Research in English Language Pedagogy (2022) 10(3): 412-436

 DOI: 10.30486/RELP.2021.1943193.1312

Original Article

A Comparative Study of the Effects of Equal and Expanding Spacing Schedules on L2 Receptive and Productive Vocabulary Retrieval

Roya Khoii[1,*], *Kobra Fallah Abed*[1]

[1]Department of Foreign Languages, Islamic Azad University, Tehran North Branch, Iran

Submission date: 23 October, 2021 Acceptance date: 19 December, 2021

Abstract

Spaced repetition seems to play a significant role in the learning and retention of words in foreign language settings. This experimental study investigated the impact of expanding and equal spacing conditions on receptive and productive English vocabulary retrieval. The participants consisted of 63 pre-intermediate EFL learners in three intact classes randomly assigned to one control and two experimental groups. A Vocabulary Knowledge Scale (VKS) was used to select 20 unknown word pairs (each word and its synonym) for teaching purposes and check the effects of the treatments at the end of the study. The word pairs were later divided into two 10-item A and B sets because of the counterbalanced design of the study. The first experimental group studied set A under an equally spaced condition (2-2-2) and set B under an expanding spaced condition (0-1-5) with the same absolute spacing value. The second experimental group studied them in the reverse order, thus counterbalancing the effect of encounter order. The control group studied the target items under the no-spacing or massed condition (0-0-0). The statistical analyses of the three groups' posttest scores indicated that all the three groups had improved their receptive and productive vocabulary retrieval. However, an equally spaced schedule had led to a significantly higher mean score only on the receptive items of the posttest and, therefore, better vocabulary retrieval. Therefore, it is suggested that EFL teachers consider spacing in general and the equal spacing condition in particular as important elements in facilitating both receptive and productive vocabulary learning.

Keywords: EFL Learners, Equal Spacing, Expanding Spacing, Productive Vocabulary, Receptive Vocabulary, Spacing Effect, Vocabulary Retrieval

Corresponding Author's E- mail: roya_kh@yahoo.com

1. Introduction

Both practitioners and researchers commonly suggest that vocabulary acquisition is essential for desirable performance in receptive and productive tasks (Lenko-Szymanska, 2019; Naismith & Juffs, 2021). However, the most influential means of attaining a sizeable vocabulary is still a mystery to practitioners in the field of second language pedagogy, and thus teachers should identify and employ the most effective strategies to help the learners in this regard. The number of repetitions necessary for acquiring a list of words is also of prime importance in this process, and different studies have found significant correlations between exposure frequency and vocabulary retention (Candry et al., 2020; Webb, 2007). Frequency plays a key role in L2 learning because the more frequently the learners encounter a linguistic item, the more solid the linguistic representation of that item will become. Hence, there will be a higher probability for that item to be incorporated into a language system which affords to accelerate the activation and processing of the target items (Crossley et al., 2019).

While vocabulary acquisition certainly depends on repeated exposures to new words, frequency by itself might not result in higher acquisition rates. Moreover, the question of spacing between repetitions remains to be answered. According to Nation (2001), spacing the act of repetition improves retention since the most forgetting happens soon after the first study of new data. He also emphasizes that, based on the findings of memory research, most forgetting happens right after the first encounter with the word. In other words, the pace of forgetting is slower when it comes to older data, which indicates that repetition should occur right after the first encounter with a lexical item, with later exposures spaced further apart.

It is emphasized that here repetition refers to the number of encounters, and the format of the encounter could differ from one situation to another. For example, Nakata (2015) used a paired-associate format to explore the effects of expanding and equal spacing on vocabulary learning. In typical paired-associate learning (PAL), learners are presented with a target L2 word paired with its L1 meaning. In the second encounter, they practice retrieval by trying to recall its meaning. This stage is followed by feedback, during which learners encounter the same word and its meaning (Karpicke & Bauernschmidt, 2011). The present study focuses on PAL because it is not only common (Nakata, 2015) but also effective for L2 vocabulary learning. Previous studies indicate that one of the

factors that could affect PAL effects may be the direction (receptive or productive) of retrieval (Nation, 2001), which is also tackled in this study.

2. Literature Review

2.1. Vocabulary Acquisition and Memory

As a constituent of the human mind, memory seems to be in close alliance with L2 learning. Smith (2017) posits that the best available facility during online processing is working memory (WM), which is a temporary storage facility prone to strict limits of capacity. Similarly, Ruiz, Rebuschat, and Meurers (2019) define working memory as an attentional system with a limited capacity that reinforces complex cognitive processing. Any theory of learning comprises memory as a key component since learning would be transitory and ineffectual without the ability to conserve any alterations in knowledge (Smith, 2017). Working memory is also vital for discerning how individuals can differentially make the best use of their prevailing resources in one or more languages. Ruiz et al. (2019) also reported that having a vaster working memory capacity contributes to handling attention-demanding learning tasks more successfully, as they require the learners to process meaning, form, and the use of linguistic forms simultaneously, similar to what they do in form-focused instruction.

The majority of studies on L2 word processing rely on Baddeley's multi-component model of working memory (2007) as it has proved to be the most influential theoretical framework utilized in investigating the relationship between working memory and L2 learning (Ruiz et al., 2019). According to Baddeley (2017), working memory is a system consisting of two storage subsystems, the phonological loop, and the visuospatial sketchpad, used for temporary keeping and processing of both verbal and visual-spatial data, respectively; a component responsible for attention control and regulation (the central executive), and an episodic buffer linking the storage subcomponents and long-term memory. In this model, a central executive directs attention to a lexical item that must be processed. If the learner attends to the word, the information enters WM, where it is controlled via a visuospatial sketchpad and a phonological loop. The sketchpad rehearses noticed input at a subvocal level, which is particularly important for lexical processing. The reason is that words need to be rehearsed so that they receive the amount of attention necessary for being sent to long-term memory.

