. doi: 10.1037/0278-7393.21.4.803.

Roediger, H. L., Putnam, A. L., & Smith, M. A. (2011). Ten benefits of testing and their applications to educational practice. In J. Mestre & B. Ross (Eds.), *Psychology of Learning and Motivation: Cognition in education* (pp. 1-36). Oxford : Elsevier. doi: 10.1016/B978-0-12-387691-1.00001-6.

Rohrer, D., Taylor, K., & Sholar, B. (2010). Tests enhance the transfer of learning. *Journal of Experimental Psychology: Learning, Memory & Cognition, 36*, 233-239. doi: 10.1037/a0017678.

Schmitt, N. (1997). Vocabulary learning strategies. In Schmitt, N., & McCarthy, M. (Eds.). *Vocabulary. Description, Acquisition and Pedagogy* (pp. 199-227). Cambridge: Cambridge University Press.

Schrooten, W., & Vermeer, A. (1994). *Woorden in het basisonderwijs. 15.000 woorden aangeboden aan leerlingen.* Tilburg: Tilburg University Press.

Scott, J., Jamieson-Noel, D. & Asselin, M. (2003). Vocabulary instruction throughout the school day in twenty-three Canadian upper-elementary classrooms. *The Elementary School Journal, 103*, 269-286.

Seabrook, R., Brown, G. D. A., & Solity, J. E. (2005). Distributed and massed practice: From laboratory to classroom. *Applied Cognitive Psychology, 19*, 107-122. doi: 10.1002/acp.1066.

Smith, S. M., & Rothkopf, E. Z. (1984). Contextual enrichment and distribution of practice in the classroom. *Cognition and Instruction, 1,* 341 - 358. doi: 10.1207/s1532690xci0103_4.

Sobel, H. S., Cepeda, N. J., & Kapler, I. V. (2011). Spacing effects in real-world classroom vocabulary learning. *Applied Cognitive Psychology*, *25*, 763-767. doi: 10.1002/acp.1747.

Spitzer, H.F. (1939). Studies in retention. *Journal of Educational Psychology, 30*, 641–656 doi: 10.1037/h0063404.

Stahl, S. A., & Fairbanks, M. M. (1986). The effects of vocabulary instruction: A model-based meta-analysis. *Review of Educational Research, 56*, 72-110. doi: 10.3102/00346543056001072.

Stahl, S. A., & Nagy, W. E. (2006). *Teaching word meanings.* Mahwah, NJ: Lawrence Erlbaum.

Stanovich, K. E. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly, 21*, 360-407.

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. doi:10.3758/BF03193305.

Thompson, I. (1987). Memory in language learning. In A. Wenden and J. Rubin (Eds.). *Learner strategies in language learning* (pp. 43-56). New York: Prentice Hall.

Toppino, T.C. (1993). The spacing effect in preschool children's free recall of pictures and words. *Bulletin of the Psychonomic Society, 31*, 27-30. doi: 10.3758/BF03334130.

Toppino, T. C. & Bloom, L. C. (2002). The spacing effect, free recall, and two-process theory: a closer look. *Journal of Experimental Psychology, 28*, 437-444. doi: 10.1037//0278-7393.28.3.437.

Toppino, T. C., & Cohen, M. S. (2009). The testing effect and the retention interval: Questions and answers. *Experimental Psychology*, *56,* 252–257. doi: 10.1027/1618-3169.56.4.252.

Toppino, T.C., & DeMesquita, M. (1984). Effects of spacing repetitions on children's memory. *Journal of Experimental Child Psychology*, *37*, 637-648. doi: 10.1016/0022-0965(84)90081-X.

Toppino, T.C., & DiGeorge, W. (1984). The spacing effect in free recall emerges with development. *Memory & Cognition, 12*, 118-122. doi: 10.3758/BF03198425.

