<table>
<thead>
<tr>
<th>Title</th>
<th>Examining an Appropriate Burden in a Vocabulary Quiz and an Optimal Interval Between Two Quiz Sessions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Author(s)</td>
<td>KASAHARA, Kiwamu; KABARA, Masaki</td>
</tr>
<tr>
<td>Citation</td>
<td>北海道教育大学紀要・人文科学・社会科学編, 69(1): 15-25</td>
</tr>
<tr>
<td>Issue Date</td>
<td>2018-08</td>
</tr>
<tr>
<td>URL</td>
<td><a href="http://s-ir.sap.hokkyodai.ac.jp/dspace/handle/123456789/9874">http://s-ir.sap.hokkyodai.ac.jp/dspace/handle/123456789/9874</a></td>
</tr>
</tbody>
</table>
Examining an Appropriate Burden in a Vocabulary Quiz and an Optimal Interval Between Two Quiz Sessions

KASAHARA Kiwamu and KABARA Masaki*

*Graduate School, Asahikawa Campus, Hokkaido University of Education
Department of English Education, Asahikawa Campus, Hokkaido University of Education

ABSTRACT

This study examined an effective way of implementing small vocabulary quizzes into EFL classroom settings. The present study dealt with two variables: the number of words and the length of interval between two learning sessions. Two groups learned the same 80 target words two times in two different conditions. Group 1 had eight weekly learning cycles: they learned 10 of them two times with the interval of one hour every week. On the other hand, Group 2 had four biweekly learning cycles: they learned 20 of them two times with the interval of one week. The total amount of learning time was the same between the groups. Both groups took a pre-test, an immediate post-test and a one-week delayed post-test to ask them for the meanings of the 80 target words. It was found that Group 2 got a significantly higher score than Group 1 in the delayed post-test. Learning a large number of words with a long interval seems to be better for long-term retention than learning a small number of words with a short interval. The present study showed that too short an interval was not good for learners in retention if the learning chances are given only twice, and that the number of words in one quiz session should not be too small.
1. Introduction

1.1 The Importance of Intentional Vocabulary Learning for Novice L2 Learners

Vocabulary is one of the most important aspects of second language (L2) learning. Wilkins (1972) skillfully described a crucial role of vocabulary in L2 learning: “Without grammar, little can be conveyed, without vocabulary, nothing can be conveyed” (p. 111). Learners can understand and produce nothing without vocabulary. Lightbown and Spada (2006) argued that learners can communicate by using words that are not placed in the proper order, pronounced perfectly, or marked with the proper grammatical morphemes, but communication often breaks down if learners do not use the correct words. No language teachers can deny the necessity of effective vocabulary teaching.

It is well known that there are two types of vocabulary learning: incidental learning and intentional learning (Hatch & Brown, 1995; Nation, 2001). Incidental vocabulary learning is to gain vocabulary knowledge through listening or reading activities (Hulstijn, 2003). The main purpose is not learning vocabulary but understanding spoken or written input. Learners are usually exposed to new lexical items in context. This is called contextualizing. On the other hand, the aim of the intentional learning is vocabulary learning itself. Paired-associate learning, which is to connect L2 forms and their first language (L1) translations through word lists or word cards, is a typical example of intentional vocabulary learning. This is called decontextualizing.

In the 1980s, decontextualizing was severely criticized by the proponents of communicative language teaching because it is not a natural way of language learning (Krashen, 1982, 1985; Oxford & Crooktail, 1990). They argued that decontextualizing cannot give L2 learners any useful information on how target words are used in actual communicative contexts. Krashen (1982; 1985) insisted that intentionally learned knowledge cannot become acquired knowledge, which can be used instantly in real communication.

However, there has been a tendency to reevaluate intentional vocabulary learning (Kasahara, 2015). A number of vocabulary researchers have argued that incidental vocabulary learning is not sufficient and should be supplemented by intentional form-focused vocabulary learning (Cobb & Horst, 2004; Laufer, 2005). Nation (2013) states that the two types learning are “complementary activities, each one enhancing the learning that comes from each other” (p. 348). Incidental learning can help learners to know how to use words in context, whereas intentional learning can help them to establish form-and-meaning connections. The latter is especially crucial for novice L2 learners because they need to know the meanings of high-frequency words before they can obtain some vocabulary knowledge by guessing from context. This is the very case with EFL learners, who cannot receive a large amount of English input outside English classrooms.