For efficient vocabulary acquisition to take place, O'Brien et al. (2006) maintain that language learners need to hold and operate upon phonological sequences in WM to process input flawlessly and add vocabulary to their word repertoire. The phonological loop connects sequences of sounds by rehearsing some of them and storing some others temporarily in the loop and retrieving them when required. Later the identified word sequences will be dispatched to long-term memory. Nevertheless, to reinforce the bonds of such sequences, the words must be repeatedly encountered and processed in the PL (Schuetze & Weimer-Stuckmann, 2011).

2.2. Receptive and Productive Vocabulary Knowledge

Vocabulary acquisition is not instantaneous and refers to two degrees of word knowledge: receptive (passive) and productive (active). Schmitt (2000) states that learners achieve receptive word knowledge prior to productive word knowledge, which highlights the significance of the amount and frequency of exposure to input. In the same vein, Lenko-Szymanska (2019) divides lexical knowledge into receptive and productive knowledge, that is, words' the meaning of which can be retrieved when reading or listening compared to words that speak of the desired meaning when writing or speaking. This division has ecological and statistical validity as it is a common classroom phenomenon for language learners to demonstrate the understanding of lexical items which they cannot produce in speaking or writing (Schmitt, 2014). It is more challenging for learners to demonstrate productive than receptive knowledge of words because receptive knowledge is acquired faster and is more easily retained (Naismith & Juffs, 2021), while productive knowledge is more complicated and demands a series of mental processes before actual production. Undoubtedly, productive knowledge is of prime importance for both speaking and writing. However, with many lexical development studies conducted in strictly controlled contexts rather than in more natural settings, more investigations are required in this regard (Juffs, 2019). As a result, there is still a gap in the field of L2 learning regarding understanding how L2 teachers can help to expedite the process of developing productive knowledge of words.

An important factor in learning both receptive and productive vocabulary is repeated retrieval (Candry et al., 2020; Nakata, 2016, Roediger & Karpicke, 2010). Lexical retrieval refers to accessing lexical representations of words with their specified syntactic properties

(Ramanujan & Weekes, 2020). In other words, retrieval refers to "the process of accessing information about L2 words from memory" (Nakata, 2016, p.2). This term, which is commonly used in the field of the spacing effect, distributed learning effect, and lag effect, refers to students' performance on repeated language tests within or between class sessions (Nakata, 2015, 2016; Nakata & Elgort, 2020; Rogers, 2021). Productive retrieval occurs when a learner wishes to convey meaning and has to recall the appropriate spoken or written form of the word to produce it (Naismith & Juffs, 2021). Receptive and productive retrieval will be more effective for learning if they enjoy variety in successive noticing, that is, varied meetings or usage in multiple contexts (Naismith & Juffs, 2021). It is noted that in the process of input spacing, each exposure to input requires retrieval, which later might lead to retention in case it is practiced in different contexts (Nation, 2020). The next section discusses various patterns of spacing encounters with input and highlights some of the most recent findings in this regard.

2.3. Spaced Repetitions

Research findings indicate that there is a significant correlation between repetition frequency and word learning (Nation, 2014). However, there is still no consensus regarding a particular minimum number of repetitions required for the learning of vocabulary items. Understandably, the higher the repetition frequency and the deeper the quality of the repetitions, the more likely it will be for L2 learners to pick up the words. Nevertheless, researchers stipulate that the spacing or interval between repetitions plays a significant role in promoting retention (Celce-Murcia et al., 2014). Elgort (2011) argues that retention rates under form-focused learning are, on average, much higher than under incidental conditions. Exposure frequency in an intentional forms-focused context affects vocabulary learning by not only increasing the likelihood of a word to be noticed and processed but also strengthening the association between the target words and the mental processing of the data. Therefore, a word with several repeated encounters will probably receive more attention from learners, will evoke more cognitive processing, will be stored more easily, and will be activated and called upon later (Zhu, 2015).

The effects of input spacing, defined as changing the duration of time intervals between multiple learning episodes, have been widely explored in the domain of psychological branches of knowledge (Rogers & Cheung, 2018). Two terms that are

closely associated with this process include 'spacing effect' and 'lag effects'. The first refers to distributed learning conditions with some learning events distributed over a longer period, which are said to result in more efficient retention and learning compared to massed conditions, where there is no time interval between learning episodes (Rogers, 2017; Rohrer, 2015). Researchers often use the term 'lag effect' when they compare two different schedules with each other, for example, one schedule with learning episodes separated by a one-day time interval and the other with learning episodes distributed over one-week time intervals (Rogers & Cheung, 2018). The term "distributed practice effect" is sometimes employed as an umbrella term referring to both lag effects and spacing (Nakata & Elgort, 2020).

The distributed practice or spacing effect refers to a memory advantage that emerges when learners are exposed to the target material over multiple separate episodes rather than a single massed condition (Sobel et al., 2011). The spacing effect is based on the assumption that learning includes retrieval, whereby learners are asked to recall L2 words from memory because research findings indicate that retrieval increases learning (Barcroft, 2007; Nakata & Webb, 2015). Retrieval can be receptive in case learners are required to remember the meaning of an L2 word in reading or listening or productive if they try to use it in speaking or writing. Spaced retrieval practice, which refers to inter-study intervals, can be either equal or expanding (Nakata, 2015).

