The effects of spaced versus massed distribution instruction on EFL learners’ vocabulary recall and retention

Ehsan Namaziandost1,*, Murad Hassan Mohammed Sawalmeh2 and Masoumeh Izadpanah Soltanabadi3

Abstract: The current study investigates the effect of massed and spaced instruction on vocabulary recall and retention. To fulfill this objective, 75 Iranian pre-intermediate EFL learners (16 to 19 years) took part in 15 sessions of 60 minutes. The participants were randomly divided into three experimental groups; a spaced distribution group \((n=25)\), a massed distribution group \((n=25)\) and a control group \((n=25)\). The massed distribution group had one intensive session on learning the target vocabulary; the spaced distribution group had three sessions at irregular time intervals, and the control group received no vocabulary-focused instruction. Using a before and after design, students were retested after 5 weeks. To collect data, a receptive vocabulary test was administered as both the pretest and the posttests. The results of One-way ANOVA indicated that the spaced distribution group significantly outperformed the massed distribution group on both immediate and delayed posttests. The results propose that EFL practitioners should synthesize spacing as a beneficial teaching technique into the curricula, instruction and educational materials to promote vocabulary learning in real classroom setting.

Subjects: Educational Research; Education Studies; Higher Education; Theories of Learning; Teachers & Teacher Education; Language & Linguistics; Language Teaching & Learning

ABOUT THE AUTHOR

Ehsan Namaziandost was born in Shiraz Province of Iran in 1985. He holds an MA degree in TEFL from Islamic Azad University of Ahvaz. Now, he is a Ph.D. candidate of TEFL at Islamic Azad University, Shahrekord, Iran. His main interests of research are TEFL, CALL, Second Language Acquisition, EFL Teaching and Learning, Language Learning and Technology, Teaching Language Skills, and Language Learning Strategies. His research papers and articles have been published by different international journals.

PUBLIC INTEREST STATEMENT

Vocabulary represents one of the most important skills necessary for teaching and learning a foreign language. It is the basis for the development of all the other skills: reading comprehension, listening comprehension, speaking, writing, spelling and pronunciation. Therefore, due to the importance of vocabulary learning, teachers are always searching for the best techniques to teach them successfully. In this study, massed instruction, in which individuals practice a task continuously without rest and spaced instruction, in which individuals are given rest intervals within the practice sessions, was implemented to check their effects on vocabulary recall and retention. After implementing the treatment sessions, it was revealed that the spaced instruction group performed better than the massed distribution group on both immediate and delayed posttests.
Keywords: distribution of instruction; spaced distribution instruction; massed distribution instruction; L2 vocabulary learning; L2 vocabulary recall and retention

1. Introduction

Learning outcomes have been evaluated in regards to variations in teaching methodology such as the distribution of learning practice. A substantial body of studies has sought to determine whether the variation in the design of distributed instruction makes a difference in learning outcomes (Lotfolahi & Salehi, 2016; Mashhadi et al., 2017; Namaziandost, Nasri, Rahimi Esfahani et al., 2019; Namaziandost, Rahimi Esfahani et al., 2018). According to Carpenter et al. (2012), most instructors and students are concerned about the timing of study to maximize carryover of previously learned materials and enhance learning/teaching outcomes. Likewise, in the field of language learning, it is unclear to many EFL teachers and learners, whether presenting and studying material across two or more sessions that are separated in time (i.e., spaced distribution instruction) results in better learning than spending the same total amount of time in a single session of massed distribution instruction (Miles, 2014; Namaziandost, Homayouni et al., 2020).

In second/foreign language learning and teaching, in particular, for the development of language skills, it is important for the learner to be exposed to repetitive and regular examples of linguistic objects or structures (DeKeyser, 2007; Mashhadi et al., 2017; Namaziandost, Hosseini et al., 2020; Segalowitz, 2010). What is not entirely clear, however, is whether this repetition of inputs should be spaced or massaged preferably (Rogers, 2017; Segalowitz, 2010). In cognitive psychology, learning was revealed to develop when the repetitions of the items to be learned occur in spaced sequences, as opposed to massive or concentrated presentations (Hosseini et al., 2017; Segalowitz, 2010). However, given the abundance of spacing effect work in cognitive psychology, few real-class experimental studies have examined the effects of spaced delivery instruction on second/foreign language learning and teaching (Etemadfar et al., 2019; Namaziandost, Nasri, Rahimi Esfahani et al., 2019; Serrano, 2012). This deficiency of real-classroom research investigating the spacing effect on language learning is specifically highlighted for further studies in the field of language learning by Ellis (2006) and Rogers (2017).

