

Paper Type: Original Article



## Improving EFL Learners' Vocabulary Knowledge through Extensive Reading and Spaced Retrieval Practice: Linking Input Hypothesis with Desirable Difficulty Theory

Fatemeh Rezaei <sup>1</sup>, Saeed Ranjbar <sup>2,\*</sup>

<sup>1</sup> Hadaf Institute of Higher Education, Sari, Iran; fatemehrezaei.75@gmail.com;

<sup>2</sup> Department of English Language, Qaemshahr Branch, Islamic Azad University, Qaemshahr, Iran; saeed.ranjbar43@gmail.com;

Received: 17 November, 2025

Revised: 22 March, 2026

Accepted: 18 May, 2026

### Abstract

In this study, the effectiveness of combining extensive reading with spaced retrieval practice was examined using digital flashcards and scheduled quizzes to enhance L2 vocabulary learning. Through linking Krashen's (1982) input hypothesis and Bjork and Bjork's (1992) theory of desirable difficulty, a quasi-experimental design was employed with 130 intermediate EFL learners in six groups, each of which receiving different combinations of interventions including extensive reading only (RG), e-cards only (EG), e-cards and extensive reading (ERG), e-cards and quizzes (EQG), extensive reading and quizzes (RQG), and all three components (ERQG). It was revealed that the ERQG significantly outperformed the other groups. ERG also had considerable gains, while the groups relying on extensive reading or e-cards only gained limited improvement. These results demonstrated that combining rich input with structured retrieval practice can have synergistic benefits, supporting the efficacy of spaced retrieval techniques for vocabulary knowledge improvement. The study also highlighted the pedagogical advantages of integrating extensive reading in EFL instruction with technology-enhanced retrieval activities.

**Keywords:** Digital flashcards, EFL learner, Extensive reading, Spaced retrieval practice, Vocabulary acquisition.

## I | INTRODUCTION

It is a common belief in the field of second language acquisition that knowing more words facilitates communication in the L2. In this regard, Mohsen and Hamed (2017) found that L2 learners generally struggle to understand and express themselves in conversations due to a shortage of vocabulary. Long before them, Krashen (1982) introduced the Input Hypothesis, which claims that language learning depends on understanding the meaning of what is read or heard. This implies that insufficient vocabulary knowledge explains L2 learners' failure in second or foreign language acquisition (Hanh, 2021). Consequently, a solid command of English words is essential to master all the aspects of English learning. Nation and Wang (1999) claimed that, to succeed in a reading or listening task, an English learner needs to recognize 98 percent of the words.

✉ Corresponding Author: [saeed.ranjbar43@gmail.com](mailto:saeed.ranjbar43@gmail.com)

Cite this article: Rezaei, F., Ranjbar, S., (2026). Improving EFL Learners' Vocabulary Knowledge through Extensive Reading and Spaced Retrieval Practice: Linking Input Hypothesis with Desirable Difficulty Theory, *Journal of Studies in Language Learning and Teaching*, 2(2), 95-109. doi: 10.22034/JSLLT.2026.23984.1102



Journal of Studies in Language Learning and Teaching. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY-NC) license.

A great number of studies in the field of L2 acquisition, such as Ayesha (2016), He and Luo (2023), Ming (2007), and Moir and Nation (2002), have recurrently emphasized the importance of L2 knowledge. However, gaining mastery of sufficient vocabulary knowledge has seemingly been a challenging goal for learners of English as a foreign language (EFL). Cennet (2019), for example, identified vocabulary shortage as the number-one obstacle for EFL learners' proficiency development. This is mainly because they find vocabulary memorization difficult (Afzal, 2019) and forget word meanings easily. They are also worried about the possibility of adopting wrong strategies to tackle vocabulary learning in English (Gao & Ma, 2011). All in all, it is fair to conclude that English learners are aware of the significance of vocabulary learning (Amiryousefi, 2015a, 2015b) but need constant help and support to acquire it (Afzal, 2019).

Numerous studies have introduced various teaching methods to improve vocabulary knowledge for EFL learners. Among them, several researchers have highlighted the importance of extensive reading as an effective strategy (Liza, 2017; Nation, 2001; Renandya, 2007). Grounded in Krashen's (1982) Input Hypothesis, extensive reading involves the enjoyable reading of a large amount of materials slightly above one's current language level ( $i + 1$ ). This approach provides the necessary comprehensible input for language acquisition. In this study, extensive reading consisted of learners engaging with graded texts on topics such as entertainment, health, culture, business, and biography. These texts were read outside of class to prepare for in-class comprehension activities.

This study recognizes the role of extensive reading in developing EFL learners' vocabulary size. However, it also aims to improve its effectiveness by combining it with spaced retrieval practice, which is seen as a promising method for committing information to long-term memory (Carpenter et al., 2012; Nation, 2001). Spaced retrieval practice is based on Bjork and Bjork's (1992) theory of desirable difficulty. This theory suggests that learning improves when tasks include the three key elements of spacing, retrieving, and interleaving. Spacing refers to distributing learning over time, retrieving involves actively recalling information, and interleaving entails switching between different topics. These elements strengthen both storage (how well information is learned) and retrieval (how easily information can be accessed).

In this study, spaced retrieval practice was implemented through two main methods. First, flashcards were created each containing a headword, its Persian meaning, synonyms (when applicable), an example sentence, and its Persian translation. These cards allowed the learners to actively recall the word meanings over several sessions. Second, six quizzes were scheduled every other session during the intervention to provide structured retrieval opportunities at spaced intervals. Vocabulary knowledge, as the dependent variable, was defined as growth in both breadth (the number of known words) and depth (the quality of word knowledge) (Nation, 2001). It was measured by the change in the participants' scores from pretest to posttest.

To investigate the effectiveness of these interventions, the study sought to answer the following research questions:

RQ1: Do EFL learners who receive different combinations of extensive reading, digital flashcards, and scheduled quizzes differ significantly in their vocabulary knowledge after the intervention?