When the learners try to recall a piece of stored data, they initially need to retrieve it from memory. Previous studies have demonstrated that the spacing schedule of word retrieval exercises has a positive effect upon long-term learning and retention of L2 vocabulary, and that repeated information is generally recalled more successfully when it is offered in spaced schedules than in a massed schedule (Nakata, 2015; Nakata & Elgort, 2020; Storm et al., 2010; Thalheimer, 2006). According to Thalheimer (2006), when repetitions are spaced, the information stored in memory is less vulnerable to forget. He also suggests that extended spacing conditions are more likely to result in more long-term recollection of data than shorter spacing conditions.

Since the timing of learning events and multiple exposures influence retention, several scholars (Cepeda et al., 2008) have proposed some techniques for varying RI (retention interval) and spacing length to enhance learning. The spacing effect has been practiced on many different types of tests, such as free recall, word completion, and

recognition with many different forms of materials, such as words and pictures, and while studying learners at different ages, for example, young learners and adults. Based on the Encoding Variability Theory, distributed practice leads to better recall since during each practice episode the learners encode the input in different ways and, therefore, will employ different retrieval cues (Serrano, 2011), and according to the Deficient Processing theory, some lag or interval is necessary for the information to be more efficiently retrieved later (Bui et al., 2019).

In one of the earlier studies of four spacing conditions with young adults, Landauer and Bjork (1978) tested the recall ability of 12 first and last names. In the massed condition, they employed a 0-0-0 schedule during which the test trials occurred consecutively with zero intervention between each retrieval effort. In the equal spacing condition, the study and test trials were distributed over an equal number of interventions (e.g., 4–4–4), while in the expanded condition they were separated by an expanding number of interventions (e.g., 1–3–9). However, in the contracting schedule, the study and test trials were distributed over a decreasing number of interventions (e.g., 9–3–1). Their findings demonstrated that the superiority of expanded retrieval over the equally spaced retrieval at the end of the treatment and the ultimate retrieval task.

Following a later study on the impact of input spacing on second/foreign vocabulary learning, Kang et al. (2014) reported that expanded retrieval spacing and equal-spaced practice resulted in similar retention of words at the end of the two-month experiment. Nevertheless, they found that the participants' average retention rate was higher in the course of the expanded trials than the whole experiment. Many researchers have often referred to the expanding condition as the most potent relative spacing condition. For instance, based on his study of the effects of equal and expanding conditions on vocabulary acquisition, Nakata (2015) reported a significant advantage for the gradual increase of interval schedules. However, Cull (2000) maintains that when average spacing is controlled, the advantages of expanded over equally spaced retrieval may be limited and, in certain cases, the outcomes of the two conditions might not be significantly different from each other. He conducted four experiments and used conditions with and without feedback after each test. He did not find any significant positive effect for expanding spacing, and in some conditions, he observed that equal spacing resulted in more efficient long-term retention following a 3-or 8-day delay.

In their study of the impact of input spacing on L2 vocabulary acquisition of primary school children, Rogers and Cheung (2018) concluded that contrary to the findings of earlier laboratory-controlled studies, shorter lags or intervals between learning trials could contribute to better word retention, as measured after a 28-day delay. However, because most of the investigations about spacing and lag effect have been carried out in laboratory settings, more research is still required to determine the optimal degree of spacing in real-world learning contexts (Rogers & Cheung, 2018). Bui et al. (2019) further stated that massed conditions might be more useful than distributed practice for L2 performance, at least regarding speaking fluency and complexity.

Vocabulary learning is an evergreen field for further studies regarding the employment of various old and new techniques, tasks, and practice types that could assist teachers in helping students enrich their word knowledge. In the Iranian context, the use of repetition-based practices is quite common in vocabulary instruction and learning. However, it seems that the decision to use such practices is to some extent uninformed in terms of the spacing of exposures to vocabulary items, frequency of exposures, and quality of exposures. Moreover, there is an undeniable lag in Iran regarding studies on the efficacy of input spacing in SLA contexts. In one of the few relevant studies that the researchers located in this regard, the authors concluded that spaced distribution was more beneficial than massed distribution in the recall and retention of vocabulary items (Namaziandost et al., 2020). Given the conflicting results of previous studies on input spacing in both instructed and naturalistic settings (Rogers, 2021), the rarity of SLA studies comparing equal spacing with expanding spacing conditions, and the growing interest in distributed or spaced practice effects in SLA (Suzuki et al., 2019), it appears that more research in this field could yield some fruitful results for language teachers. Thus, this study was carried out to provide an answer to the following questions:

RQ1: Is there a statistically significant difference between the effects of equal and expanding spacing conditions on EFL learners' receptive vocabulary retrieval?

RQ2: Is there a statistically significant difference between the effects of equal and expanding spacing conditions on EFL learners' productive vocabulary retrieval?

3. Methodology

3.1. Design and Context of the Study

This study employed a counterbalanced non-equivalent quasi-experimental design to investigate the effects of spacing on vocabulary retrieval. Counterbalancing was employed to control the effect of the order of exposure to different spacing conditions. The independent variables were equal, expanding, and massed spacing conditions and the dependent variables were receptive and productive types of vocabulary retrieval, which were tested through the Vocabulary Knowledge Scale (VKS) test (Wesche & Paribakht, 1996) before and after the treatment. The gender, language proficiency, and age of the participants were the moderator variables. This study was conducted in 2017 in a girls' language school in Tehran. The design of the present research is illustrated below:

Table 1.