Meanwhile, one of the controversial problems in vocabulary instruction to second/foreign language learners is that while the concept of spaced distribution instruction and its supremacy over massed distribution instruction have been checked and emphasized by numerous researchers (Namaziandost, Rezvani et al., 2020; Serrano & Munoz, 2007; Stoltzfus & Sukseemuang, 2018); such studies have concentrated on second/foreign grammar and language skills (Mashhadi et al., 2017; Namaziandost, Pourhosein Gilakjani et al., 2020), but rare studies have been done to check whether the same superiority applies to vocabulary learning, specifically in English as a foreign language (EFL) classroom contexts. Hence, Ellis (2006) suggested that the problem of the massed vs. spread delivery instruction remained unanswered, at least with respect to vocabulary instruction, and further work should, therefore, explore this topic further.

In a line of previous research, Allen (1983) stated that foreign language (FL) teachers need additional help with vocabulary instruction because even where teachers have devoted a lot of time to vocabulary teaching, much of the core vocabulary is not retained even after weeks, months, or even years of English instruction. Cameron (2001) noted two issues with regard to the teaching of L2 vocabulary in an ESL situation. The first issue is that meeting and understanding a new word is just the onset of a learner’s vocabulary acquisition process. The second and more challenging issue is how to create a long-term memory of a word so that it will be available to the learner for future usage. She further added that learners need to use memorizing activities at the point of learning new vocabulary words for the first time, and that they need to regularly review those words at intervals (Cameron, 2001). Although most available literature is in agreement that for FL/L2 learning one should implement word repetition at the time of study and at intervals
during continued learning, it is not as clear how these repetitions should best be implemented in foreign vocabulary teaching as a memory aid.

Exploring the beneficial effects of spacing in learning has been an active area of research in psychological sciences under the name of the spacing effect. The spacing effect refers to a memory advantage whereby memory is enhanced when learning episodes are spread over longer periods of time rather than being massed in one single session (e.g., Cepeda et al., 2006). In general, it is necessary to distinguish between two types of repetitions, namely restudy and retrieval practice (Goossens et al., 2014). The research in cognitive psychology has shown that using retrieval practice leads to better memory than restudy in the learning phase (Roediger & Karpicke, 2006). This phenomenon is commonly referred to as the retrieval practice effect or testing effect. The testing effect refers to a memory phenomenon whereby testing has a more reinforcing impact on memory than restudying. The testing effect is a very well-documented phenomenon (Bjork, 1975; Goossens et al., 2014), but its tremendous effects are unknown to many scholars outside the domain of cognitive psychology. For instance, Ur (2012) stated that reviewing words results in more effective learning than testing words. She also stated that the tests merely assess students’ knowledge, but do not promote further learning. In this example, Ur (2012) fails to make the definite distinction between recognition and understanding. With respect to learning, the former reflects superficial initial encoding and the latter the functional ability to retrieve and use a word in context.

2. Literature review
 Encoding instability and flawed transmission are two basic theories that account in reality for the observable dominance of the spread distribution (Greene, 1989). The theory of encoding instability indicates that spaced materials are easier to remember than massed materials because each appearance in the spaced distribution is presented in a novel fashion and thus provides more reminder hints to the learner. Nevertheless, this theory focuses on the position of the particular circumstance and asserts that the sense in which an object is offered is coded along with its meaning (Anderson & Bower, 1972; Namaziandost, Rahimi Esfahani et al., 2018). However, the flawed processing hypothesis (Challis, 1993) suggests that the massed display does not require sufficient information processing. In the case of massed distribution, all of the material is presented at once and the learner does not have to engage in anything more than superficial recall of the current presentation. On the other hand, when a subject is introduced after a certain period of time has passed and after certain additional objects have been shown since the initial introduction, a more thorough analysis will inevitably arise because the previous presentation will not be as easily accessible as in the case of mass instruction (Mashhadi et al., 2017; Namaziandost & Çakmak, 2020).

In the meantime, the superior learning impact of spaced instruction over massed instruction has been shown in some of the previous researches on learning L2 grammar (Mashhadi et al., 2017; Miles, 2014), reading comprehension (Namaziandost, Hashemifardnia et al., 2018), and vocabulary (Bohrick et al., 1993; Çekiç & Bakla, 2019; Chukharev-Hudilainen & Klepička, 2016; Goossens et al., 2012; Kornell, 2009; Lotfolahi & Salehi, 2016; Nakata, 2015; Sobel et al., 2011). Notably, although these studies highlighted the spacing effect on vocabulary learning, the outcomes vary in terms of the superiority of one method over the other. The present research has aimed to contribute to the discussion with empirical results from a real-classroom context, which creates a solid connection between learning in the educational setting and the existing research (Sobel et al., 2011).