Toppino, T. C., Kasserman, J. E., & Mracek, W. A. (1991). The effect of spacing repetitions on the recognition memory of young children and adults. *Journal of Experimental Child Psychology, 51,* 123-138. doi: 10.1016/0022-0965(91)90079-8.

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. doi: 10.1016/S0022- 5371(67)80092-6.

Van Berkel, S. & Alberts, N. (2009). *LOVS Woordenschat groep 4*. Arnhem: Cito.

Van Berkel, S. & Hilte, M. (2009). *LOVS Woordenschat groep 5*. Arnhem: Cito.

Van de Gein, J., Van de Guchte, C., & Kouwenberg, B. (2008). *Zin in Taal Nieuw A, B, C, D, E.* Tilburg: Zwijsen.

Verhallen, M. (2009). *Meer en beter woorden leren. Een brochure over de relatie tussen taalvaardigheid en schoolsucces.* Utrecht: Projectbureau Kwaliteit PO-Raad.

Verhallen, M., & Verhallen, S. (1994). *Woorden leren, woorden onderwijzen.* Hoevelaken: CPS.

Verhoeven, L. & Vermeer, A. (1993). *Taaltoets Allochtone Kinderen Bovenbouw.* Tilburg: Zwijsen.

Verhoeven L. & Vermeer, A. (1996). *Taalvaardigheid in de bovenbouw: Nederlands van autochtone en allochtone leerlingen in het basis- en mlk- onderwijs*. Tilburg: Tilburg University Press.

Verkoeijen, P. P. J. L., Rikers, R. M. J. P., & Schmidt, H. G. (2005). Limitations to the spacing effect: Demonstration of an inverted U-shaped relationship between interrepetition spacing and free recall. *Experimental Psychology, 52*, 257–263. doi: 10.1027/1618-3169.52.4.257.

Verkoeijen, P.P.J.L., Tabbers, H.K., & Verhage, M.L. (2011). Comparing the effects of testing and restudying on recollection in recognition memory. *Experimental Psychology, 58*, 490-498. doi: 10.1027/1618-3169/a000117.

Vermeer, A. (2000). Lexicale rijkdom, tekstmoeilijkheid en woordenschatgrootte. Beschrijving van de MLR, een woordenschat-analyseprogramma. *Toegepaste Taalwetenschap in Artikelen, 64*, 95-105. doi: 10.1075/ttwia.64.10ver.

Vermeer, A. (2001). Breadth and depth of vocabulary in relation to L1/L2 acquisition and frequency of input. *Applied Psycholinguistics, 22*, 217-234.

Webb, L. W. (1921). A comparison of two methods of studying with application to foreign language. *The School Review: A Journal of Secondary Education, 29,* 58-67.

Wenger, S. K., Thompson, C. P., & Bartling, C. A. (1980). Recall facilitates subsequent recognition. *Journal of Experimental Psychology: Human Learning and Memory, 6*, 135–144. doi: 10.1037/0278-7393.6.2.135.

Wheeler, M. A., Ewers, M., & Buonanno, J. F. (2003). Different rates of forgetting following study versus test trials. *Memory, 11,* 571–580. doi: 10.1080/09658210244000414.

Dit proefschrift was er niet geweest zonder vele personen om mij heen. In dit dankwoord sta ik graag stil bij de personen die mij hebben geholpen bij het schrijven van dit proefschrift.

Allereerst wil ik graag Gino, Huib, Peter en Rolf bedanken voor de betrokkenheid en het enthousiasme waarmee zij mij hebben begeleid. Gino, dankjewel voor het steeds maar weer spelen van de advocaat van de duivel als ik een onderzoeksplan had dat meer theoretisch fundament behoefde, of als ik mijn redeneerlijnen nog niet helemaal had uitgedacht of uitgeschreven. Je gaf hierbij de pijnpunten feilloos aan, maar gaf me ook de vrijheid om eigen initiatief te nemen. Bedankt voor de goede tips die je me gaf voor de begeleiding van studenten. Peter en Huib, dankjewel voor alle extra hulp, dit was echt een verrijking. Bedankt voor de vlugge feedback op manuscripten en voor alle keren dat jullie deuren open stonden voor mij, als ik binnenkwam voor een statistische of praktische vraag! Rolf, dankjewel dat jij zo goed de grote lijnen van het onderzoeksproces in de gaten hield. Ook waardeer ik het enorm dat jij je promovendi zo stimuleert om kritisch te kijken naar de wetenschap.