In fact, several studies have shown that intentional paired-associate word learning is an effective way for novice learners, who are required to learn high frequency words in a limited time. This is because they can focus on the link between L2 word forms and their meanings in their L1 (Nation, 2013; Nakata, 2008). Paired-associate learning is thought to give less burden to learners than context learning. Prince (1996) compared paired-associate learning and context learning for novice L2 learners. In his context learning
condition, participants learned 44 English words in example sentences including the target words. There were no L1 translations beside the sentences. In the paired-associate learning condition, participants received the list of the same 44 English words to learn, accompanied by their L1 (French) equivalents. He found that the paired-associate learning condition group got significantly higher scores than the context learning condition group. The participants in the paired-associate learning group memorized more words by linking the L2 targets and their L1 equivalents than those in the context learning group.

Laufer and Shumueli (1997) compared the following four groups: the list group, the sentence group, the text group, and the elaborated text group. The list group learned 20 target items with the word list including the 20 target words and their meanings. In this list, 10 words were attached to their L1 translations, and the other 10 words were paired with their L2 definitions or synonyms. The sentence group learned the same 20 word pairs as the list group used. In addition, they were given 20 example sentences, each of which had one of the target words. The text group read a passage which was chosen from a British course book. All the 20 target words appeared in the text and were glossed in L1 or L2 in the left margin of the text. The elaborated text group learned the 20 target items in an elaborated text, a revised version of the passage the text group read. The elaborated text had either a synonym or a brief explanation after each target word. This elaboration aimed to facilitate the participants to understand the text and the meaning of the target words while reading. The results of their study showed that the most effective methods for long-term retention were list learning and sentence learning.

According to these studies, it is effective for novice L2 learners to learn vocabulary from paired-associate learning such as word lists or word cards because learners can focus on the forms and the meanings of the target items. The novice learners have low L2 processing abilities and they do not process a large amount of information at a time. Therefore, it is important for them to focus on the forms and the meanings of the target items.

1.2 The Effectiveness of Small Vocabulary Quizzes

Learning new words in small recall quizzes is an effective way to enhance the effect of paired-associate learning. Letting learners retrieve L1 meanings from their L2 forms or vice versa can consolidate the form-and-meaning connections. Retrieving learned information from memory is called retrieval practice, which is thought to be a powerful tool to enhance long-term retention of the learned item (Barcroft, 2007). Kanayama and Kasahara (2015) shows two reasons why retrieval practice is effective for long-term retention. First, retrieving stored information from memory requires a great mental effort, which can lead to better performance later. The great mental effort in a small vocabulary quiz can create a deeper trace in a learner’s mental lexicon. Second retrieval practice can help learners to have an effective learning plan. During the retrieval practice, learners can notice which items have already stored and which items have not. This experience can help them to focus on the items they have not acquired in the next learning session.

Repetition is indispensable for vocabulary learning and repeated encounters with the same lexical items can increase a possibility to master
them (Coxhead, 2006; Nation, 2013, Webb, 2007). This is also the case with vocabulary quizzes: recall tests on the same vocabulary should be repeated with an interval between each test. This is called spaced retrieval. Kanayama and Kasahara (2017) examined how to conduct spaced retrieval in repeated sessions of intentional vocabulary learning. They compared three types of distributed learning: equal space learning, expanding space learning, and contracting space learning. Equal spacing means that the intervals between the sessions were always equal. Expanding spacing means that the intervals between sessions become longer and longer toward the end of the learning period. Contracting spacing means that the intervals between sessions become shorter and shorter toward the end of the learning period. They found that if learners were given three or more times to learn, there was no significant difference between the three types of spaced learning.

However, the results of the immediate tests showed that expanding space group got a higher mean score in the second trial than the other groups. If learners are given only two times to learn, it is not clear which types of interval were better. Therefore, this study focused on how the length of interval between two learning sessions affect retention of target items. In addition, few studies have examined optimal learning burden in one small quiz. Learning burden in small quizzes means the number of words in one small quiz. Hence, the intervals of small quizzes in two sessions and the number of words in one quiz session are the research questions of this study.

2. Method

2.1 Research Question

The present study set up the following research question:

What is the appropriate number of words in a vocabulary quiz and what intervals are optimal between two quiz sessions?

The present study investigated which condition was more effective for intentional vocabulary learning: (a) 10 words with an hour interval and (b) 20 words with a week interval. Through investigating this research question, this study was expected to clarify the appropriate number and the optimal interval in a vocabulary quiz.