Design of the Study

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
EX I	PET	_	O1	AB	B	_	AB	_	_	A	_	_	AB	O2		
EX II	PET	_	O3	AB	A	_	AB	_	_	B	_	_	AB	O4		
C	PET	_	O5	AB	AB	AB	AB	O6								

O1, O3, O5 = Pretests — O2, O4, O6= Posttests

1-16 = Sessions — A/B= Sets of target items A & B

Ex1 & Ex2= Experimental groups — C= Control group

PET= Preliminary English Test

3.2. Participants

The participants of this study consisted of 63 (selected out of 70) Iranian female pre-intermediate EFL learners (13-17 years old) studying English in three intact B1 classes at a girl' language school in Tehran. The classes were randomly assigned to one control (C) and two experimental (EX I, EX II) groups with 24, 21, and 25 students in each, respectively. The experimental groups studied the target items under both the equal and expanding spacing conditions, while the control group studied them under a massed

learning or no spacing condition. All the students studied the book *Solutions: Preintermediate* (2012), units 9 and 10, in a 16-session semester.

Table 2.

Demographic Background of the Participants

Initial Number of the Participants	70
Final Number of the Participants	63
Gender	Female
Age	13-17
Native Language	Persian
Language Level	B1
Educational Level	High School Students
Academic Years	2017-2018

3.3. Instruments

The required data for the study were collected using the following tests:

1. A complete Preliminary English Test (PET) adopted from the Cambridge Assessment English website was used to select participants who were homogenous in terms of English proficiency before the treatment. This test is aimed at B1 level EFL learners based on the classification of CEFR. The Chronbach α reliabilities of the reading and listening sections of this test were 0.82 and 0.89, respectively. The interrater reliabilities of the speaking and writing sections were 0.89 and 0.87, respectively. Of the 70 students who took the test, 63 in three intact classes with scores within 1.5 standard deviations above and below the mean were selected as the main participants of the study.
2. A 35-item Vocabulary Knowledge Scale (VKS) pre-test made by the researchers was administered to choose 20-word pairs (each word with its synonym) which were completely unfamiliar to the students.
3. A VKS post-test identical with the pretest was employed to measure the effects of the treatment. It is noted that the participants through a VKS scale had no access to any clue regarding the meanings of the words. Besides, the sameness of the pretest and posttest posed no threat here since the students were

repeatedly exposed to the same input over different spacing conditions. Each exposure was supposed to reinforce retrieval.

In the VKS test, each item is followed by a 5-option scale to evaluate the students' productive and receptive word knowledge. The VKS test eliminates the danger of guessing the meaning of words as it emphasizes 'what the students already know' instead of 'what they might not know' by allowing them to show their partial knowledge of a word (Schmitt, 2000, p.175). It is noted that the VKS used in this study was quite different from the one developed by Wesche and Paribakht (1996) because it only considered the participants' knowledge of synonyms for each given item, and the students were not required to provide L1 translations of the target items. Since the pretest and the posttest were the same, this strategy was employed to prevent unwanted consciousness-raising on the part of the students.

The book *Oxford Word Skills, Intermediate Book* (Gain & Redman, 2008) was used to select the target word pairs as recommended and commonly used by the language school where this study was conducted. It is emphasized that the researchers deliberately chose items which the students had not studied the previous semesters. To identify the target words of the study, a 5-level VKS pre-test consisting of 35 vocabulary items was given to the three groups. Based on the scoring scale, a score of 1 or 2 was given to levels 1 and 2, a score of 3 to satisfactory synonyms, a score of 4 to using the word, however incorrect in terms of accuracy, in an appropriate context, and a score of 5 to use the target item accurately in an appropriate context. The students' responses to Levels 3 and 4 options, which indicated that they were able to recognize the words and provide a synonym for each, though imperfectly, were used to measure their receptive knowledge of the target words. Finally, their responses to level 5 options, whereby they employed the target words in a correct sentence, were used to measure their productive knowledge. After identifying the unfamiliar 20 word-pairs (those receiving a score of 1 or 2), the researchers divided them into two A and B sets each containing 10 items.

3.4. Data Collection Procedure

Initially, a PET was given to all the participants to check their homogeneity in terms of English language knowledge. After the statistical analysis of their scores, 63 of them were chosen as the ultimate participants of the study. They were in three intact classes

randomly assigned to the experimental and control conditions. Then, they received the VKS pretest so that the researchers could identify the items (20) which were completely unfamiliar to all of them.

The whole research project took place over a 16-session semester during regular class hours (90 minutes each session) in eight weeks. The same teacher taught all the three groups to control the effect of the teacher factor. Twenty teacher-made flashcards were designed and used to expose the learners to the word pairs, and each exposure lasted 40 minutes. Each flashcard had a target item on one side and its synonym on the other side. The treatment for the equal and expanding experimental groups began from the fourth session onwards and continued until the 14th session. However, the treatment for the control group (massed condition) finished in the eighth session before administering the post-test the following session. It is emphasized that the number of exposures to the target words was the same for this group, and it was only the spacing condition that was different. Therefore, all three groups were exposed to the input four times over diverse spacing conditions.

The equal and expanding experimental groups were exposed to both A and B sets under two different spacing conditions. EX I studied set A under an equal spacing condition (2-2-2) and set B under an expanding one (0-1-5). In the equally spaced schedule, the students had further retrieval practice in sessions 7, 10, and 13, while in the expanding schedule they did so in the 5th, 7th, and 13th sessions of the experiment. The same procedure was followed for the second experimental group but in the reverse direction, set A under the expanding and set B under equally spaced schedules, to counterbalance the effects of the order of encounter with spacing conditions. The absolute value or the total amount of spacing was 6 in both conditions. The third group, however, was exposed to both word sets under massed or no-spacing conditions. Here, the participants studied all the target items four times in the 4th, 5th, 6th, and 7th sessions based on a 0-0-0 schedule. The three groups studied the words using flashcards.

As mentioned previously, all three groups studied each item four times during the treatment. In the first two exposures, the students were introduced to the new word and its synonym through flashcards (e.g., the word *detest* and its synonym *hate*), while in the second two exposures, they practiced productive retrieval by trying to make some

sentences orally using the new word. The errors at this stage were treated through resorting to self-correction, peer correction, and teacher feedback in all the groups.