RQ2: Which combination of instructional strategies, extensive reading only, digital flashcards only, scheduled quizzes only, or their various combinations, is most effective for enhancing EFL learners' vocabulary knowledge?

Accordingly, the following null hypotheses were formulated:

H01: There is no significant difference among EFL learners receiving different combinations of extensive reading, digital flashcards, and scheduled quizzes in terms of their post-intervention vocabulary knowledge.

H02: The combination of extensive reading, digital flashcards, and scheduled quizzes does not produce significantly higher vocabulary gains compared to other intervention types.



## II. Literature review

Krashen's input hypothesis and the theory of desirable difficulty were the theoretical foundations of this study which mainly aimed at building vocabulary knowledge of EFL learners effectively. The following sections encompass the description of these two theoretical perspectives.

### 2.1. The theoretical framework of the study

#### 2.1.1. Krashen's input hypothesis

Krashen's input hypothesis, inspired by child language acquisition, proposes that an L2 is acquired via extensive exposure to the target language (Krashen, 1982). That is, adequate input in the L2 is necessary for foreign language learning, and this input must be comprehensible to the learner. Comprehensibility of input was defined as the input that is slightly above the learner's current language ability ( $i + 1$ ). Accordingly, to improve the participants' vocabulary knowledge, this research relied on extensive reading as a realization of the input hypothesis in an EFL class. This approach is based on the findings established long ago by Krashen (1989) as well as Nation and Wang (1999).

#### 2.1.2. The theory of desirable difficulty

Bjork and Bjork (1992) conducted a series of experiments within the field of cognitive psychology and demonstrated that, to perform a learning activity successfully, a learner must be involved in tasks with a desirable level of difficulty. In other words, the learning task should not be so hard that the learner quits the task. Bjork and Bjork (1992) put forward three ways to achieve a desirable level of difficulty, including a) spacing, which implies extending a learning task over a long period of time, if possible, b) retrieving, which was explained as testing learners according to the learning objective as primary learning events, as frequently as possible, and c) interleaving, which was defined as making shifts among different learning topics. Accordingly, instructors were recommended to manipulate the learner's current learning conditions to create a desirable level of difficulty to achieve enhanced long-term retrieval of the learned material (Bjork & Bjork, 1992).

Within the framework of desirable difficulty, Bjork and Bjork (1992) made a sharp distinction between retrieval strength and storage strength to provide a better understanding of how the results of a learning activity are interpreted. While the former reflects the ease of retention, the latter measures the extent to which information was learned. For example, rereading a certain amount of information as a passive learning activity raises storage strength; however, it contributes little to retrieval strength (Hopkins et al., 2016). Bjork and Bjork (1992) claimed that calibrating the desirable difficulty of given learning activities via manipulating interleaving, retrieving, and spacing results in a greater increase in not only storage strength but also retrieval strength.

Bjork and Bjork's (1992) suggestion to achieve a desirable level of difficulty through interleaving, retrieving, and spacing was implemented in this quasi-experimental research to recall vocabulary. This was done with reference to the results of some previous empirical studies indicating the positive effects of performing vocabulary activities after reading tasks on EFL learners' vocabulary retention (Mirghasempoor, 2017). To operationally accomplish retrieving, spacing, and interleaving in this study, electronic flashcards (henceforth, e-cards) were used.

### 2.2. Review of the related empirical evidence

This section provides empirical evidence and a discussion of the theoretical foundation of the study, i.e., the input hypothesis and the theory of desirable difficulty, as described above. It is organized around four axial themes. First, extensive reading can improve vocabulary retention. Second, post-reading focused task of vocabulary contributes to L2 vocabulary retrieval. Third, spaced retrieval practice secures the retention of learning materials. Fourth, e-card is a promising tool for spaced retrieval practice assisting vocabulary retention.

### 2.2.1. Efficacy of (extensive) reading for L2 vocabulary learning

According to the input hypothesis, which is a long-held and well-referenced view of L2 learning, pleasurable reading is thought to improve the vocabulary repertoire of L2 learners, and an EFL teacher's job is to provide them with reading materials aligned with their competence (Liza, 2017; Renandya, 2007). Such a claim has been recurrently tested. For example, De'Ath (2001) used storybooks for a year and found that this helped the learners outperform their peers who undertook the audiolingual method for oral sentence repetition and word recognition. In a more recent study, Suk (2017) administered 15-week sessions after a course lasting for 168 hours of extensive reading, which led to the participants' better performance in reading speed and vocabulary knowledge. Similarly, AlRamadhan (2020), using textual glosses, found that incidental vocabulary learning among EFL learners could be facilitated if repetition were embedded in instruction.

There is a compelling body of evidence demonstrating how intentional vocabulary learning occurs via reviewing words in focus-on-form tasks, which can be added to extensive reading activities. Hung (2015), for example, claimed that, if learners notice the form and meaning of words, their vocabulary learning is optimized. Nation (2001) also contended that learners' acquisition of vocabulary accelerates provided the spelling and direct definitions of words are focused on. Mirghasempoor (2017) has recently shown that post-reading vocabulary-focused tasks, more specifically spelling-focused ones, could bring about better performance on a vocabulary posttest than meaning-focused ones. Similarly, Laufer and Rozovski-Roitblat (2015) experimented with the effects of using a variety of post-reading vocabulary tasks on the learners' vocabulary retention tested two weeks after the course had finished. The test included form, recognition, and word meaning. It was found that those who used a dictionary as a part of their post-reading activities outperformed their counterparts who did not. Moreover, among the dictionary users, those who had more vocabulary exercises performed better on the posttest than those who did fewer exercises but met the words more often.