Graag wil ik de Stichting Bestuur Openbaar Onderwijs Rotterdam bedanken voor de mogelijkheid die ze mij hebben geboden om het onderzoek uit te voeren. Paul Kamps en Linda van Tuyl wil ik bedanken voor hun betrokkenheid bij de ontwikkeling van de Toolbox. Linda, jouw doortastendheid en vermogen om mensen met elkaar te verbinden hebben mij geïnspireerd. Graag wil ik de directeuren, intern begeleiders en de leerkrachten van OBS De Wilgenstam, OBS De Plevier, OBS Jan Antonie Bijloo, OBS De Driehoek, OBS Nelson Mandela, OBS De Klimop en OBS De Schalm hartelijk bedanken voor hun gastvrijheid. Ook de kinderen en hun ouders wil ik bedanken voor hun deelname aan het onderzoek.

I would like to thank the members of the dissertation committee, prof. dr. Liesbeth Kester, dr. Jeffrey Karpicke, dr. Diane Pecher, prof. dr. Rob Martens en dr. René Zeelenberg. Thank you for taking the time and effort to read my dissertation. I am looking forward to discuss the content during the public defense.

Zonder de C&L-groep was dit project er niet geweest, dus graag wil ik Anique, Gino, Huib, Peter, Samantha en Jan bedanken voor de prettige samenwerking bij de ontwikkeling van de Toolbox. Gabriela, Gerdien, Mario, Noortje en Stijn, dankjewel voor jullie extra hulp hierbij. Heel de C&L-groep wil ik bedanken voor de interessante discussies en de gezellige etentjes! Samantha, dankjewel voor je enthousiasme, je statistische hulp en je inspiratie voor praktijkgericht onderzoek. Het was leerzaam om met jou samen te werken in het onderzoek naar diepte van woordkennis en in het onderzoek uit hoofdstuk 5 van dit

proefschrift. Jan, dankjewel voor de prettige samenwerking bij onze BOOR-activiteiten. Het was fijn dat jij altijd bereid was om nog eens met een kritische blik naar een tekst of presentatie te kijken.

De 'lab'-collega's, Anita, Jacqueline, Jan, Karen, Lisa, Lysanne, Rolf en Wim, wil ik graag bedanken voor onze wekelijkse bijeenkomsten en voor de schrijfweek. Ik heb veel geleerd van onze gezamenlijke projecten en van onze discussies over de wetenschap. Daarnaast wil ik de collega's van B&C en O&O graag bedanken voor de belangstelling tijdens presentaties. De O&O-pubgroup wil ik bedanken voor de aangename sfeer waarin we elkaars manuscripten bespraken. Graag wil ik dan ook Bonnie, Diane, Katinka, Kim, Lisette, Lydia, Martine, René, Sofie, Tamara en Vincent bedanken voor hun constructieve feedback.

Kamergenoten Benjamin en Lysanne wil ik bedanken voor de fijne tijd in kamer T13-15. Benjamin, bedankt voor het geven van de praktische tips en de minicolleges statistiek. Lysanne, dankjewel voor de prettige gesprekken, de peptalks, de vreugdedansjes, de dropjes en de theetjes! Jan, Kimberley en Martine, 'de club van 1 december', het was zeer vermakelijk om onze ervaringen te delen tijdens onze etentjes, om carnaval te vieren en om aardbeien te plukken. Lysanne en Martine, ik vind het top dat jullie mijn paranimfen zijn! Birgit, dankjewel voor de bemoedigende gesprekken die we hebben gevoerd. Ook wil ik graag de promovendi van onderzoeksschool ICO bedanken, in het bijzonder Mariëtte en Kim. Kim, dankjewel voor je input, je enthousiasme en je gastvrijheid.