To date, Kornell (2009) investigated the superiority of spacing vs massed distribution with 20 words and their synonyms. The spaced learning group significantly outperformed the massed group in a recalling of word pair task.

Sobel et al. (2011) had 39 middle-school children study eight new English words during two sessions with a one-week break between study sessions. The children learned the words under two
different learning conditions (massed vs. spaced). In the massed condition, the two study sessions took place in immediate succession during one encounter. In the spaced condition, however, the two learning sessions were separated by a one-week break in between study sessions. Thirty-five days after the second learning session, a cued recall test assessed children’s performance. The results revealed that the recall for spaced items was vastly better than the recall for massed items.

Similarly, in a study by Goossens et al. (2012), 48 elementary school children were exposed to 15 unfamiliar words in the massed session and 15 other unfamiliar words in the spaced intervals. In the massed condition, the target words were divided into three sets of five words each and children practiced each set three times in one of the three study sessions. In the spaced condition, the children studied the words across three consecutive sessions during which the children studied the words once in each of the three study sessions. A retention test assessed children’s recall 7 days and 35 days after the last study session. The results showed that children in the study recalled the spaced words better than the massed words.

Lotfolahi and Salehi (2016) explored learners’ perceptions of different spacing schedules (massed vs. spaced). To achieve the purpose of the study, the researchers taught 30 children 24 English–Farsi word pairs utilizing different spacing schedules. Later, they administered a questionnaire to explore learners’ perceptions of both massed and spaced schedules. The findings indicated that the children perceived spaced instruction to be more effective than massed instruction.

Nakata and Suzuki (2018) have also examined the effects of massing and spacing on the learning of semantically related and unrelated words. One hundred and thirty-three Japanese university students studied 48 English-Japanese word pairs in their study, under two conditions: massed and spaced. Half the words were semantically interrelated while the other half were not. Although there were no significant differences in post-test scores between semantically linked and non-related items, semantically related items contributed to more intrusion errors than unrelated items. In addition, contrary to the authors’ hypothesis that spacing is especially beneficial for semantically related items, spacing benefited unrelated items more than related items.

More recently, Çekiç and Bakla (2019) investigated the effect of different spacing instructions on L2 incidental vocabulary learning. Seventy-seven Turkish learners of English with an intermediate level of proficiency were exposed to 20 target words embedded in short reading texts in three groups. Each group had the same treatment, and material differing only in distribution design. The first group had readings and target words once per session for 9 weeks (fixed spacing group). The second group had readings and target words for 7 weeks with a two-week interval (spaced massing with fixed intervals) and the third group had the same intervention in different distribution designs: three times in each session with one-week intervals (spaced massing with expanded intervals). The Vocabulary Knowledge Scale (VKS) and a multiple-choice test immediately after intervention were employed at specific phases of the research. The results showed that the fixed spacing group exceeded the spaced massing at fixed intervals and spaced massing at extended intervals.

The findings of former research investigating input spacing in relation to second-language learning have given an uncertain picture of whether exposure to a second language, whether in educated or naturalistic contexts, should be massaged or spaced. Some studies have shown that massed (i.e., intensive) distribution instructions appeared better than scattered distribution instructions when tested only on immediate posttests (Namaziandost, Nasri, Rahimi Esfahani et al., 2019; Serrano, 2011; Serrano & Munoz, 2007; Stoltzfus & Sukseemuang, 2018). Conversely, there is evidence that spaced distribution instructions are more effective in preserving target language structures than mass distribution instructions. This measurement is achieved with a subsequent delayed posttest (Schuetze, 2015).
This paper aims to examine the impact of spaced and massed distribution on EFL learners’ vocabulary learning. Therefore, the study addresses the following research questions:

1. Does variation in distributed instruction (spacing, mass, and no distributed instruction) have any significant effects on Iranian EFL learners’ vocabulary recall and retention?
2. Is there a significant difference between Iranian EFL learners’ vocabulary recall and retention learning through spacing instruction versus massed instruction?

Taking existing theoretical arguments and findings of spacing effect on vocabulary learning into consideration, the current study aimed to test a hypothesis in relation to research questions. The hypothesis is premised on the theoretical foundation of the Encoding Variability Theory (Glenberg, 1979) that spaced distribution makes a difference in recall performance and that knowledge is retained better when provided with distributed instruction (Suzuki & DeKeyser, 2017). It was expected that participants exposed to spaced instruction would outperform those exposed to the massed instruction in terms of vocabulary recall and retention. Additionally, findings of existing studies show that spacing yields better performance in vocabulary learning and retention both in immediate tests after the experiences and over time (Namaziandost, Nasri, Rahimi Esfahani et al., 2019; Pagan & Nation, 2019; Serrano, 2011; Serrano & Munoz, 2007; Stoltzfus & Sukseemuang, 2018).