### 2.2.2. Contribution of post-reading focused tasks of vocabulary to L2 vocabulary retrieval

Post-reading tasks have become an important pedagogical strategy to improve vocabulary learning and use. Atay and Kurt (2006) revealed the inadequacy of reading without follow-up practice, with L2 learners in EFL settings often producing "poor" lexicons after multiple years of classroom education. Their research compared isolated written tasks with combined written-interactive tasks and reported that the latter resulted in a significant increase in young learners' vocabulary. That is, interactive post-reading processing promotes deeper levels of lexical processing and retrieval. Additional support for post-reading tasks is also provided by Kang and Nicol (2025), who used an eye-tracking technique to investigate the effects of word-focused activities on reading behaviors and vocabulary acquisition. They found that the activities that involved fill-in-the-blank exercises and summary-writing were far more effective for vocabulary retention than comprehension-only activities. Importantly, the summarizing group's processing times for target words during rereading were significantly related to vocabulary test scores, suggesting that deeper cognitive mobilization for recalling information at post-reading contributed to retrieval. Similarly, Yang et al. (2017) reported that sentence-writing and gap-fill were superior to comprehension only in both immediate and delayed vocabulary tests. They also found the mediating effect of working memory, such that the learners with high working memory capacity actually improved more than those with low working memory through form-focused post-reading activities. These results are consistent with the Task Involvement Load Hypothesis, which suggests that tasks that involve higher cognitive loads may be beneficial to vocabulary learning.

The type of post-reading activity is also important. Göksoy and Dollar (2017) compared the effects of post-reading writing-based, speaking-based, and integrated writing-with-speaking tasks. They found that integrated skills (reading, speaking, and writing) tasks were generally more effective. However, speaking tasks appeared to contribute in particular to the improvement of learners' L2 proficiency over the longer term, perhaps through engaging learners and their motivation. It may be concluded that interactive and multimodal post-reading tasks could maximize vocabulary recall by tapping into cognitive and affective

channels. Explicitly connecting vocabulary instruction to post-reading activities has been promoted by [Sonbul and Schmitt \(2010\)](#), who argued that a combined approach of incidental learning and direct instruction of the meanings of words has more positive effects on different levels of vocabulary knowledge. Their results highlight that good post-reading activities are those that promote explicit lexical form and meaning awareness, which can be used to strengthen retrieval.

In addition to vocabulary acquisition, post-reading activities have been found to support more general aspects of L2 learning. [Ajideh et al. \(2024\)](#) reported that some activities enhance reading motivation, self-regulation, and comprehension, which indirectly facilitate lexical retrieval by contributing to stronger and more autonomous learning. More recently, [Nezhad and Hasan \(2022\)](#) demonstrated the diverse flexibility of post-reading tasks by revealing that title selection tasks, in which participants infer the central message of a text, also contributed to vocabulary learning. It seems that higher-order thinking can deepen lexical processing and retrieval.

### 2.2.3. The effectiveness of spaced retrieval practice for better retention

A few decades ago, spacing and retrieval were proved to have a constructive long-term impact on retention ([Carpenter et al., 2012](#)). Spacing (scheduling restudy sessions of certain material over a period of time) and retrieval (studying for recalling certain material) are well-researched and well-established in cognitive psychology. However, the focus of subsequent studies has been on merging these two practices as spaced retrieval practice ([Carpenter et al., 2012](#)). In other words, spacing is operationally defined as studying material repetitively in multiple consecutive sessions rather than in one single long session. [Sobel \(2011\)](#), for example, examined spacing with regard to the retention of uncommon words by comparing the performance of the students receiving material twice a day with that of the students who received the same material with a one-week interval. Interestingly enough, the mean posttest scores of the students receiving the material twice on a single day were half of those of the other group. Retrieval goes hand-in-hand with the spacing effect and implies that testing learners' memory of information rather than simply reviewing it better consolidates what they are supposed to learn ([Brown et al., 2014](#)).

Meanwhile, a large body of research has focused on merging spacing and retrieval to form the spaced retrieval practice approach to learning ([Hopkins et al., 2016; Rowland, 2014](#)). These studies showed that holding multiple sessions separate in time, in which information is retrieved repeatedly, can enhance long-term retention ([Smith & Karpicke, 2014](#)). These studies either explored the effects of retrieval under different conditions or attempted to discover optimal study schedules.

Research on optimal study schedules indicated that the longer the intervals between sessions, the more consolidated the retention of information (e.g., [Karpicke & Roediger, 2007; Storm et al., 2010](#)). Research results, however, are controversial when it comes to whether the intervals between learning sessions should be increasingly spaced or kept constant. That is, while some research suggests expanding learning intervals accomplish better retention ([Karpicke & Roediger, 2007](#)), others are in favor of consistent learning intervals ([Storm et al., 2010](#)). In sum, the results of both groups of studies are aligned with the desirable difficulty theory previously put forward by [Bjork and Bjork \(1992\)](#). Generally, it can be concluded that, based on this theory, a) learning happens in multiple sessions, b) longer intervals between learning sessions entail more lasting retention, and c) learning improves if it is accompanied by recalling sessions.

### 2.2.4. Role of spaced retrieval practice in L2 vocabulary learning

Research has consistently shown that spaced learning is better than massed practice for long-term retention ([Anthony, 2010](#)). [Kim and Webb \(2023\)](#) discovered that spaced practice improved delayed post-test performance compared to massed practice across various activities (such as fill-in-the-blanks and flashcards). [Gordon's \(2020\)](#) review recognized retrieval-based and spaced practice as effective strategies for maximizing word learning in both educational and clinical settings. [Lotfolahi and Salehi \(2017\)](#) raised concerns about spacing in EFL classrooms. They argued that, while spacing might initially hinder short-term retention, learners benefit in the long run from better vocabulary development. [Kanayama and Kasahara \(2016\)](#) compared expanding intervals to equally spaced retrieval practice and found that the early

benefits of expanding intervals decreased with repeated reviews. This challenges the belief that longer intervals are always better, emphasizing instead the importance of sufficient retrieval attempts over time. The format of practice also influences retrieval outcomes. Nakata (2016) looked at different retrieval formats (recognition, recall, combination, and generative recall) and found that recall-based exercises were the most effective for spelling knowledge. Nakata (2017) further showed that repeated retrieval within a session boosted long-term retention, but the benefits declined after five repetitions, suggesting a need to balance frequency and mental effort. Higham et al. (2023) studied the relationship between retrieval and restudying, finding that restudying sometimes led to better immediate recall. However, this effect vanished when controlling for memory ratings, indicating that learners' perceptions play a significant role. Lotfolahi and Salehi (2016) noted that students view spaced practice as more effective than cramming, which enhances its educational value. Technology, especially flashcard apps, has made spaced retrieval practice easier. Nakata's (2011) review of nine flashcard programs identified multilingual features and smart scheduling as crucial, though differences in design standards show room for improvement in digital learning tools.