2.2 Participants

The participants were 57 Japanese university students. They belonged to two different regular English classes. They took 15 weekly lessons in the English course. There were 40 students in Group 1. However, this study excluded students who were absent from any lesson from the analyses, so Group 1 consisted of 33 participants. Following the same procedure as Group 1, Group 2 excluded 14 students out of 38 from the analyses. Group 2 consisted of 24 participants in the end.

The two groups learned the same 80 target words two times under different conditions. Group 1 learned every 10 words in two learning sessions with about a one-hour interval weekly. They took an immediate test after each learning session. They repeated this cycle 8 times. Group 2 learned every 20 words in two learning sessions with a one-week interval biweekly: They learned 20 words in one lesson, and they learned the same 20 words in the next week. They also took an immediate test after each learning session. They repeated this cycle 4 times.

2.3 Material

The target words were selected from the following three academic word lists: Coxhead
(2000), Gardner and Davies (2014), and Browne (2013). One reason why these three academic word lists were selected was that they are well-known and proved to be useful word lists. Another reason was that the participants were university students. This experiment was expected to be of some benefit for the participants.

Coxhead (2000) described the development and evaluation of her new academic word list. It was compiled from a corpus of 3.5 million running words of academic text by examining the range and frequency of words outside the first 2,000 most frequently occurring words of English. She removed the 2,000 most frequent words in daily life. By highlighting the words that university students meet in a wide range of academic texts, the list shows learners with academic goals which words are most worth studying. The list also provides a useful basis for further research into the nature of academic vocabulary.

Gardner and Davies (2014) used 120 million words of academic texts in the Corpus of Contemporary American English (COCA). They used lemmas, not word families, to make initial counts and analyses, and then word families were formed from the lemmas to support certain academic needs. This list provided a good coverage of academic English. It turned out to be more highly oriented towards an academic genre than other genres. For example, it covers 14.0% of academic texts in COCA, 7.3% of the 85 million words of newspapers in COCA, and just 3.4% of the 86 million words of fiction texts in COCA.

The list that Browne (2013) compiled is called New Academic Word List. It was based on a carefully selected academic corpus of 288 million words: academic journals, non-fictions, student essays, academic discourses, Michigan Corpus of Academic Spoken English (MICASE), British Academic Spoken English (BASE), and Corpus of 100s of top-selling academic textbooks. This list aimed to create a list of the most important high-frequency words useful for L2 learners of English, ones which gives the highest possible coverage of English texts with the fewest words possible.

The candidates for the target words were 182 words which are included in all the three academic word lists above. This study planned to be conducted for 15 weeks, and it needed a large number of target words, which were not familiar to the participants. In order to examine which words were less known to university students, a preliminary survey was conducted to 28 university students who were 1st, 2nd, and 4th grade students in the English education major. The survey asked them whether they knew the meanings of the 182 words. According to the results of the survey, the least known 100 words were selected. All the 100 words were planned to be used in this study. However, it turned out that the experiment could deal with only 80 words out of these because of restriction of the time to implement the experiment.

Next, word lists with which the participants studied the 80 target items were made. For Group 1, the authors made eight lists, each of which included 10 target words. The target words and their L1 meanings were written in the lists for Group 1. An immediate recall test was also made to examine how well the participants remembered the meanings of the target words. The test asked the participants write down the meanings of the target words in blanks next to the target words. Each test for Group 1 had 10 items; each test for Group 2 had
20 items.

In order to see difference in vocabulary gain between the groups, a pre-test and two post-tests were made. The pre-test and the immediate post-test were the same yes-no tests. They asked the participants whether they knew the meanings of the 80 target words. The delayed post-test was different from these two tests. That was a yes-no test with 4-option multiple choices. This also asked whether they knew the meanings of the 80 target words. If participants chose yes, they had to choose the correct Japanese meaning of the target item out of four choices. The reason why we made it different from the pre-test and the immediate post-test was to confirm whether they really remembered the meanings of the target items. If possible, it would be ideal to let them write down the meanings of the target items. However, it was difficult to carry it out because this experiment was conducted in a given limited time in the regular English courses.

2.4 Procedure

The pre-test was conducted in the first lesson. This test asked the participants whether they knew the meanings of the 80 target words. The format of this test was a yes-no test. The participants chose "yes" if they knew the meaning of each target word, and chose "no" if they did not. They did not have to write down the meanings of the items. They were given enough time (about 10 minutes) to deal with all the items.