3. Methodology

3.1. Participants
The study consisted of 75 participants selected from three classes among Iranian students between the ages of 16 and 19 years old. They were selected based on the outcomes of an Oxford Quick Placement Test (OQPT). The English proficiency level of the participants was pre-intermediate. The participants were male and they were native speakers of Persian. As all classes were at the same level, the researchers divided them as follows: one class as Spaced Distribution Group (SDG) (n = 25), one as Massed Distribution Group (MDG) (n = 25), and one as a Control Group (CG) (n = 25).

3.2. Instruments
The first instrument which was utilized in the present study was the OQPT. This test was used to homogenize the participants. At the beginning of the study, the researchers applied OQPT to determine the level of the students before the treatment. The paper-based version of the test has 60 multiple-choice items testing grammar, vocabulary and reading skills. It includes two parts. Part one (questions 1–40) fits for all students below advanced level and considers participants scoring 30 to 39 as pre-intermediate, and part two (questions 40–60) which is designated for high achievers. The reliability of the original test was reported as 0.9 for the 60 item test and 0.85 for the 40 item test (Geranpayeh, 2003). It took approximately 40 minutes for participants to finish the test. Based on the results of this test, 75 pre-intermediate learners were considered as the target participants of the present study. The reliability value of the test used in this research was 0.89.

The second instrument for collecting sufficient data to answer the research question was a researcher-developed vocabulary pre-test. This test measures the receptive vocabulary knowledge of the participants employing a multiple-choice format. The participants were exposed to a sentence as a context where a missing target word is to be provided. They were instructed to choose the correct option appropriate to the context from among four options. The test consisted of 50 multiple-choice items. The test was piloted on a similar group of students from another institute. The reliability value of the pre-test as calculated through the KR-21 formula was $r = .936$.

The third instrument which was used in this study to determine the impacts of the treatment on the participants’ vocabulary was a researcher-made post-test test which was used as both an
immediate and a delayed post-test. Only minor vocabulary and question order variations differentiated the immediate and delayed post-tests. It should be mentioned that the time interval between the pretest and the delayed posttest was about 2 months; thus, it was unlikely that the participants could recall the content of the pretest. Like the pretest, the immediate and delayed post-tests also included a receptive vocabulary knowledge test with 50 multiple-choice items. The validity of the immediate and delayed posttest was confirmed by those who validated the pretest; and the reliability was computed through KR-21 formula as \( r = .901 \) and \( r = .816 \), respectively.

3.3. Data collection procedure
First, the homogeneity of participants across conditions was ensured and then their L2 receptive vocabulary knowledge was measured by a vocabulary pre-test. The students in the experimental groups received the same treatment but in a different way. The new words were taught to the experimental groups through spacing instruction and massed instruction. Sixty minutes were allocated as the teaching time for all groups. In the massed class, the words were taught during a single weekly 60-minute session to the students. In the spacing class, the 60 minutes were divided into three 20-minute sessions delivered at intervals throughout the week. The 20-minute sessions were spaced with two intervening days between sessions in order to deliver three spaced sessions over a seven-day week. Each session introduced new exercises in order to maintain the participants' interest and provide a wider range of vocabulary sample sentences. The total time of instruction for both experimental groups was 60 minutes per week which included the three main parts for the treatment including an explanation stage, vocabulary exercises, along with quizzes and were accompanied by communicative activities. The whole treatment lasted 12 weeks. In terms of the time spent on task, the three groups received equal instruction time. Specifically, the three groups each received a total of 720 minutes of instruction. Over 12 weeks both groups were taught 225 words appropriate to the intermediate level. The Massed Distribution Group received 15 words in each weekly session (60 minutes each session lasted). The Spaced Distribution Group was also taught 15 words per week, 5 words in the first short session (20 minutes lasted), 5 words in the second short session (20 minutes lasted), and 5 words in the last short session (20 minutes lasted). The control group received the same amount of target words with no specific distribution of instruction.

After the first week of instruction, the first posttest was conducted. This helped the researchers to measure the immediate recall of the participants' vocabulary knowledge. After a 4-week delay, the delayed posttest was given to the participants of both experimental groups to assess the retention of the participants' vocabulary knowledge.

3.4. Data analysis
After collecting the data, the researcher used the mean and standard deviation to point out the differences between the performances of the three groups during the pretest. In order to analyze the data after the posttest quantitatively, one-way ANOVA was run to determine the differences between the three groups. SPSS (Statistical Package for Social Sciences) version 25 was used to analyze the data in the pretest, immediate posttest, and delayed post-test of the study.

