Journal of Cognitive Psychology

Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/pecp21

Spreading the words: A spacing effect in vocabulary learning

Nicole A. M. C. Goossens, Gino Camp, Peter P. J. L. Verkoeijen, Huib K. Tabbers & Rolf A. Zwaan

Department of Psychology, Erasmus University Rotterdam, Rotterdam, The Netherlands
Published online: 27 Sep 2012.

To cite this article: Nicole A. M. C. Goossens, Gino Camp, Peter P. J. L. Verkoeijen, Huib K. Tabbers & Rolf A. Zwaan (2012): Spreading the words: A spacing effect in vocabulary learning, Journal of Cognitive Psychology, 24:8, 965-971

To link to this article: http://dx.doi.org/10.1080/20445911.2012.722617

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: http://www.tandfonline.com/page/terms-and-conditions

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae, and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand, or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.
Spreading the words: A spacing effect in vocabulary learning

Nicole A. M. C. Goossens, Gino Camp, Peter P. J. L. Verkoeijen, Huib K. Tabbers, and Rolf A. Zwaan

Department of Psychology, Erasmus University Rotterdam, Rotterdam, The Netherlands

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 generalises 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 practised 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.

Keywords: Learning strategies; Memory; Spacing effect; Vocabulary learning.

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 we will call massed repetition, but it is also possible to distribute repetition across multiple learning sessions, a procedure 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 memorising 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 generalisation 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 metre 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 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 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.

### 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, 13 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 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 a 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.

<table>
<thead>
<tr>
<th>Session 1</th>
<th>Session 2</th>
<th>Session 3</th>
<th>Session 4</th>
<th>Session 5</th>
<th>Session 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Presentation of all items (Items 1–30)</td>
<td>Items 1–15</td>
<td>Items 1–15</td>
<td>Items 1–15</td>
<td>Open-ended question test of all items after 1 week (Items 1–30)</td>
<td>Open-ended question test of all items after 5 weeks (Items 1–30)</td>
</tr>
<tr>
<td>Exercise 1</td>
<td>Exercise 2</td>
<td>Exercise 3</td>
<td>Feedback</td>
<td>Feedback</td>
<td>Feedback</td>
</tr>
<tr>
<td>Feedback</td>
<td>Feedback</td>
<td>Feedback</td>
<td>(spaced)</td>
<td>(spaced)</td>
<td>(spaced)</td>
</tr>
<tr>
<td>Items 16–20</td>
<td>Items 21–25</td>
<td>Items 26–30</td>
<td>Exercise 1</td>
<td>Exercise 1</td>
<td>Exercise 1</td>
</tr>
<tr>
<td>Exercise 1</td>
<td>Exercise 2</td>
<td>Items 26–30</td>
<td>Feedback</td>
<td>Feedback</td>
<td>Feedback</td>
</tr>
<tr>
<td>Items 16–20</td>
<td>Items 21–25</td>
<td>Items 26–30</td>
<td>Exercise 3</td>
<td>Exercise 3</td>
<td>Exercise 3</td>
</tr>
<tr>
<td>Exercise 3</td>
<td>Exercise 2</td>
<td>Items 26–30</td>
<td>Feedback</td>
<td>Feedback</td>
<td>Feedback</td>
</tr>
<tr>
<td>Items 16–20</td>
<td>Items 21–25</td>
<td>Items 26–30</td>
<td>(massed)</td>
<td>(massed)</td>
<td>(massed)</td>
</tr>
</tbody>
</table>
In Session 2, the children practised 15 items (three thematic sets) in the spaced condition and 5 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. Sessions 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 practised 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.

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 inter rater 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 analysed 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, $t(28) = -0.891$, $p = .380$, $d = 0.166$.

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) × 2 (retention interval) repeated measures ANOVA. The analysis revealed a main effect of condition, $F(1, 32) = 10.118$, $p = .003$, $\eta^2_p = .240$: Retention in the spaced condition was better than retention in the massed condition. Also, there was an effect of retention interval, $F(1, 32) = 13.103$, $p = .001$, $\eta^2_p = .291$. 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 | After 1 week | After 5 weeks |
| Massed | 46.46% (25.85) | 42.22% (23.07) |
| Spaced | 55.96% (26.24) | 49.49% (27.13) |

Final test
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, whereas 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 one 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 hypothesise 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 optimisation 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.

Original manuscript received February 2012
Revised manuscript received August 2012
First published online September 2012

REFERENCES