From the second lesson to the ninth lesson, the participants had learning sessions and immediate tests. In Group 1, the participants were given a list including 10 words with their Japanese translations. They remembered the meanings of the targets for two minutes and half. Then, after the lists were collected, an immediate test was conducted for one minute and half. This session was conducted two times at the beginning and at the end of each weekly lesson. Therefore, they learned 10 words at one time (a light burden) and the interval between the two sessions was about one hour (a short interval). In Group 2, the participants were given a list including 20 words with their Japanese equivalents. They remembered the meanings of the target items for five minutes. Then, after the lists were collected, an immediate test was conducted for three minutes. A week later, in the next lesson, they had the same learning session and the immediate test. These two sessions were conducted every two weeks. Therefore, they learned 20 words at one time (a heavy burden) and the interval between the two sessions was one week (a long interval). Group 1 had eight cycles of double learning sessions with 10 words in each session; Group 2 had four cycles of double learning sessions with 20 words in each session. The participants in Group 1 and Group 2 had the same time to learn the same target items in total. The procedural difference between the group is shown in Figures 1 & 2.

In the 10th lesson, the post-test was conducted in each group. This was the same as the pre-test. The participants were asked again whether they knew the meanings of the target words after the whole experimental intervention.

In the 11th lesson, the delayed post-test was conducted in each group. This test was different from the pre-test and the post-test as mentioned in the material section. The participants were asked to show whether they knew the meanings and to choose the correct Japanese meaning of the item from 4 options if they showed that they knew the meaning of the item.
An Appropriate Burden and an Optimal Interval in Vocabulary Quizzes

2.5 Scoring

One point was given to each correct answer and zero point to each wrong answer in the immediate tests. The maximum score was 10 points for Group 1 and 20 points for Group 2. In the pre-test and the post-test, one point was given if a participant chose ‘yes’ and zero if a participant chose ‘no’. The maximum score was 80 points. In the delayed post-test, one point was given for ‘yes’ with the correctly chosen answer, and zero point was given to ‘no’ or ‘yes’ with a wrong choice.

2.6 Data Analysis

In order to know if there was an interaction and any main effects between the groups, a two-way ANOVA was conducted. In addition, t tests were conducted to know the difference between the groups in each test.

3. Results

Table 1 shows the means and the standard deviations for the pre-tests, the post-tests, and the delayed post-tests of Group 1 and Group 2. In addition, Figure 3 shows the changes of the means of each group over the three tests.

<table>
<thead>
<tr>
<th></th>
<th>Pre</th>
<th>Post</th>
<th>Delayed</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Group 1 (n=33)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M</td>
<td>12.88</td>
<td>24.03</td>
<td>23.64</td>
</tr>
<tr>
<td>SD</td>
<td>10.11</td>
<td>11.19</td>
<td>10.45</td>
</tr>
<tr>
<td><strong>Group 2 (n=24)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M</td>
<td>17.54</td>
<td>31.17</td>
<td>35.42</td>
</tr>
<tr>
<td>SD</td>
<td>13.35</td>
<td>18.45</td>
<td>14.57</td>
</tr>
</tbody>
</table>

The result of the two-way ANOVA found that there was a significant interaction between the groups, $F(1,55) = 3.25, p = .04, \eta^2 = .56$. In addition, there was a significant main effect between the groups, $F(1,55) = 6.60, p = .013, \eta^2 = .11$. Moreover, there was a significant main effect between the tests, $F(1,55) = 50.42, p = .0001, \eta^2 = .65$.

Furthermore, t tests showed that there were no significant difference between Group 1 and Group 2 for the pre-test, $t(55) = -1.50, p = .14, r = .19$, or the post-test, $t(55) = -1.81, p = .08, r = .23$. However, there was a significant difference between the groups for the delayed post-test, $t(55) = -3.56, p = .001, r = .42$. The effect size was medium.

The results seem to show that Group 1 performed better than Group 2. Though there was no significant interaction or any significant
difference in the pre-test or the post-test between the groups, the \( t \) test showed that Group 2 had a significantly higher mean score than Group 1 in the delayed post-tests.