In een promotieproject is hulp bij concrete zaken belangrijk. Graag wil ik Anne Vermeer van de Universiteit van Tilburg bedanken voor zijn praktische woordenschattips, maar bovenal voor het overbrengen van zijn enthousiasme over woordenschatonderzoek, waardoor ik op zoek ben gegaan naar een promotieplaats. Ik dank de studenten Pim, Priscillia, Rinske, Nathan, Quinlan, Tehilah, Luca, Sonja en Elise voor hun bijdragen aan het onderzoek. De medewerkers van het EBL wil ik bedanken voor hun hulp bij het programmeren van experimenten. Speciale dank gaat uit naar de secretaresses Mirella, Iris en Angelique.

Linda, dankjewel voor het ontwerpen van de prachtige kaft van dit proefschrift! Ik vond het heel verfrissend om te zien hoe jij de talige informatie over het onderzoek grafisch hebt verwerkt. Eveline, Marloes en Josien, dankjewel voor jullie betrokkenheid! Anne, Eefje, Mariët en Wendy, wat is het fijn om al zo lang met jullie lief en leed te delen. Dankjewel voor alle feestjes, etentjes en weekendjes die voor ontspanning zorgden!

Ten slotte wil ik mijn familie bedanken. Miranda en Stefan, Patricia en Teun, Janny en Hubert, dankjewel voor alle fijne momenten. Miranda en Patricia, jullie blijven mijn grote zussen waar ik naar opkijk. Dankjewel 'dat ik met jullie kan lachen en huilen', om Patricia maar te citeren. Fabienne, Joy, Sven, Lars, Kim, Jort en Sven, jullie zorgen voor veel plezier. Zo moet ik nog steeds lachen om de manier waarop jullie mij uitlegden wat een 'ovaal' of een 'voorlader' was. Toon en Riet, bedankt dat jullie altijd voor Johan en mij klaarstaan. Betere buren kunnen wij ons niet wensen! Papa en mama, dankjewel voor jullie luisterend oor. Jullie hebben mij in alle stappen van mijn leven gestimuleerd en gemotiveerd. Ik vind het fijn dat ik altijd bij jullie terecht kan. Als laatste, lieve Johan, dankjewel voor je onvoorwaardelijke liefde, steun en begrip. Jij hebt me geleerd om beter te relativeren, om beter te ontspannen en je laat me bovenal heel veel lachen. We zijn een topteam en daarom ben ik blij dat jij aan mijn zijde hebt gestaan tijdens dit promotieproject! Ik hoop dat we ons hele leven nog aan elkaars zijde mogen staan, bij welk project of bij welke gebeurtenis dan ook.

Curriculum Vitae

Nicole Goossens was born in Breda, The Netherlands, on January 14th, 1987. She completed her secondary education in 2005 at the Katholieke Scholengemeenschap Etten-Leur. She entered the Communication and Information Sciences program at Tilburg University in 2005. In 2007, she spent four months abroad as an exchange student at Karlstad University in Sweden. She obtained her Bachelor's and Master's degree in Communication Design in 2008 and 2009 respectively. For her Bachelor's thesis, she investigated the effectiveness of a primary school vocabulary learning computer program, and for her Master's thesis she studied the role of primary school vocabulary knowledge in text comprehension.