4. Results
The Kolmogorov-Smirnov test of normality was run on the data obtained from the above-mentioned tests. The test indicated that the distributions of scores for the pretest, immediate, and delayed post-tests obtained from all three groups had been normal. The statistic was .14 (min) and .29 (max), \( D(25) \), \( p > 0.05 \).

Parametric test (i.e. Paired sample t-test and One-way ANOVA in this case) was computed to make further comparisons between the participating groups.

4.1. One-Way ANOVA for pretests
Table 1 indicates the descriptive statistics of the vocabulary scores on the pretest.
Table 1. Descriptive Statistics Results Comparing SDG, MG and CDG Mean Scores on the Pretests

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Mean</th>
<th>Std. Deviation</th>
<th>Std. Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>SG</td>
<td>25</td>
<td>14.04</td>
<td>1.65</td>
<td>.33</td>
</tr>
<tr>
<td>MG</td>
<td>25</td>
<td>13.44</td>
<td>1.26</td>
<td>.25</td>
</tr>
<tr>
<td>CG</td>
<td>25</td>
<td>13.92</td>
<td>1.70</td>
<td>.34</td>
</tr>
<tr>
<td>Total</td>
<td>75</td>
<td>13.80</td>
<td>1.55</td>
<td>.17</td>
</tr>
</tbody>
</table>

Note. SDG (Spaced Distribution Group); MDG (Massed Distribution Group); CG (Control Group) p > 0.05

Table 2. Results of One-Way ANOVA for Comparing SDG, MDG and CG Mean Scores on the Pretests

<table>
<thead>
<tr>
<th>SUM OF SQUARES</th>
<th>df</th>
<th>MEAN SQUARE</th>
<th>F</th>
<th>SIG.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between Groups</td>
<td>5.04</td>
<td>2</td>
<td>2.52</td>
<td>1.04</td>
</tr>
<tr>
<td>Within Groups</td>
<td>173.46</td>
<td>72</td>
<td>2.60</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>178.50</td>
<td>74</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 1 depicts the descriptive statistics of all the three groups on the pre-tests. Based on the above table, the mean of Spaced Group on the pre-test is 14.04; the mean of the Massed Group on the pre-test is 13.44 and the mean of Control Group is 13.92. As it is shown in the table, all the three groups had almost an equal performance on the pre-tests. To see whether the difference between these mean scores, and thus the three groups on the pretest, were statistically significant or not, the researcher had to examine the p value under the Sig. (2-tailed) column in the one-way ANOVA test in Table 2.

Based on the information presented in Table 3., there was not a statistically significant difference in the pretest for SDG (M = 14.04, SD = 1.65), MDG (M = 13.44, SD = 1.26), and CG (M = 13.92, SD = 11.70), p = .35 (two-tailed).

4.2. One-Way ANOVA for immediate posttests

One-way ANOVA was run between-groups to compare the posttest vocabulary scores of the SDG, MDG and CG on the immediate posttest. The descriptive results of the comparison of the three groups are displayed in Tables 3 and 4.

The mean scores of the SDG (M = 17.62), MDG (M = 15.62), and CG (M = 14.12) were different from one another on the posttest. To figure out whether the differences among these mean scores were significant, one-way ANOVA was computed for the posttest scores.

As is displayed in Table 4., there was a statistically significant difference in the posttest scores for SDG (M = 17.62, SD = 1.16), MDG (M = 15.62, SD = 1.76), and CG (M = 14.12, SD = 2.06) (p < .05). It can be claimed that the treatment affected the performance of all three groups in the post-test and further that performance of the three groups did significantly differ on the posttest. However,
to further clarify precise differences, researchers applied the post-hoc Scheffe Test. The results are presented in Table 5.

As shown in Table 5, Scheffe post-hoc analyses indicated that on the immediate posttest, both massed and spaced distribution groups scored significantly higher than the control group \((p < 0.05)\). Moreover, there was a significant difference between the spaced \((M = 17.62)\) and the massed group \((M = 15.62)\), on the immediate posttest. In fact, the spaced group performed significantly better than both the massed group and the control group. It can be concluded that in this case, spacing instruction positively affected Iranian EFL learners’ vocabulary recall.

### 4.3. Results of three groups’ delayed post-test

Another objective of the study was to find out whether using mass or spaced instruction affected the long-term retention of words. For this purpose, the delayed posttest vocabulary scores of the SDG, MDG, and CG were compared by computing one-way between-groups ANOVA. The results of the comparison of the three groups on the delayed posttest are displayed in Table 6, 7, and 8.