4. Discussion

According to these results, learning a large number of words with a long interval seems to be better than learning a small number of words with a short interval for the long-term memory. Group 2 got significantly higher scores than Group 1 in the delayed post-test though there was no significant difference between the groups in the post-test. A possible reason is that the delayed post-test took a different test format from the post-test. The delayed post-test asked the participants to choose the correct meaning of a target word from four choices. Giving them the choices including the key might have helped them to retrieve the correct meaning of the target word. This might have been the case especially when their memory of the target meaning was vague. The delayed post-test may have tapped their partial knowledge on the target word meanings that the post-test could not tap, because the post-test asked them whether they knew the meanings of the targets without giving any choices. The delayed post-test may have revealed subtle difference in memory between the groups, which was not shown in the post-test. It could be said that Group 2 had more vocabulary gain than Group 1.

Another possible reason is that Group 1 may have got overconfident on how well they remembered the meanings of the target items in the course of each learning session. In each immediate test, Group 1 got a higher mean score than Group 2. The score was close to full marks in the second trial of each immediate test (See Figure 4). They may have thought that they had already learned all the target words during the first learning session, and they may have paid less attention to each item toward the end of the second learning session. Furthermore, there were no feedbacks from the instructor after each immediate test. They could not confirm whether they really remembered the target items, which could have helped them to become overconfident with their learning performance in the first session. If they had had some feedbacks after the immediate test, they might have noticed that there were still some target words that they had not mastered.

On the other hand, Group 2 had the 20 words to learn in each session, and therefore they may have thought that they needed a certain amount of time and effort to remember the meanings of the target items. They may have tried to learn the items harder than Group 1. Group 2 got a lower mean score than Group 1 in each immediate test. In the first trial of each immediate test, they answered less than 70 percent of the items correctly. They might have had the image that they had not remembered all the meanings of the items. This image may have led them to making more effort to remember the same items again in the second trial.

![Figure 4. Means of the immediate tests (the percentage the correct answers in each test)](image-url)
In addition, Group 2 had a one-week interval between the first learning session and the second learning session. They may have noticed their memory attrition of the target items in the second learning session, and therefore they may have tried to learn items harder than Group 1. This may also have led to the long-term retention.

There are three limitations in this study. The first limitation is the small number of the participants. In addition, the gap of the number between the groups is also problematic. The participants reduced to 33 in Group 1 and 27 in Group 2 because this experiment was a longitudinal study. It was conducted for 12 weeks. Some students who were absent from one of more lessons were excluded from the data analyses. Further studies need more participants, and the gap of the number between the groups should be controlled.

Secondary, there were two variables in this study: the number of words and the length of interval. The study could not reveal which variable affected the results stronger. Therefore, further studies that will focus on one of the variables are necessary.

Thirdly, the formats of the post-test and the delayed post-test were different. This study was conducted in the regular English course, and available time in each class for the experiment was limited. Because letting the participants write down the meanings of all the items was time-consuming, the pre-test and the post-test took a yes-no test format. However, to see whether they really remembered the meanings of the items, the delayed post-test was changed into a multiple choice test format. To get rid of this test-format difference, further studies should use the same multiple choice test for the pre-test, the post-test and the delayed post-test.

5. Conclusion

The main purpose of this study was to examine the appropriate number of words in a vocabulary quiz and what intervals are optimal between two quiz sessions. Learning 10 words with a one-hour interval and learning 20 words with a one-week interval were compared. It was found that learning 20 words with a one-week interval was more effective for long-term retention of target word meanings than learning 10 words with a one-hour interval. Learning a large number of words with a long interval seems to be better for long-term retention than learning a small number of words with a short interval. According to the Ebbinghaus’ forgetting curve, the first interval was thought to be made short because most forgetting occurs soon after the first learning occasion. However, this study showed that too short an interval was not good for learners in retention if the learning chances are given only twice. The number of words in one quiz session should not be too small. Thus, giving the learners appropriate learning burden is important for their long-term memory.

To apply the results of the present study to L2 vocabulary teaching, there are two pedagogical implications. First, English teachers should give the learners an appropriate learning burden when they learn new words. Too light a learning burden is not good for learners. They may have the feeling that they can remember all target words without a large amount of effort and they may not concentrate on learning the target words toward the end of a learning session. In a vocabulary small quiz, teachers should make quizzes which include an appropriate number of words.

Second, teachers should give an optimal
interval between different small quizzes. Too short an interval might not be good for learners' long-term memory. If the interval is very short, learners still remember the target words in the second trial. Retrieving word meanings without difficulty may not lead to their long-term retention of the target words. If teachers have only two quiz sessions, the second trial should be conducted a certain period of time after the first trial.

REFERENCES


（笠原 究 旭川校教授）
（鹿原 真生 旭川校大学院生）