The reviewed literature highlights three key foundations for this study. First, extensive reading, based on Krashen's (1982) input hypothesis, has been shown to support L2 vocabulary development by offering comprehensible input in meaningful contexts (Suk, 2017; AlRamadhan, 2020). Second, tasks focused on reading significantly boost vocabulary retention by fostering deeper cognitive processing and retrieval (Laufer & Rozovski-Roitblat, 2015; Kang & Nicol, 2025; Yang et al., 2017). Third, spaced retrieval practice, based on Bjork and Bjork's (1992) theory of desirable difficulty, has proved effective in transferring information into long-term memory. Digital flashcards are a practical tool to apply this practice in schools (Nakata, 2011, 2016; Kim & Webb, 2023). While these areas have been explored independently, few studies have looked at the combined effects of extensive reading, spaced retrieval practice, and digital flashcards with scheduled quizzes in one intervention for EFL vocabulary learning. This study aims to address the gap by examining whether and how combining these instructional strategies can better enhance EFL learners' vocabulary knowledge than any single method alone. Specifically, the research compares six instructional conditions to find the best combination for maximizing vocabulary gains among Iranian EFL learners.

### III. METHODOLOGY

#### 3.1. Research design

This study used a quasi-experimental design with a pretest-posttest control group layout. A quasi-experimental approach was chosen because the random assignment of the individual participants was not possible due to the existing class structures. Instead, six intact classes were randomly assigned to six experimental conditions. The study compared six groups receiving different combinations of instructional methods including a) e-cards only (EG), b) extensive reading only (RG), c) e-cards and extensive reading (ERG), d) e-cards, extensive reading, and scheduled quizzes (ERQG), e) e-cards and scheduled quizzes (EQG), and f) extensive reading and scheduled quizzes (RQG). The independent variable was the type of instructional intervention with six levels, while vocabulary knowledge, measured by posttest scores, was the dependent variable. The pretest scores were used as a covariate to control for the initial differences among the groups. The intervention lasted for 10 sessions during the first semester of the 2023-2024 academic year, and a posttest was given one week after the intervention.

#### 3.2. Participants

The participants in this research were selected through convenience sampling. More specifically, the students who enrolled in three different classes of general English courses in the first semester of the 2023-2024 academic year were asked to take the placement test. Totally, in three classes, 236 students participated in the placement test session, and finally 130 of them got a score  $\pm 1$  standard deviation higher and lower than the mean. These students were considered as the sample of this research and were



randomly grouped into six groups as described in Table 1. It should be noted that these groups were students of different engineering fields and humanities, and their average age was 21.5.

**Table 1.** The distribution of the participants

Groups	N		
	Male	Female	Total
1 e-cards only (EG)	8	11	19
2 Extensive reading only (RG)	9	12	21
3 e-cards and extensive reading (ERG)	11	12	23
4 e-cards, extensive reading and scheduled quizzes (ERQG)	7	15	22
5 e-cards and scheduled quizzes (EQG)	8	14	22
6 Extensive reading and scheduled quizzes (RQG)	13	10	23
Total	56	74	130

### 3.3. Instruments

#### 3.3.1. Placement test

This test was designed under the title of Total English Placement Test by Longman Institute in 2006 to determine the level of the students' English proficiency in EFL. This test consists of one hundred three-choice questions that measure students' grammatical and vocabulary knowledge. Each item in this test has one point, and the total score is calculated out of 100. The reliability of the test in this research was 0.83 calculated with the KR-21 formula, which showed the high reliability.

#### 3.3.2. Vocabulary knowledge test

Among the words available in the texts selected for the students, one hundred words were selected after consultation with two English language teachers who had more than 15 years of experience in teaching general language courses in Azad universities. Those items were placed in a list with a 5-point Likert scale, which consisted of Likert type five-point scale: 1) I did not see these points at all, 2) I saw this word, but I do not know its meaning, 3) I feel I know its meaning, but I am not sure, 4) I know its meaning, but I cannot make a sentence with it. 5) I can make a sentence with this word. This scale was previously used to identify unknown points by different researchers (e.g., Alamer, 2022; Wesche & Paribakht, 1996). The words that were in the first three categories were used as target words in this research.

#### 3.3.3. Vocabulary test

This test included 38 four-choice items to measure the students' knowledge of target words. The items of this test were selected from 52 initial items. Initially, the test that was piloted, and 14 items were removed due to either the high level of difficulty of the item or the defective options. Finally, in the second test, i.e., the refined version, the remaining 38 items were of acceptable consistency considering that the reliability of the test calculated with the KR-21 formula was 0.77. It should be noted that the validity of the test was approved by three English language teaching experts who had a PhD degree in this field.

### 3.4. Materials

Three types of instructional materials were used in this research, including reading comprehension texts, e-cards and scheduled quizzes. To compile reading comprehension passages, materials from a popular website ([www.er-lentral.com](http://www.er-lentral.com)) were used. By joining this website, one can access a large number of texts at different levels of difficulty and in different topics. For this training course, 32 texts from the main topics of entertainment, health, culture, business and biography were selected and given to the students in turn.



More specifically, the texts were selected from a pack of 47 texts after consulting a group of four university teachers who had been teaching general English courses at Islamic Azad University for at least ten years by the time the data were being collected.