The mean scores of the CG \((M = 13.06)\), MDG \((M = 14.66)\), and SDG \((M = 16.32)\) were found to be different from one another on the delayed posttest.

<table>
<thead>
<tr>
<th>(I) Groups</th>
<th>(J) Groups</th>
<th>Mean Difference (I-J)</th>
<th>Std. Error</th>
<th>Sig.</th>
<th>95% Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>SDG</td>
<td>MDG</td>
<td>2.00</td>
<td>.48</td>
<td>.00</td>
<td>.79 [3.20]</td>
</tr>
<tr>
<td></td>
<td>CG</td>
<td>3.50</td>
<td>.48</td>
<td>.00</td>
<td>2.29 [4.70]</td>
</tr>
<tr>
<td>MDG</td>
<td>SDG</td>
<td>−2.00</td>
<td>.48</td>
<td>.00</td>
<td>−3.20 [−.79]</td>
</tr>
<tr>
<td></td>
<td>CG</td>
<td>1.50</td>
<td>.48</td>
<td>.01</td>
<td>2.29 [2.70]</td>
</tr>
<tr>
<td>CG</td>
<td>SDG</td>
<td>−3.50</td>
<td>.48</td>
<td>.00</td>
<td>−4.70 [−2.29]</td>
</tr>
<tr>
<td></td>
<td>MDG</td>
<td>−1.50</td>
<td>.48</td>
<td>.01</td>
<td>−2.70 [−.29]</td>
</tr>
</tbody>
</table>

Note: Statistically significant at 0.05 level

### Table 4. Results of One-Way ANOVA for Comparing SDG, MDG and CG Mean Scores on the Posttest

<table>
<thead>
<tr>
<th></th>
<th>Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between Groups</td>
<td>154.16</td>
<td>2</td>
<td>77.08</td>
<td>26.43</td>
<td>.00</td>
</tr>
<tr>
<td>Within Groups</td>
<td>209.92</td>
<td>72</td>
<td>2.91</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>364.08</td>
<td>74</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Table 5. Results of the Scheffe Post-Hoc Test for Comparing SDG, MDG, and CG Mean Scores on the Posttests

<table>
<thead>
<tr>
<th>(I) Groups</th>
<th>(J) Groups</th>
<th>Mean Difference (I-J)</th>
<th>Std. Error</th>
<th>Sig.</th>
<th>95% Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>SDG</td>
<td>MDG</td>
<td>2.00</td>
<td>.48</td>
<td>.00</td>
<td>.79 [3.20]</td>
</tr>
<tr>
<td></td>
<td>CG</td>
<td>3.50</td>
<td>.48</td>
<td>.00</td>
<td>2.29 [4.70]</td>
</tr>
<tr>
<td>MDG</td>
<td>SDG</td>
<td>−2.00</td>
<td>.48</td>
<td>.00</td>
<td>−3.20 [−.79]</td>
</tr>
<tr>
<td></td>
<td>CG</td>
<td>1.50</td>
<td>.48</td>
<td>.01</td>
<td>2.29 [2.70]</td>
</tr>
<tr>
<td>CG</td>
<td>SDG</td>
<td>−3.50</td>
<td>.48</td>
<td>.00</td>
<td>−4.70 [−2.29]</td>
</tr>
<tr>
<td></td>
<td>MDG</td>
<td>−1.50</td>
<td>.48</td>
<td>.01</td>
<td>−2.70 [−.29]</td>
</tr>
</tbody>
</table>

Note: Statistically significant at 0.05 level

### Table 6. Descriptive Statistics Results Comparing SDG, MDG, and CG Mean Scores on the Delayed Posttest

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Mean</th>
<th>Std. Deviation</th>
<th>Std. Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>SDG</td>
<td>25</td>
<td>16.32</td>
<td>.88</td>
<td>.17</td>
</tr>
<tr>
<td>MDG</td>
<td>25</td>
<td>14.66</td>
<td>2.00</td>
<td>.40</td>
</tr>
<tr>
<td>CG</td>
<td>25</td>
<td>13.06</td>
<td>2.64</td>
<td>.52</td>
</tr>
<tr>
<td>Total</td>
<td>75</td>
<td>14.68</td>
<td>2.37</td>
<td>.27</td>
</tr>
</tbody>
</table>
Table 7. Results of One-Way ANOVA for Comparing SDG, MDG, and CG Mean Scores on the Delayed Posttest

<table>
<thead>
<tr>
<th></th>
<th>Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between Groups</td>
<td>132.86</td>
<td>2</td>
<td>66.43</td>
<td>16.90</td>
<td>.00</td>
</tr>
<tr>
<td>Within Groups</td>
<td>282.96</td>
<td>72</td>
<td>3.93</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>415.82</td>
<td>74</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 8. Results of the Scheffe Post-Hoc Test for Comparing SDG, MDG, and CG Mean Scores on the Delayed Posttest