Finally, a VKS post-test identical with the pretest was given to the three groups to examine their progress in vocabulary knowledge at the end of the experiment. The scoring process was also the same as the pre-test, and the main focus was only on the 20 target items of the test.

3.5. Data Collection Procedure

After collecting the required data, they were fed into SPSS. Descriptive statistics were computed for all the tests and normality checks were run. The Cronbach's α formula was used to compute the reliability of the reading and listening sections of PET. The Pearson Product Moment Formula was used to compute the inter-rater reliability of the speaking and writing sections of PET. The pretest measured lack of knowledge; hence, no statistical checks were required. However, because the posttest was a criterion-referenced achievement test, the dependability of the scores was computed using the Threshold Loss Agreement Formula. To answer the first and second research questions, an ANOVA, followed by a post hoc Tukey test, and two independent samples t-tests were run. Finally, the magnitude of mean difference was computed using the eta squared method.

4. Results

4.1. Results of the Homogeneity Test

The descriptive statistics of the three groups' PET scores were given in Table 3.

Table 3.

Descriptive Statistics of PET

Groups	N	Mean	Std. Error.Mean	SD	Skewness Statistic	Skewness Std.Error
EX I	21	62.90	1.042	4.77	0.422	0.501
EX II	25	65.72	2.52	12.61	-0.002	0.464
C	24	64.42	1.642	8.04	0.56	0.472

The results of skewness analysis confirmed the normality of score distributions. Then, a one-way ANOVA was performed to establish the homogeneity of the groups concerning English proficiency (Table 4).

Table 4.

ANOVA for PET

	Sum of Squares	df	Mean Square	F	Sig
Between Groups	90.46	2	45.23	0.526	0.594
Within Groups	5764.683	67	86.04		

With F (2, 67) = 0.526, p = 0.594 > 0.05 (two-tailed), no statistically significant difference was detected between the means of the three groups on PET.

4.2. Results of the Pretest

As mentioned before, a VKS pretest was administered to select 20 unfamiliar word pairs before the treatment. Based on the scoring scale of the VKS, scores of 1 and 2 indicated a lack of vocabulary knowledge. Thus, the words which had been scored between 20-40 for each group were chosen as the target words of the study. The descriptive statistics of the 20 target items on the pre-test were given in Table 5. It is worth noting that the scores of seven absentees (one in EXI, and five in EX II, and one in C) during the treatment, were excluded from the next analyses.

Table 5.

Descriptive Statistics of the Target Items of VKS Pretest

Groups	N	Mean	Std. Error Mean	SD	Skewness Statistic	Std. Error
EX I	20	26	0.82	3.67	0.454	0.481
EX II	20	27.45	0.84	3.76	0.136	0.512
C	23	26.52	0.62	2.98	0.905	0.512

The results of the skewness analysis confirmed the normality of score distributions within each group.

4.3. Results of the Post-test

After four encounters with the target items by each group, the same VKS pre-test was given to the three groups to check the changes in their word knowledge. The scoring process for the posttest was similar to that of the pretest. As mentioned previously, 10 target items in each experimental group (EX I and EX II) were studied under the equal spacing condition (2-2-2), and the other 10 items were studied under the expanding condition (0-1-5), thus the number of participants in each group was multiplied by the number of the target word sets (i.e. 2) for the equal and expanding schedules in the following analyses. Table 6 presents the descriptive statistics of the posttest scores of the three spacing conditions. The dependability of the posttest, which was calculated through the Threshold Loss Agreement Formula, was equal to 0.82, which was desirable.

Table 6.

Descriptive Statistics of VKS Posttest of the Three Spacing Conditions

Spacing Conditions	n	Mean	Std. Error Mean	SD	Skewness Statistic	Skewness Std.Error
Massed	46	33.3696	1.123373	8.36755	-0.307	0.350
Equal	40	39.60	1.51268	9.56704	-0.139	0.374
Expanding	40	34.325	1.29168	8.16932	-0.053	0.374

*. n = number of participants × 2 (sets of words studied under each condition)

The mean of the equal spacing group (*M* = 39.60, *SD* = 9.567) was larger than that of the massed spacing group (*M* =33.369, *SD* = 8.367) and the expanding spacing group *(M = 34.325, SD = 8.169).* However, further statistical analyses were required to see whether the differences were statistically significant. Therefore, an ANOVA was run to compare the participants' receptive and productive word gain mean scores in the three spacing conditions after the treatment (Table 7).

Table 7.

ANOVA for VKS Post-test of the Three Spacing Conditions

	Sum of Squares	df	Mean Square	F	Sig
Between Groups	933.543	2	466.771	6.158	0.003
Within Groups	9323.092	123	75.797		

With F (2, 123) = 6.158, p = 0.003< 0.05 (two-tailed), it was decided that the learners' mean scores in the three spacing conditions were significantly different from each other; hence, a post hoc Tukey test was run to identify the exact location of the differences (Table 8).

Table 8.

Post Hoc Test for the Three Spacing Conditions

(I) groups	(J) groups	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Upper	Lower
Massed	Equal	-6.2304*	1.88	0.003	-10.6958	-1.7650
	Expanding	-0.9554	1.88	0.868	-5.4208	3.5100
Equal	Massed	6.2304*	1.88	0.003	1.7650	10.6958
	Expanding	5.2750*	1.95	0.021	0.6565	9.8935
Expanding	Massed	0.9554	1.88	0.868	-3.515	0.4208
	Equal	-5.2750*	1.95	0.021	-9.8935	-0.6565

*. α = 0.05

Post-hoc comparisons demonstrated that the study of words under equally spaced conditions (M=39.60, 2-2-2 condition) had significantly contributed to vocabulary learning compared to the massed and expanding interval schedules.