From these texts, 100 difficult vocabulary items were selected with the help of two experienced professors from the pool of unknown or unfamiliar words identified via the vocabulary knowledge scale data. Then, their e-cards were prepared and sent to the students in the form of an Android/IOS application. The e-cards included the headword, its meaning in Persian, its synonyms if available, an exemplar sentence, and its translation in Persian. All the students could use one of the two applications, Flashcards Learn Languages 4.11.9 (for Android) or Flashcard Hero (for IOS), according to the operating system of their smartphones. It has to be emphasized that the main features of these two apps were the same and served the purpose of the study. Moreover, the contents of the e-cards were identical. It should also be noted that the target words included 100 items and were provided to the students all at once at the beginning of the semester, but the texts were provided to them gradually in each session. Scheduled quizzes were also prepared in the form of separate PDF files. They were sent to the students on a regular basis every other session. Full instructions including how the quiz would be taken and the time allotted, among others, were provided in Persian on the first page. The second page included some items of different test formats, mainly, fill-in-the-blank items, matching items, and cloze passages with multiple-choice items. In order to increase the seriousness of the students, they were told that the mean score of the six scheduled quizzes would be factored into their final course score. The students had a 15-minute quiz time allotted right from the moment the file was sent. They were supposed to answer the quiz which usually included 17 to 20 test items of the words extracted from the reading passages, both familiar and unfamiliar. According to the instructions, if they had sent the answers later than the allotted time, their answer sheets would not have been corrected.

### 3.5. Procedure

This research was actually done in two stages. In the first stage, which included piloting the instruments, the materials and content necessary to hold the class were prepared and compiled. First, with the help and advice of experienced professors who had the experience of teaching general English courses in Islamic Azad University in Mazandaran, appropriate texts were prepared for the students. Then, a list of 197 words was prepared from these texts and given to the same professors. According to their comments, these words were reduced to 172 words and were provided to the students in the form of a list as explained above. According to the answers given by the students attending the first session of the general English courses in the first semester of the academic year 2024-2025, one hundred words that were among the unfamiliar words were selected as the target words. After that, their e-cards were prepared in the form of two smartphone software programs, Flashcards Learn Languages 4.11.9 (for Android) or Flashcard Hero (for IOS). They were provided to the students who were supposed to receive e-cards as a part of the intervention.

The second phase of the study included sampling and homogenizing the participants. To this end, the placement test described above was distributed among the learners. They were told that the test was distributed for a diagnostic purpose and they were expected to do their best to answer the test items correctly. It should be noted that the students took the placement test and the pretest in the first and second sessions of the semester when the class was over, and those whose placement test scores were within one standard deviation above and below the average were considered as the participants in this research.

An explanation sheet was sent to each student in the form of a PDF file, in which the purpose of the course, the method of doing their work or duties, the expectations of the course instructor, the dates of the exams and quizzes and the grading rubrics at the end of the course were explained. In the fourth session, according to the previous notification, the students were asked to participate in the pre-test. They knew that this test would not affect their grade at the end of the course, and it was explained to them that the test was administered to estimate their vocabulary level. The selected texts were provided to them from the fifth session onwards, such that two texts were given to the students a week before the class and they were expected to read them so that they could answer the while-class comprehension

questions in the class. The questions were of the yes/no type or were wh-questions, and the students had to answer them upon teacher’s call. The questions were often designed to cover the sentences in which the target words were located. At the end of the class, the students were reminded to review the important words on the flash cards. The intervention lasted for 10 sessions, and the posttest was given to the participants one week after the intervention course.

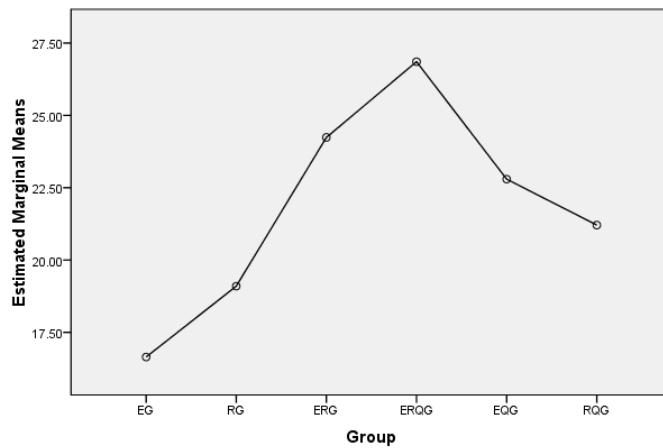
### IV. Results

One-way ANCOVA was performed to examine the effects of interventions and compare and contrast their effectiveness on the vocabulary knowledge of the participants after the intervention. Having confirmed the assumptions including normality of distribution and homogeneity of variance, the null hypothesis that the groups receiving interventions (Table 1) were similar in terms of their vocabulary posttest scores was tested. The results are shown below.

**Table 2.** Descriptive statistics of the posttest scores

Group	N	Mean	Std. deviation
EG	19	16.63	1.42
RG	21	19.09	2.50
ERG	23	24.21	3.05
ERQG	22	26.90	3.10
EQG	22	22.77	2.06
RQG	23	21.21	2.17

Table 2 shows that ERQG had the highest mean score after the intervention, whereas EG the lowest. In general, the mean scores imply that the groups spread over a continuum in terms of their knowledge of target vocabulary items. That is, the interventions had diverse effects as far as their effect on vocabulary knowledge improvement was concerned, as shown in Figure 1.



**Figure 1.** Distribution of the groups in terms of their post-intervention achievement

To compare the interventions in terms of their impact on vocabulary knowledge improvement, one-way ANCOVA was run in which pretest score was the covariate.

**Table 3.** One-way ANCOVA for vocabulary posttest scores

Source	Type III Sum of squares	df	M <sup>2</sup>	F	p	η <sup>2</sup>
Group	1378.74	5	275.74	44.80	.00	.64

The statistics in Table 3 confirm the difference among the groups receiving different interventions [ $F(130, 5) = 44.80, p = .00 < .05$ ]. Moreover, the observed partial eta square determining the effect size implies that the effect of intervention was very large ( $\eta^2 = .64$ ). Further pairwise comparison was done to compare the groups and see which group outperformed the other ones significantly.