<table>
<thead>
<tr>
<th>(J) Groups</th>
<th>(J) Groups</th>
<th>Mean Difference (I-J)</th>
<th>Std. Error</th>
<th>Sig.</th>
<th>95% Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>SDG</td>
<td>MDG</td>
<td>1.66</td>
<td>.56</td>
<td>.01</td>
<td>.25 (3.06)</td>
</tr>
<tr>
<td></td>
<td>CG</td>
<td>3.26</td>
<td>.56</td>
<td>.00</td>
<td>1.85 (4.66)</td>
</tr>
<tr>
<td>MDG</td>
<td>SDG</td>
<td>-1.66</td>
<td>.56</td>
<td>.01</td>
<td>-3.06 (-.25)</td>
</tr>
<tr>
<td></td>
<td>CG</td>
<td>1.60</td>
<td>.56</td>
<td>.02</td>
<td>.19 (3.00)</td>
</tr>
<tr>
<td>CG</td>
<td>SDG</td>
<td>-3.26</td>
<td>.56</td>
<td>.00</td>
<td>-4.66 (-1.85)</td>
</tr>
<tr>
<td></td>
<td>MG</td>
<td>-1.60</td>
<td>.56</td>
<td>.02</td>
<td>-3.00 (-.19)</td>
</tr>
</tbody>
</table>

Note: Statistically significant at 0.05 level

As is could be observed in Table 7, there was a statistically significant difference in the delayed posttest scores for CG (M = 13.06, SD = 2.64), MDG (M = 14.66, SD = 2.00), and SDG (M = 16.32, SD = .88) (p < 0.05) on the delayed posttest of vocabulary. This indicates that the three groups significantly differed in terms of vocabulary retention after a lapse of time. Pair-wise comparisons of the groups (in Table 8) reveals which two groups were significantly different on the delayed posttest.

Scheffe post-hoc analyses also demonstrated that on the delayed posttest, the massed and spaced groups scored significantly higher than the control group (p < 0.05). Furthermore, on the delayed posttest, the difference between SDG (M = 16.32) and MDG (M = 14.66) was statistically significant (p < 0.05). This revealed that the spaced distribution group scored significantly higher than the massed distribution group.

These findings suggest that both spaced and massed distribution instruction could have a significant effect on the long-term retention of words. Yet, between spaced and massed distribution groups, the spaced proved to be more effective.

5. Discussion and conclusion

This study investigated whether distributed instruction could enhance EFL learners’ vocabulary retention and recall and if that were to be the case, which distributed instruction (i.e., spaced vs massed) would have a more significant effect on recall and retention of EFL vocabulary. After collecting and analyzing the data, the results indicated that the spaced group improved on their immediate post-test compared to their pre-test. In fact, the results revealed a significant difference between the massed and spaced distribution groups on the participants’ recall of the target words in favor of the spaced distribution group. The spaced distribution group made more progress on the immediate posttest. Furthermore, when measured on the 4-week-delayed posttest (i.e., retention test), the results revealed that the spaced group also significantly outperformed the
massed group on the delayed posttest. The findings imply that spaced distribution instruction may positively progress EFL learners’ long-term vocabulary learning.

The results suggest that the more spaced two items are, the more likely it is that they will be interpreted differently in the memory of the learners, in accordance with the theoretical arguments of the present study. This inconsistency in memory representation, which is enabled by the different contexts in which spaced items are demonstrated, gives the learner a more retrievable indicator (Mashhadi et al., 2017). This is not the case when the massed distribution is practiced. As a result, one could argue that remembering is more actively facilitated in spaced distribution instruction. Moreover, the additional processing mechanisms entailed when teaching over intervals simplifies learning and retention by parsing the information into more manageable units (through varied context and repeated exposure). In contrast, when participants are presented with multiple items concomitantly, as is the case with massed distribution, they might not dedicate as much consideration to these items individually as they would have in the spacing context.

The finding is consistent with previous studies in cognitive psychology (Namaziantost, Rahimi Esfahani et al., 2018; Pavlik & Anderson, 2005; Seabrook et al., 2005) which confirmed the effect of spaced distribution instruction in different domains of learning. Moreover, the result also corroborates some previous studies (e.g., Miles, 2014; Rohrer & Pashler, 2007; Smolen et al., 2016) showing that the spaced distribution instruction improved foreign language learning, and specifically vocabulary learning (Goossens et al., 2012; Kornell, 2009; Sobel et al., 2011).