Following this stage, two independent-samples *t-tests* were run to compare the two spaced exposure groups' receptive and productive word gain mean scores under the two equally spaced and expanding schedules (Tables 9 and 10). It is noted that the massed condition was not included because it did not involve any lag or space between every two exposures.

Table 9.

Independent Samples T-test for Productive Word Gain on VKS Post-test in the Spaced Conditions

	Levene's Test for Equality of Variances		t-test for Equality of Means					
	F	Sig	T	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the

								Difference	
								Lower	Upper
Equal Variances assumed	1.126	.292	-0.275	78	.784	-1.00	3.6357	-8.2382	6.2382
Equal variances not assumed			-0.275	77.27	.784	-1.00	3.6357	-8.2393	6.2393

After establishing the equality of variances of the two conditions on the posttest, with t (78) = -0.275, p = 0.784 (two-tailed), no statistically significant difference was observed between the means of the participants' productive word gain under the two spacing conditions.

Table 10.

Independent Samples T-test for Receptive Word Gain on VKS Post-test in the Spaced Conditions

	Levene's Test for Equality of Variances		t-test for Equality of Means						
	F	Sig	T	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
								Lower	Upper
Equal Variances assumed	.139	.710	2.147	78	.035	5.50	2.5615	.4004	10.5996
Equal variances not assumed			2.147	77.987	.035	5.50	2.5615	.4004	10.5996

With F (78) =0.139, p=0.71 (two-tailed), the equality of the variances of the students' receptive word gain scores under the two schedules on the post-test was determined. Nevertheless, with t (78) =2.147, p=0.035 (two-tailed), the researchers concluded that the means of the participants' receptive vocabulary gain were significantly

different under the two treatment conditions. The difference magnitude under the two spacing schedules (mean difference=5.50, 95% *CI*: 0.4 to 10.60) was moderate ($\eta 2 \approx 0.06$). Hence, it was decided that the equal spacing condition had significantly contributed to improving the participants' receptive word retrieval.

5. Discussion

The present study was conducted to examine the effects of massed, equal, and expanding spacing conditions on the Iranian EFL learners' knowledge of receptive and productive vocabularies. In response to the posed research questions, it is reported that, although all the three groups, regardless of the encounter conditions, had performed better on their post-test, the participants had significantly better vocabulary retention in an equally-spaced (2-2-2) schedule in comparison to expanding and massed schedules. In other words, repeated exposure to input over equal intervals had helped the students to improve their vocabulary retrieval significantly. However, further analysis of the data demonstrated that an equally-spaced schedule was more effective than an expanding schedule in helping L2 learners develop their receptive knowledge of vocabulary. Such a difference was not observed regarding productive vocabulary knowledge. This finding could be due to the short duration of the study or the limited number of exposures to the target words. It also highlights the fact that productive vocabulary retrieval is a more daunting task than its receptive counterpart. This is why L2 learners might require a more extended spacing schedule to be more successful in this regard.

Nevertheless, until quite recently, it has widely been assumed that expanding spacing provides an optimal method for scheduling practices (e.g., Karpicke & Roediger, 2010; Nakata, 2015, Nakata & Elgort, 2020; Serrano & Huang, 2018), and it has been recommended that a gradual increase in the spacing intervals may facilitate vocabulary learning. Other researchers, however, have questioned this assumption (e.g., Cull, 2000, Karpicke & Roediger, 2007; Schuetze & Weimer-Stuckmann, 2011) and demonstrated that when memory is tested following a longer retention interval, uniform (equal) spacing, under some conditions, results in better performance than expanding spacing. When the target material is open to forgetting, interventions separated by expanding intervals can contribute to both successful recall of desirable data and prevention of unwanted recall of wrong information (Storm et al, 2010).

In contrast to most previous research endeavors comparing spacing schedules within the limits of a single learning session followed by a post-test given immediately after the experiment (e.g. Nakata, 2015), in the present study, the interventions and the post-test were given over several spaced learning episodes in a real-world classroom setting rather than in laboratory conditions. As Kang et.al, (2014) argued spaced repetitions and immediate post-test in a single session, irrespective of the spacing schedule, do not represent what happens in real-world learning episodes and are rarely enough for long-term retention. Besides, if the words are tested just a few minutes after the exposure, the loop is still working actively, at which point it is not easy to decide if the learner has yet established a record (Schuetze, 2015). Moreover, productive retrieval is more challenging and complicated than receptive retrieval.

What made this study further different from other related research was that, in all existing L2 vocabulary studies which compared the effects of equal and expanding spacing, the participants practiced receptive retrieval rather than productive retrieval, while here the effects of the two conditions were investigated on both receptive and productive vocabulary retrieval. The present research took place over a 16-session semester. Therefore, the spacing schedules had to be self-paced (2-2-2 and 0-1-5) due to the regulations regarding the use of time in the class. Future research could focus on the desirable spacing condition for vocabulary learning and retrieval.

Most of the studies regarding spacing effects, including the present one, only employed three retrieval trials and examined a greater number of retrieval attempts that could reveal the sequence of learning episodes in more applied settings. It would also be desirable to administer a delayed post-test to assess long-term vocabulary retention since in most real-world learning contexts the learned material is accessible for a long time. Finally, due to the limited resources of the institute in which the study was conducted, the students did not have access to computers. Hence, the researchers used teacher-made flashcards to present the target word pairs. Follow-up studies could benefit from computer-generated flash cards. This may guarantee the equivalence of retrieval times between the equal and expanding spacing conditions.