**Table 4.** Pairwise comparison of the groups

(I) Group	(J) Group	Mean difference (I-J)	Std. error	p
<b>EG</b>	RG	-2.45*	.78	.03
	ERG	-7.59*	.76	.00
	ERQG	-10.20*	.78	.00
	EQG	-6.14*	.77	.00
	RQG	-4.56*	.77	.00
<b>RG</b>	EG	2.45*	.78	.03
	ERG	-5.14*	.74	.00
	ERQG	-7.75*	.75	.00
	EQG	-3.69*	.75	.00
	RQG	-2.11	.74	.08
<b>ERG</b>	EG	7.59*	.76	.00
	RG	5.14*	.74	.00
	ERQG	-2.61*	.74	.00
	EQG	1.43	.74	.80
	RQG	3.03*	.73	.00
<b>ERQG</b>	EG	10.20*	.78	.00
	RG	7.75*	.75	.00
	ERG	2.61*	.74	.00
	EQG	4.06*	.75	.00
	RQG	5.64*	.74	.00
<b>EQG</b>	EG	6.14*	.77	.00
	RG	3.69*	.75	.00
	ERG	-1.43	.74	.80
	ERQG	-4.06*	.75	.00
	RQG	1.58	.74	.50
<b>RQG</b>	EG	4.56*	.77	.00
	RG	2.11	.74	.08
	ERG	-3.03*	.73	.00
	ERQG	-5.64*	.74	.00
	EQG	-1.58	.74	.50

\* The mean difference is significant at the .05 level.

Table 4 confirms that ERQG clearly outperformed the other five groups, whereas EG had the poorest achievement. Moreover, ERG ranked second and outperformed EQG and RQG who performed similarly ( $I-J = 1.58, p = .50 > .05$ ). In addition, EQG outperformed RG who achieved vocabulary knowledge similar to that of RQG ( $I-J = 2.11, p = .08 > .05$ ).

## V. DISCUSSION

The results indicate that combining extensive reading with spaced retrieval practice through digital flashcards and scheduled quizzes, significantly affects L2 vocabulary acquisition. This was evident in the strong performance of the ERQG group. Regarding the first research question, i.e., whether there were significant differences in vocabulary knowledge among the six instructional groups, the results confirmed that there was a significant difference, leading to the rejection of the first null hypothesis. About the second research question, i.e., which combination was most effective, pairwise comparisons revealed a clear ranking; ERQG ranked the highest, significantly surpassing all the other groups. ERG was the second, doing significantly better than the single-intervention groups, but not significantly different from EQG. EQG and RQG were in the middle tier. RG (reading only) and EG (e-cards only) were the least effective. Thus, the second null hypothesis was rejected, with the full combination (ERQG) emerging as the best instructional approach for vocabulary learning among Iranian EFL learners.

The results are supported by the findings of previous research showing the efficacy of spaced retrieval practice (Carpenter et al., 2012; Gordon, 2020) and supporting the theory of desirable difficulty (Bjork & Bjork, 1992). It is emphasized that spaced and retrieval-based tasks enhance long-term retention. Moreover, integration of multiple modalities, such as reading, digital flashcards, and quizzes, was shown to create a fruitful learning environment that reinforces vocabulary knowledge through varied and repeated exposures. Scheduled quizzes may promote a sense of accountability and goal-directed learning (Ajideh et al., 2024). Additionally, the multimodal approach probably catered to diverse learning preferences, enhancing participation and retention (Göksoy & Dollar's, 2017).

Previous studies such as the one conducted by Suk (2017) have already confirmed the benefits of extensive reading (e.g., Suk, 2017) for developing language skills, especially, general reading skills. What is more, retrieval practice in isolation was shown to be of great help to learners (Karpicke & Roediger, 2007). However, few studies have integrated these two lines of instruction in EFL contexts, which makes the present study innovative. Indeed, the present study expanded the scope of our understanding in that what has been discussed in the related literature separately can be creatively integrated in a constructive manner. That is, the results of the present study demonstrated that a hybrid intervention which is based on sound underpinnings can outperform a single-method approach. More specifically, as far as vocabulary learning of EFL learners are concerned, the findings reported in the previous section mainly challenge the assumption that vocabulary development of EFL learners in general English courses can be maximized through input alone. The results, however, do not degrade what the participants in ERG achieved. In spite of their notable gains, their counterparts' gain in ERQG was more notable which emphasized the promising role of retrieval practice as well as the benefits of reading. The latter point deserves being tried and tested more meticulously not just by replication but by reconciling SLA theories, other than the one adopted in this research, with those of cognitive psychology to develop more holistic models of second language skill and sub-skill acquisition.

ERG's performance is supported by Nakata's (2016, 2017) studies, which highlighted the benefits of recall-based tasks for vocabulary retention. The use of e-cards could possibly facilitate active retrieval (Karpicke & Roediger, 2007). However, adding scheduled quizzes (as conducted in ERQG) amplified this effect, possibly by introducing structured intervals for retrieval and corrective feedback (Lotfolahi & Salehi, 2017). This highlights how important it is to combine extensive input with deliberate, spaced practice so as to enhance vocabulary knowledge.

Relying solely on extensive reading (RG) or e-cards (EG) was partially advantageous in comparison to the other four interventions. RG's performance was consistent with Krashen's (1982) input hypothesis, implying that reading alone is insufficient for deep vocabulary retention. Atay and Kurt's (2006) findings echoed these findings demonstrating the limitations of passive exposure and the need for supplementary post-reading activities. Similarly, EG's lower scores indicated that flashcards alone, without contextualized input or additional retrieval tasks, are not likely to be beneficial, echoing Kim and Webb's (2023) observation that the effectiveness of spaced practice varies depending on the activity type. RQG and EQG's performance suggests that quizzes play a significant role in vocabulary retention, but their impact is maximized if paired with extensive reading and digital flashcards. This resonates Sonbul and Schmitt's

(2010) argument that explicit vocabulary instruction yields deeper lexical knowledge, if it is combined with incidental learning. The quizzes may give learners chances for retrieval and self-assessment (Higham et al., 2023).