In December 2009, Nicole started as a PhD student at Erasmus University Rotterdam, in a project co-funded by the Board of Public Education in Rotterdam (Stichting BOOR) and the Institute of Psychology of Erasmus University Rotterdam. In this PhD project she investigated the use of distributed practice and retrieval practice in primary school vocabulary learning, which resulted in the present dissertation. In addition to performing research, she supervised several Bachelor and Master theses and she was involved in teaching courses on psychology, statistics, and writing skills. Nicole is also a co-author of the 'Toolbox', a practical guide for teachers that presents ten useful research-based guidelines for classroom instruction. She has given several presentations and workshops about this Toolbox to teachers in primary schools across Rotterdam.

In April 2014, Nicole started working as a post-doc researcher at the Open University of The Netherlands. Currently, she does further research into the use of distributed practice and retrieval practice in primary school vocabulary learning. She does this research in collaboration with teachers of primary schools in Rotterdam, and in co-operation with the Board of Public Education in Rotterdam (Stichting BOOR).

Peer-Reviewed Publications

Goossens, N.A.M.C., Camp, G., Verkoeijen, P.P.J.L., Tabbers, H. K., & Zwaan, R.A. (2014). The benefit of retrieval practice over elaborative restudy in primary school vocabulary learning. *Journal of Applied Research in Memory and Cognition*, *3*, 177-182. doi: 10.1016/j.jarmac.2014.05.003.

Goossens, N.A.M.C., Camp, G., Verkoeijen, P.P.J.L., & Tabbers, H.K. (2014). The effect of retrieval practice in primary school vocabulary learning. *Applied Cognitive Psychology*, *28*, 135-142. doi: 10.1002/acp.2956.

Goossens, N.A.M.C., Camp, G., Verkoeijen, P.P.J.L., Tabbers, H.K., & Zwaan, R.A. (2012). Spreading the words: A spacing effect in vocabulary learning. *Journal of Cognitive Psychology*, *24*, 965-971. doi: 10.1080/20445911.2012.722617.

Dutch Publications

Bouwmeester, S., de Bruin, A.B.H., Camp, G., Engelen, J.A.A., Goossens, N.A.M.C., Tabbers, H.K., & Verkoeijen, P.P.J.L. (2012). *Toolbox: 10 oefenstrategieën uit de geheugenpsychologie voor in de klas.* Rotterdam, Okay Color.

Goossens, N. & Vermeer, A. (2009). Wat is een optimale tekstdekking? Woordkennis en tekstbegrip in groep 6. *Toegepaste Taalwetenschap in Artikelen*, *82*, 81-92.

Submitted Manuscripts

Goossens, N.A.M.C., Camp, G., Verkoeijen, P.P.J.L., Tabbers, H.K., Bouwmeester, S., & Zwaan, R.A. (2014). Distributed practice and retrieval practice in primary school vocabulary learning: A multi-classroom study. *Manuscript submitted for publication.*

In the ICO Dissertation Series dissertations are published of graduate students from faculties and institutes on educational research within the ICO Partner Universities: Eindhoven University of Technology, Leiden University, Maastricht University, Open University of the Netherlands, University of Amsterdam, University of Twente, Utrecht University, VU University Amsterdam, and Wageningen University, and formerly University of Groningen (until 2006), Radboud University Nijmegen (until 2004), and Tilburg University (until 2002). The University of Groningen, University of Antwerp, University of Ghent, and the Erasmus University Rotterdam have been 'ICO 'Network partner' in 2010 and 2011. From 2012 onwards, these ICO Network partners are full ICO partners, and from that period their dissertations will be added to this dissertation series. Over two hundred dissertations have been published in this series. The most recent ones are listed below.

207. Zitter, I.I. (04-02-2010). *Designing for learning: Studying learning environments in higher professional education from a design perspective*. Utrecht: Utrecht University.

208. Koopman, M. (11-02-2010). *Students' goal orientations, information processing strategies and knowledge development in competence-based pre-vocational secondary education*. Eindhoven: Eindhoven University of Technology.