6. Conclusion

The findings of the study demonstrated the greater impact of equal spacing over expanding and massed spacing in vocabulary learning. In particular, they also verified the superiority of equal spacing over expanding spacing for receptive vocabulary retrieval. These findings are useful because they could assist L2 teachers in the process of optimizing L2 vocabulary learning from retrieval. Moreover, they could function as some guidelines for using flashcards in the process of vocabulary instruction, which could be greatly facilitated in a digital context, where more frequent exposures to input are possible. Repetition is a major element in doing different types of vocabulary exercises (True/False, fill-in-the-blank, rote repetition, word list memorization, matching, etc.) in Iran. Therefore, the findings of this study could benefit teachers and material developers in the appropriate spacing of related tasks and activities so that the students could improve their vocabulary knowledge more efficiently.

Input spacing can also inform course designers regarding how frequently L2 classes should meet during the week or semester, as the number of meetings directly affects the number of practice opportunities. An increased number of meetings enables teachers to have a greater chance of reviewing previously taught material over more distributed practice sessions. Given the findings of this study, a shift from back-to-back instruction (massed distribution) to an equally spaced instructional schedule could benefit the students in terms of vocabulary learning in general, and receptive word retrieval in particular. As emphasized by Cepeda et al. (2008), the compression of learning into a very short period will probably produce misleadingly high levels of immediate mastery that will not last for a long time.

However, as SLA has borrowed the concept of spaced practice from cognitive psychology, more robust research is required in this field with respect to the dichotomy of receptive and productive lexical knowledge. The reason is that several variables such as the frequency and the time interval between the spacings, the kind of feedback provided to the students after each retrieval, and the type of exposure could affect the findings of similar studies.

References

Baddeley, A. D. (2007). *Working memory, thought, and action.* Oxford University Press.

Baddeley, A.D. (2017). Modularity, working memory, and language acquisition. *Second Language Research*, *33*, 299–311.

Barcroft, J. (2007). Effects of opportunities for word retrieval during second language vocabulary learning. *Language Learning*, *57*, 35 – 56.

Bui, G., Ahmadian, M.J., & Hunter, A. M. (2019). Spacing effects on repeated L2 task performance. *System, 81*, 1-13.

Candry, S., Decloedt, J., & Eyckmans, J. (2020). Comparing the merits of word writing and retrieval practice for L2 vocabulary learning. *System*, *89*, DOI.org/10.1016/j.system.2020.102206

Celce-Murcia, M., Brinton, D. M., & Snow, M. A. (2014). *Teaching English as a second or foreign language* (4th ed.). National Geographic Learning.

Cepeda, N., Vul, E., Rohrer, D., Wixted, J., & Pashler, H. (2008). Spacing effects in learning: A temporal ridgeline of optimal retention. *Association for Psychological Science*, *19*(11), 1095-1101.

Crossley, S. A., Skalicky, S., Kyle, K., & Monteiro, K. (2019). Absolute frequency effects in second language lexical acquisition. *Studies in Second Language Acquisition, 41*(4), 721-744. 1-24. DOI:10.1017/S0272263118000268.

Cull, W. L. (2000). Untangling the benefits of multiple study opportunities and repeated testing for cued recall. *Applied Cognitive Psychology*, *14*, 215–235.

Elgort, I. (2011). Deliberate learning and vocabulary acquisition in a second language. *Language Learning, 61*(2), 367–413.

Fall, T., & Davies, P. (2012). *Solutions:Pre-Intermediate* (3rd ed.). Oxford University Press.

Gains, R., & Redman, S. (2008). *Oxford word skills: Intermediate*. Oxford University Press.

Juffs, A. (2019). Lexical development in the writing of English language program students. In R. M. DeKeyser, & G.P. Botana (Eds.), *Reconciling methodological demands with pedagogical applicability* (pp. 179–200). John Benjamins.

Kang, H.K., Lindesy, R.V, Mozer, M.C., & Pashler, H. (2014). Retrieval practice over the long term: Should spacing be expanding or equal-interval? *Psychonomic Bulletin & Review, 21,* 1544-1550.

Karpicke, D.J., & Roediger, L.H. (2007). Expanding retrieval practice promotes short-term retention, but equally spaced retrieval enhances long-term retention. *Journal of Experimental Psychology, 33*(4), 704-419.

Karpicke, D.J., & Roediger, L.H. (2010). Is expanding retrieval a superior method for learning text materials? *Memory & Cognition, 38*(1), 116-124.

Kapricke , J. D., & Bauernschmidt , A. (2011). Spaced retrieval: Absolute spacing enhances learning regardless of relative spacing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 37, 1250 – 1257.

Landauer, T. K., & Bjork, R. A. (1978). Optimum rehearsal patterns and name learning. In M. Gruneberg, P. E. Morris, & R. N. Sykes (Eds.), *Practical aspects of memory* (pp. 625–632). Academic Press.

Lenko-Szymanska, A. (2019). *Defining and assessing lexical proficiency.* Routledge.

Naismith, B., & Juffs, A. (2021). Finding the sweet spot: Learner's productive knowledge of mid-frequency lexical items. *Language Teaching Research, OnlineFirst*. https://doi.org/10.1177/13621688211020412.

Nakata, T. (2011). Computer-assisted second language vocabulary learning in a paired-associate paradigm: A critical investigation of flashcard software. *Computer Assisted Language Learning*, *24*, 17 – 38.

Nakata, T. (2015). Effects of expanding and equal spacing on second language vocabulary learning: Does gradually increasing spacing increase vocabulary learning? *Studies in Second Language Acquisition, 37*(4), 677-711. DOI:10.1017/S0272263114000825.

Nakata, T., & Webb, S. (2015). Does studying vocabulary in similar sets increase learning? The effects of part and whole learning on second language vocabulary acquisition. *Studies in Second Language Acquisition. 38*(3), 523-552. DOI:10.1017/S0272263115000236.randal

Nakata, T. (2016). Does repeated practice make perfect? The effects of within-session repeated retrieval on second language vocabulary learning. *Studies in Second Language Acquisition*. *39*(4), 653-679.