## VI. Conclusion

This research provided compelling evidence regarding the effectiveness of combining extensive reading with spaced retrieval practice, particularly through digital flashcards and scheduled quizzes, in improving L2 vocabulary learning. The findings, supported by Krashen's (1982) input hypothesis and Bjork and Bjork's (1992) theory of desirable difficulty, have significant pedagogical implications. To create more effective vocabulary learning environments, EFL teachers should invest in teaching extensive reading in combination with spaced retrieval activities, such as flashcards and periodic quizzes. The use of technology (e.g., digital flashcards) offers a practical way to implement spaced retrieval practice in educational settings. Additionally, scheduled quizzes can foster accountability and motivation due to their structured nature, helping learners overcome vocabulary attrition and learner disengagement (Afzal, 2019; Cennet, 2019).

Future research can be conducted to address the limitations of this study. First, beginners or advanced learners could be the focus of study to gain new insights. Second, a longer intervention could explore the sustainability of vocabulary gains over extended periods. Third, other technological aspects besides those included in this study (e.g., multimedia support, adaptive scheduling) deserve deeper investigation. Finally, individual differences, such as working memory capacity or motivation, could be considered as mediating variables.

## AUTHORS' BIOGRAPHIES

**Fatemeh Rezaei** is currently a PhD student of TEFL at Islamic Azad University, Qaemshahr Branch, and teaches general English courses at Hadaf Institute of Higher Education in Sari. She has been interested in researching various aspects of L2 reading comprehension in EFL contexts and has published research papers examining EFL learners L2 reading strategies, and L1-L2 reading strategies transfer in academic journals and conferences.

**Saeed Ranjbar** is an assistance professor and a retired faculty member of Islamic Azad University, Qaemshahr Branch. He has been teaching English and translation for over 30 years. He and his colleague(s) published a number of books in the area of teaching general English to students and published some research articles in applied linguistics journals. He also ran some workshops covering skills and topics such as Think Aloud Protocols (TAP) presented at Allameh Tabatabaei University. His major interests are teaching language skills particularly reading and speaking, translation and TAP.

## REFERENCES

Afzal, N. (2019). A study on vocabulary-learning problems encountered by BA English majors at the university level of education. *Arab World English Journal (AWEJ)*, 10.

Ajideh, P., Zohrabi, M., & Khojand, B. (2024). Exploring the effectiveness of pre-and post-reading activities on the development of reading motivation and self-regulation as essentials for reading comprehension. *Journal of Modern Research in English Language Studies*, 11(3), 25-49.

Alamer, A. (2022). Basic psychological needs, motivational orientations, effort, and vocabulary knowledge: A comprehensive model. *Studies in Second Language Acquisition*, 44(1), 164-184.



- AlRamadhan, M. H. (2020). L1 Textual Glosses and Word Repetition: Facilitative Interventions for Incidental Vocabulary Acquisition. *International Journal of Instruction*, 13(4), 815-832.
- Amiryousefi, M. (2015). Iranian EFL learners' and teachers' beliefs about the usefulness of vocabulary learning strategies. *Sage Open*, 5(2).
- Amiryousefi, M. (2015). Iranian EFL teachers and learners' beliefs about vocabulary learning and teaching. *International Journal of Research Studies in Language Learning*, 4(4), 29-40.
- Atay, D., & Kurt, G. (2006). Elementary school EFL learners' vocabulary learning: The effects of post-reading activities. *Canadian Modern Language Review*, 63(2), 255-273.
- Ayesha, S. (2016). Teaching and learning vocabulary: Students' attitudes. *The EDRC Journal of Learning and Teaching*, 1(2), 81-91.
- Bjork, R. A., & Bjork, E. L. (1992). A new theory of disuse and an old theory of stimulus fluctuation. In A. Healy, S. Kosslyn, & R. Shiffrin (Eds.), *From learning processes to cognitive processes: Essays in honor of William K. Estes* (Vol. 2, pp. 35-67). Hillsdale, NJ: Erlbaum.
- Brown, P. C., Roediger, H. L., III., & McDaniel, M. A. (2014). *Make it stick: The science of successful learning*. Cambridge, Massachusetts: The Belknap Press of Harvard University Press.
- Carpenter, S. K., Cepeda, N. J., Rohrer, D., Kang, S. H. K., & Pashler, H. (2012). Using spacing to enhance diverse forms of learning: Review of recent research and implications for instruction. *Educational Psychology Review*, 24(3), 369-378. <https://doi.org/10.1007/s10648-012-9205-z>
- Cennet, A. (2019). Integrating a computer-based flashcard program into academic vocabulary learning. *TOJET: The Turkish Online Journal of Educational Technology*, 18(1), 44-62.
- De'Ath, P. (2001). The Niue literacy experiment. *International Journal of Educational Research*, 35(2), 137-146. [https://doi.org/10.1016/S0883-0355\(01\)00011-8](https://doi.org/10.1016/S0883-0355(01)00011-8)
- Gao, X., & Ma, Q. (2011). Vocabulary learning and teaching beliefs of pre-service and in-service teachers in Hong Kong and mainland China. *Language Awareness*, 20(4), 327-342. <https://doi.org/10.1080/09658416.2011.579977>
- Göksoy, E., & Dollar, Y. K. (2017). Comparative study of acquisition rates between English vocabulary sets intentionally taught via post-reading writing and post-reading speaking skills. *Turkish Online Journal of English Language Teaching*, 2(2), 85-102.
- Gordon, K. R. (2020). The advantages of retrieval-based and spaced practice: Implications for word learning in clinical and educational contexts. *Language, Speech, and Hearing Services in Schools*, 51(4), 955-965.
- Hanh, L. T. T. (2021). Vocabulary Learning Strategy and Vietnamese University Students' Learning Experience in English as Medium of Instruction Classrooms. *International Journal of Instruction*, 14(3), 117-132.
- He, L., & Luo, Q. (2023). Developing Thai speaking skills of Chinese students using vocabulary learning strategies. *International Journal of Instruction*, 16(1), 983-998.
- Higham, P. A., Fastrich, G. M., Potts, R., Murayama, K., Pickering, J. S., & Hadwin, J. A. (2023). Spaced retrieval practice: Can restudying trump retrieval? *Educational Psychology Review*, 35(4), 98.