209. Platteel, T. (11-02-2010). *Knowledge development of secondary school L1 teachers on concept-context rich education in an action-research setting*. Leiden: Leiden University.

210. Mittendorff, K. M. (12-03-2010). *Career conversations in senior secondary vocational education*. Eindhoven: Eindhoven University of Technology.

211. Moolenaar, N.M. (01-06-2010). *Ties with potential: Nature, antecedents, and consequences of social networks in school teams*. Amsterdam: University of Amsterdam.

212. Duijnhouwer, H. (04-06-2010). *Feedback effects on students' writing motivation, process, and performance*. Utrecht: Utrecht University.

213. Kessels, C.C. (30-06-2010). *The influence of induction programs on beginning teachers' well-being and professional development*. Leiden: Leiden University.

214. Endedijk, M.D. (02-07-2010). *Student teachers' self-regulated learning*. Utrecht: Utrecht University.

215. De Bakker, G.M. (08-09-2010). *Allocated online reciprocal peer support as a candidate for decreasing the tutoring load of teachers*. Eindhoven: Eindhoven University of Technology.

216. Groenier, M. (10-09-2010). *The decisive moment: Making diagnostic decisions and designing treatments*. Enschede: University of Twente.

217. Bonestroo, W.J. (24-09-2010). *Planning with graphical overview: Effects of support tools on self-regulated learning*. Enschede: University of Twente.

218. Vos, M.A.J. (30-09-2010). *Interaction between teachers and teaching materials: On the implementation of context-based chemistry education*. Eindhoven: Eindhoven University of Technology.

219. Kostons, D.D.N.M. (05-11-2010). *On the role of self-assessment and task-selection skills in self-regulated learning*. Heerlen: Open University of the Netherlands.

220. Bruin-Muurling, G. (21-12-2010). *The development of proficiency in the fraction domain: Affordances and constraints in the curriculum*. Eindhoven: Eindhoven University of Technology.

221. Slof, B. (28-01-2011). *Representational scripting for carrying out complex learning tasks*. Utrecht: Utrecht University.

222. Fastré, G. (11-03-2011). *Improving sustainable assessment skills in vocational education*. Heerlen: Open University of the Netherlands.

223. Min-Leliveld, M.J. (18-05-2011). *Supporting medical teachers' learning: Characteristics of effective instructional development.* Leiden: Leiden University.

224. Van Blankenstein, F.M. (18-05-2011). *Elaboration during problem-based small group discussion: A new approach to study collaborative learning.* Maastricht: Maastricht University.

225. Dobber, M. (21-06-2011). *Collaboration in groups during teacher education.* Leiden: Leiden University.

226. Jossberger, H. (24-06-2011). *Towards self-regulated learning in vocational education: Difficulties and opportunities.* Heerlen: Open University of the Netherlands.

227. Schaap, H. (24-06-2011). *Students' personal professional theories in vocational education: Developing a knowledge base.* Utrecht: Utrecht University.

228. Kolovou, A. (04-07-2011). *Mathematical problem solving in primary school.* Utrecht: Utrecht University.

229. Beausaert, A.J. (19-10-2011). *The use of personal developments plans in the workplace. Effects, purposes and supporting conditions*. Maastricht: Maastricht University.

230. Favier, T.T. (31-10-2011). *Geographic information systems in inquiry-based secondary geography education: Theory and practice.* Amsterdam: VU University Amsterdam.

231. Brouwer, P. (15-11-2011). *Collaboration in teacher teams.* Utrecht: Utrecht University.

232. Molenaar, I. (24-11-2011). *It's all about metacognitive activities; Computerized scaffolding of self-regulated learning.* Amsterdam: University of Amsterdam.

233. Cornelissen, L.J.F. (29-11-2011). *Knowledge processes in school-university research networks.* Eindhoven: Eindhoven University of Technology.