Nakata, T., & Elgort, I. (2020). Effects of spacing on contextual vocabulary learning: Spacing facilitates the acquisition of explicit, but not tacit, vocabulary knowledge. *Second Language Research, 37*(2), 233-260.

Namaziandost, E., Mohammed Sawalmeh, M. H., & Izadpanah Soltanabadi, M. (2020). The effects of spaced versus massed distribution instruction on EFL learners' vocabulary recall and retention, *Cogent Education, 7*(1), DOI:10.1080/2331186X.2020.1792261

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

Nation, I. S. P. **(**2014). How much input do you need to learn the most frequent 9,000 words? *Reading in a Foreign Language*, *26*(2), 1–16.

Nation, P. (2020). The different aspects of vocabulary knowledge. In S. Webb (Ed.), *The Routledge handbook of vocabulary studies* (pp. 15-29). Routledge.

O'brien, I., Segalowitz, N., Collenine, J., & Freed, B. (2006). Phonological memory and lexical, narrative, and grammatical skills in second language oral production by adult learners. *Applied Psycholinguistics, 27,* 377-402.

Ramanujan, K., & Weekes, B. S. (2020). Predictors of lexical retrieval in Hindi–English bilingual speakers. *Bilingualism: Language and Cognition, 23*(2), 265-273.

Roediger, H. L., & Karpicke, J. D. (2010). Intricacies of spaced retrieval: A resolution. In A. S. Benjamin (Ed.), *Successful remembering and successful forgetting: A festschrift in honor of Robert A. Bjork* (pp. 23- 47). Psychology Press.

Rogers, J. (2017). The spacing effect and its relevance to second language acquisition. *Applied Linguistics*, *38*, 906–911.

Rogers, J., & Cheung, A. (2018). Input spacing and the learning of L2 vocabulary in a classroom context. *Language Teaching Research.* Advance online publication. DOI: 10.1177/1362168818805251.

Rogers, J. (2021). Input spacing in second language classroom settings: Replications Bird (2010) and Serrano (2011), *Language Teaching*, *54*, 424-433.

Rohrer, D. (2015). Student instruction should be distributed over long time periods. *Educational*

Psychology Review, *27*, 635–643.

Ruiz, S., Rebuschat, P., & Meurers, D. (2019). The effects of working memory and declarative memory on instructed second language vocabulary learning: Insights from intelligent CALL. *Language Teaching Research.* Advance online publication. DOI: 10.1177/1362168819872859.

Schmitt, N. (2000).*Vocabulary in language teaching*. Cambridge university press.

Schmitt, N. (2014). Size and depth of vocabulary knowledge: What the research shows. *Language Learning*, *64*, 913–951.

Schuetze, U. (2015). Spacing techniques in second language vocabulary acquisition: Short-term gains vs. long-term memory. *Language Teaching Research*, *19*(1), 28–42.

Schuetze, U., & Weimer-Stuckmann, G. (2011). Retention in SLA processing. *CALICO Journal*, *28*, 460–472.

Serrano, R. (2011). The time factor in EFL classroom practice. *Language Learning*, *61*(1), 117–145.

Serrano, R., & Huang, H. Y. (2018). Learning vocabulary through assisted repeated reading: How much time should there be between repetitions of the same text? *TESOL Quarterly*, *52*, 971–94.

Smith, S. M. (2017). Working with working memory and language. *Language Teaching Research*, *33*(3), 291-297.

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

Storm, B., Bjork, R. & Storm, J. (2010). Optimizing retrieval as a learning event: When and why expanding retrieval practice enhances long-term retention. *Memory & Cognition*, *38*(2), 244-253. DOI:10.3758/MC.38.2.244.

Suzuki, Y., Nakata, T., & DeKeyser, R. (2019). The desirable difficulty framework as a theoretical foundation for optimizing and researching second language practice. *Modern Language Journal*, *103*, 713–720.

Thalheimer, W. (2006). Spacing learning events over time: What the research says. http://www.work-learning.com/catalog/

Webb, S. (2007). The effects of repetition on vocabulary knowledge. *Applied Linguistics, 28*(1), 46–65.

Wesche, M., & Paribakht, T. S. (1996). Assessing second language vocabulary knowledge: Depth vs. breadth. *Canadian Modern Language Review*, *53,* 13-39.

Zhu, Y. (2015). Revisiting word exposure frequency and incidental vocabulary acquisition. *Theory and Practice in Language Studies, 5*(5), 968-973.

Article I. Appendix A

Target words in two sets

Set A	**Set B**
1. well off = wealthy / rich	**1.** appalling = terrible
2. manufacture = produce	**2.** vanish = disappear
3. detest = hate	**3.** surrender = give up
4. thrilled = delighted	**4.** dull = boring
5. drawback = disadvantage	**5.** offender = criminal
6. complicated = complex	**6.** identical = exactly the same
7. postpone = put off	**7.** astonished = amazed
8. straightaway = immediately	**8.** roughly = approximately
9. scared stiff = terrified	**9.** filthy = very dirty
10. crucial = very important	10. Objective = goal/aim

Appendix B

Sample VKS items

There are five statements for each word. Read them carefully and choose the best statement or statements. **If you do section 5, please also do section 4.**

Fancy

1. I don't remember having seen this word before.
2. I have seen this word before but I don't know what it means.
3. I have seen this word before and I think it means …… (synonym)
4. I know this word. It means …***feel like***…… (synonym)
5. I can use this word in a sentence, e.g:..***Do you fancy a cup of coffee?***..……