Hopkins, R. F., Lyle, K. B., Hieb, J. L., & Ralston, P. A. S. (2016). Spaced retrieval practice increases college students' short- and long-term retention of mathematics knowledge. *Educational Psychology Review*, 28(4), 853–873. <https://doi.org/10.1007/s10648-015-9349-8>

Hung, H. T. (2015). Intentional Vocabulary Learning Using Digital Flashcards. *English Language Teaching*, 8(10), p107. <https://doi.org/10.5539/elt.v8n10p107>

Kang, H., & Nicol, J. (2025). Effects of post-reading word-focused activities on L2 learners' reading behavior and vocabulary learning: an eye-tracking study. *The Language Learning Journal*, 1-19.

Karpicke, J. D., & Roediger, H. L. (2007). Expanding retrieval practice promotes short-term retention, but equally spaced retrieval enhances long-term retention. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 33(4), 704–719. <https://doi.org/10.1037/0278-7393.33.4.704>

Khoii, R., & Fallah Abed, K. (2022). A Comparative Study of the Effects of Equal and Expanding Spacing Schedules on L2 Receptive and Productive Vocabulary Retrieval. *Research in English Language Pedagogy*, 10(3), 412-436.

Kim, S. K., & Webb, S. (2023). Does spaced practice have the same effects on different second language vocabulary learning activities? Fill-in-the-blanks versus flashcards. *The Modern Language Journal*, 107(4), 944-964.

Krashen, S. (1989). We acquire vocabulary and spelling by reading: Additional evidence for the input hypothesis. *The Modern Language Journal*, 73(4), 440–464. <https://doi.org/10.2307/326879>

Krashen, S. D. (1982). Second language acquisition theory. In *Principles and practice in second language acquisition* (pp. 9-45). Oxford, MS: Pergamon Press.

Laufer, B., & Rozovski-Roitblat, B. (2015). Retention of new words: Quantity of encounters, quality of task, and degree of knowledge. *Language Teaching Research*, 19(6), 687–711. <https://doi.org/10.1177/1362168814559797>

Liza, E. M. (2017). Adult Hispanic ESL students and graded readers. *The CATESOL Journal*, 29(2), 41–59.

Lotfolahi, A. R., & Salehi, H. (2016). Learners' perceptions of the effectiveness of spaced learning schedule in l2 vocabulary learning. *Sage Open*, 6(2), 2158244016646148.

Lotfolahi, A. R., & Salehi, H. (2017). Spacing effects in vocabulary learning: Young EFL learners in focus. *Cogent Education*, 4(1), 1287391.

Ming, W. (2007). An examination of vocabulary learning of college-level learners of English in China. *The Asian EFL Journal*, 9(2), 93–114.

Mirghasempoor Ahmadi, S. R. (2017). The impact of incidental and intentional L2 vocabulary learning on depths and breadth dimensions of vocabulary knowledge. *World Journal of English Language*, 7(1), 50. <https://doi.org/10.5430/wjel.v7n1p50>

Mohsen, Z., & Hamed, B. (2017). Incorporating intelligent flashcards for fostering EFL learners' vocabulary knowledge: The case of Anki Software. *Technics Technologies Education Management*. 12(3): 184-190.



- Moir, J., & Nation, I. S. P. (2002). Learners' use of strategies for effective vocabulary learning. *Prospect*, 17(1), 15–35.
- 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(1), 17-38.
- Nakata, T. (2016). Effects of retrieval formats on second language vocabulary learning. *International Review of Applied Linguistics in Language Teaching*, 54(3), 257-289.
- Nakata, T. (2017). 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.
- Nation, I. S. P. (2001). *Learning Vocabulary in Another Language*. Cambridge: Cambridge University Press. doi:10.1017/CBO9781139524759
- Nation, P., & Wang, K. M. (1999). Graded readers and vocabulary. *Reading in a Foreign Language*, 12(2), 355-380.
- Nezhad, M., & Hasan, A. (2022). Effect of title selection technique as a post-reading task on intermediate language learners' vocabulary achievement. *Journal of new advances in English Language Teaching and Applied Linguistics*, 4(2), 1002-1013.
- Renandya, W. A. (2007). The power of extensive reading. *RELC Journal*, 38(2), 133–149. <https://doi.org/10.1177/0033688207079578>
- Rowland, C. A. (2014). The effect of testing versus restudy on retention: A meta-analytic review of the testing effect. *Psychological Bulletin*, 140(6), 1432–1463. <https://doi.org/10.1037/a0037559>
- Smith, M. A., & Karpicke, J. D. (2014). Retrieval practice with short-answer, multiple-choice, and hybrid tests. *Memory*, 22(7), 784–802. <https://doi.org/10.1080/09658211.2013.831454>
- Sobel, H. S., Cepeda, N. J., & Kapler, I. V. (2011). Spacing effects in real-world classroom vocabulary learning. *Applied Cognitive Psychology*, 25(5), 763–767. <https://doi.org/10.1002/acp.1747>
- Sonbul, S., & Schmitt, N. (2010). Direct teaching of vocabulary after reading: Is it worth the effort?. *ELT journal*, 64(3), 253-260.
- Storm, B. C., Bjork, R. A., & Storm, J. C. (2010). Optimizing retrieval as a learning event: When and why expanding retrieval practice enhances long-term retention. *Memory & Cognition*, 38(2), 244–253. <https://doi.org/10.3758/MC.38.2.244>
- Suk, N. (2017). The effects of extensive reading on reading comprehension, reading rate, and vocabulary acquisition. *Reading Research Quarterly*, 52(1), 73–89. <https://doi.org/10.1002/rrq.152>
- Wesche, M., & Paribakht, T. S. (1996). Assessing second language vocabulary knowledge: Depth versus breadth. *Canadian Modern Language Review*, 53(1), 13-40.
- Yang, Y., Shintani, N., Li, S., & Zhang, Y. (2017). The effectiveness of post-reading word-focused activities and their associations with working memory. *System*, 70, 38-49.