234. Elffers, L. (14-12-2011). *The transition to post-secondary vocational education: Students' entrance, experiences, and attainment*. Amsterdam: University of Amsterdam.

235. Van Stiphout, I.M. (14-12-2011). *The development of algebraic proficiency.* Eindhoven: Eindhoven University of Technology.

236. Gervedink Nijhuis, C.J. (03-2-2012). *Culturally Sesitive Curriculum Development in International Cooperation*. Enschede: University of Twente.

237. Thoonen, E.E.J. (14-02-2012). *Improving Classroom Practices: The impact of Leadership School Organizational Conditions, and Teacher Factors*. Amsterdam: University of Amsterdam.

238. Truijen, K.J.P (21-03-2012). *Teaming Teachers. Exploring factors that influence effictive team functioning in a vocational education context.* Enschede: University of Twente.

239. Maulana, R.M. (26-03-2012). *Teacher-student relationships during the first year of secondary education. Exploring of change and link with motivation outcomes in The Netherlands and Indonesia*. Groningen: University of Groningen.

240. Lomos, C. (29-03-2012). *Professional community and student achievement.* Groningen: University of Groningen.

241. Mulder, Y.G. (19-04-2012). *Learning science by creating models.* Enschede: University of Twente.

242. Van Zundert, M.J. (04-05-2012). *Optimising the effectiveness and reliability of reciprocal peer assessment in secondary education*. Maastricht: Maastricht University.

243. Ketelaar, E. (24-05-2012). *Teachers and innovations: on the role of ownership, sense-making, and agency*. Eindhoven: Eindhoven University of Technology.

244. Logtenberg, A. (30-5-2012). *Questioning the past. Student questioning and historical reasoning.* Amsterdam: University of Amsterdam.

245. Jacobse, A.E. (11-06-2012). *Can we improve children's thinking?* Groningen: University of Groningen.

246. Leppink, J. (20-06-2012). *Propositional manipulation for conceptual understanding of statistics.* Maastricht: Maastricht University.

247. Van Andel, J (22-06-2012). *Demand-driven Education. An Educational-sociological Investigation.* Amsterdam: VU University Amsterdam.

248. Spanjers, I.A.E. (05-07-2012). *Segmentation of Animations: Explaining the Effects on the Learning Process and Learning Outcomes.* Maastricht: Maastricht University.

249. Vrijnsen-de Corte, M.C.W. (13-09-2012). *Researching the Teacher-Researcher. Practice-based research in Dutch Professional Development Schools.* Eindhoven: Eindhoven University of Technology.

250. Van de Pol, J.E. (28-09-2012). *Scaffolding in teacher-student interaction. Exploring, measuring promoting and evaluating scaffolding.* Amsterdam: University of Amsterdam.

251. Phielix, C. (28-09-2012). *Enhancing Collaboration through Assessment & Reflection [Samenwerking Verbeteren door middel van Beoordeling en Reflectie]*. Utrecht: Utrecht University.

252. Peltenburg, M.C. (24-10-2012). *Mathematical potential of special education students*. Utrecht: Utrecht University.

253. Doppenberg, J.J. (24-10-2012). *Collaborative teacher learning: settings, foci and powerful moments*. Eindhoven: Eindhoven University of Technology.

254. Kenbeek, W.K. (31-10-2012). *Back to the drawing board. Creating drawing or text summaries in support of System Dynamics modeling.* Enschede: University of Twente.

255. De Feijter, J.M. (09-11-2012). *Learning from error to improve patient safety*. Maastricht: Maastricht University

256. Timmermans, A.C. (27-11-2012). *Value added in educational accountability: Possible, fair and useful?* Groningen: University of Groningen.

257. Van der Linden, P.W.J. (20-12-2012). *A design-based approach to introducing student teachers in conducting and using research.* Eindhoven: Eindhoven University of Technology.

258. Noroozi, O. (11-01-2013). *Fostering Argumentation-Based Computer-Supported Collaborative Learning in Higher Education*. Wageningen: Wageningen University.