These results are in contrast with previous studies (Collins & White, 2011; Lee & Choe, 2014; Miles, 2014; Snoder, 2017), which found no clear advantage of spaced conditions over massed conditions on immediate posttests.

To sum up, our findings complement earlier studies by that show the beneficial effects of spacing in vocabulary learning, and seriously question the belief that learners benefit from cramming more than spacing. It can be concluded that learning through spaced distribution instruction gives the learners a better opportunity to encode and retain new knowledge gained from instruction until the next chance for review occurs. Utilizing spacing instruction permitted students more time to rest, gave them more time to think, and provided them with more opportunities to study; this extra processing time, neural recovery during spacing (Smolen et al., 2016) and the dispersion of input exposure over different sessions may explain the students’ notably better L2 vocabulary development. Moreover, several spaced practice sessions seem to better reinforce memory over time. It is in fact highly likely that practice strengthens the development of declarative knowledge. Structuring practice through spaced distribution at regular sessions and spaced instruction enhances learning outcomes for simple tasks and is useful during the initial stage of declarative learning (Suzuki & DeKeyser, 2017). It should be noted that the mechanism for the observed benefit is debatable. While both processing speed and accuracy vary among learners, breaking declarative knowledge into smaller units delivered at an appropriate pace should be beneficial to all learners. One would expect these benefits to most impact novice and intermediate learners. These learners are usually also the ones with the greatest room for improvement in terms of testing performance. It should also be noted that the current study tested for vocabulary recognition is the recall task. Recognition is reflective of superficial learning but is not as indicative of learning as a pure recall test with no cueing. Further research addressing the mechanism for the noted benefits of spaced instruction as well as its applicability to more complex learning tasks would be beneficial. Clarification would allow for more targeted and effective curriculum design.

Based on the findings, some implications can be proposed for EFL practitioners. First of all, homework assignments could be used to re-expose learners to the content they have already mastered. If class time and chances for review are limited, it may be particularly appropriate. Second, teachers may offer combined tests and quizzes. Cumulative tests and quizzes provide a good incentive for learners to revisit material on their own, requiring active recall and extra detail
in order to strengthen new knowledge. Third, teachers should empower and allow learners to exert conscious control over their preferred metacognitive strategies. This can be achieved using specific self-regulatory techniques, such as setting up a regular timetable of instructions. Third, to increase the amount of their studying, learners should spread their self-study sessions out in time. It might be a good idea for teachers to plan learning tasks in the classroom according to a spread schedule in order to improve the success of the assessments among the learners. Last but not least, as spacing has a positive impact on the long-term memory of learners, syllabus designers and curriculum creators can take the opportunity to determine when and where to repeat a word in a series of teaching materials. We should integrate spacing into the creation of learning curricula and educational materials as an instructional technique.

The study suffered from limitations. The sampling approach used in this analysis was based on the participants' availability. Similar studies can yield more generalizable findings with a more representative sample. Another drawback is that this study was conducted only with pre-intermediate learners; the next studies will work with high school students, intermediate and also elementary learners in order to get more robust findings. A small population of Iranian EFL learners also attended the current study. In comparison, care for both classes was primarily limited to explicit instruction and examination. Therefore, it could be helpful to add communicative methods to language teaching (i.e., emphasis on the form) or unconscious learning. Furthermore, only male students participated in this study since the researcher did not access to female learners. The current research has only covered vocabulary learning, so the forthcoming studies can investigate other skills and sub-skills as language learners with the younger groups. Finally, the current study merely looked at retaining the words in more than 15 sessions. Future studies will repeat the analysis using longer time periods and several post-tests with delays. These data can provide a deeper insight into the word-learning processes and the spacing learning of young learners.

Funding
The authors received no direct funding for this research.

Author details
Ehsan Namaziondost
E-mail: e.namaz75@yahoo.com
ORCID ID: http://orcid.org/0000-0002-8393-2537
Murad Hassan Mohammed Sawalmeh
E-mail: msawalmehd@edu.edu.om
Masoumeh Izadpanah Soltanabadi
E-mail: sizadpanho27@gmail.com
1 Department of English, Faculty of Humanities, Shahrekord Branch, Islamic Azad University, Shahrekord, Iran.
2 Linguistics and Translation, Dhofar University, Salalah, Oman.
3 Department of English Language, Payam Noor University, Tehran, Iran.

Citation information
Cite this article as: The effects of spaced versus massed distribution instruction on EFL learners' vocabulary recall and retention, Ehsan Namaziondost, Murad Hassan Mohammed Sawalmeh & Masoumeh Izadpanah Soltanabadi, Cogent Education (2020), 7: 1792261.

References