259. Bijker, M.M. (22-03-2013). *Understanding the gap between business school and the workplace: Overconfidence, maximizing benefits, and the curriculum.* Heerlen: Open University of the Netherlands.

260. Belo, N.A.H. (27-03-2013). *Engaging students in the study of physics.* Leiden: Leiden University.

261. Jong, R.J. de (11-04-2013). *Student teachers' practical knowledge, discipline strategies, and the teacher-class relationship.* Leiden: Leiden University.

262. Verberg, C.P.M. (18-04-2013). *The characteristics of a negotiated assessment procedure to promote teacher learning.* Leiden: Leiden University.

263. Dekker-Groen, A. (19-04-2013). *Teacher competences for supporting students' reflection. Standards, training, and practice.* Utrecht: Utrecht University.

264. M.H. Knol (19-04-2013). *Improving university lectures with feedback and consultation.* Amsterdam: University of Amsterdam.

265. Diggelen, M.R. van (21-05-2013). *Effects of a self-assessment procedure on VET teachers' competencies in coaching students' reflection skills.* Eindhoven: Eindhoven University of Technology.

266. Azkiyah, S.N. (23-05-2013). *The effects of Two Interventions - on Teaching Quality and Student Outcome.* Groningen: University of Groningen.

267. Taminiau, E.M.C. (24-05-2013). *Advisory Models for On-Demand Learning.* Heerlen: Open University of the Netherlands.

268. Milliano, I.I.C.M. de (24-05-2013). *Literacy development of low-achieving adolescents. The role of engagement in academic reading and writing.* Amsterdam: University of Amsterdam.

269. Vandyck, I.J.J. (17-06-2013). *Fostering Community Development in School-University Partnerships.* Amsterdam: VU University Amsterdam.

270. Hornstra, T.E. (17-06-2013). *Motivational developments in primary school. Group-specific differences in varying learning contexts.* Amsterdam: University of Amsterdam.

271. Keuvelaar-Van den Bergh, L. (26-06-2013). *Teacher Feedback during Active Learning: The Development and Evaluation of a Professional Development Programme.* Eindhoven: Eindhoven University of Technology.

272. Meeuwen, L.W. van (06-09-13). V*isual Problem Solving and Self-regulation in Training Air Traffic Control.* Heerlen: Open University of the Netherlands.

273. Pillen, M.T. (12-09-2013). *Professional identity tensions of beginning teachers.* Eindhoven: Eindhoven University of Technology.

274. Kleijn, R.A.M. de, (27-09-2013). *Master's thesis supervision. Feedback, interpersonal relationships and adaptivity.* Utrecht: Utrecht University.

275. Bezdan, E. (04-10-2013). Graphical *Overviews in Hypertext Learning Environments: When One Size Does Not Fit All.* Heerlen: Open University of the Netherlands.

276. Bronkhorst, L.H. (04-10-2013). *Research-based teacher education: Interactions between research and teaching.* Utrecht: Utrecht University.

277. Popov, V. (08-10-2013). *Scripting Intercultural Computer-Supported Collaborative Learning in Higher Education.* Wageningen: Wageningen University.

278. Dolfing, R. (23-10-2013). *Teachers' Professional Development in Context-based Chemistry Education. Strategies to Support Teachers in Developing Domain-specific Expertise.* Utrecht: Utrecht University.

279. Lucero, M.L. (21-11-2013). *Considering teacher cognitions in teacher professional development: Studies involving Ecuadorian primary school teachers*. Ghent: Ghent University.

280. Kamp, R.J.A. (28-11-2013). *Peer feedback to enhance learning in problem-based tutorial groups.* Maastricht: Maastricht University.

281. Cviko, A. (19-12-2013). *Teacher Roles and Pupil Outcomes. In technology-rich early literacy learning*. Enschede: University of Twente.