Issues in Language Teaching (ILT), Vol. 6, No. 2, 239-258, December 2017

Studying the Effect of Retrieval Direction during Reading on Productive and Receptive Knowledge of Vocabulary

Sasan Baleghizadeh*
Associate Professor, Shahid Beheshti University

Samaneh Shafeie
M.A. in TEFL, Shahid Beheshti University

Abstract
Retrieval tasks provide learners with an opportunity to focus both on meaning and on form. There are four different retrieval directions. The present study aimed to identify the optimal direction of recall type retrievals during reading and to investigate the outcomes of each one. Forty-eight intermediate EFL learners took part in the study. One of the experimental groups was provided with the productive retrieval version of the reading text, and the other experimental group with the receptive retrieval version. A posttest was conducted in both productive and receptive directions for all the groups. Delayed posttests were administered two weeks after the treatment without prior notice to evaluate the time effect on participants' performance in each group. The results showed that there was a statistically significant difference in participants' performance in immediate and delayed posttests based on the method of retrieval. Further explanation of the effect of each retrieval direction on immediate and delayed posttests has been presented.

***Keywords*:** Retrieval, receptive recall, productive recall

Corresponding author: Sasan Baleghizadeh (sasanbaleghizadeh@yahoo.com)

INTRODUCTION
Retrieval practice plays an important role in learning in general (e.g. Carpenter, 2012; Endres, Carpenter, Martin & Renkl, 2017) and learning vocabulary in particular (e.g. Barcroft, 2007; Goossens, Verkoeijen &

Tabbers, 2014). Barcroft (2015) believes that retrieval tasks can facilitate vocabulary leaning during implicit reading. Retrieval tasks help learners not to break the flow of meaning in order to focus on form. In fact, focusing both on meaning and on form results in increased vocabulary learning. Retrieval can modify one's existing knowledge and memory tracing pattern (e.g. Barcroft, 2015; Migueles & Garcia-Bajos, 2007) which in turn results in effective and long term learning (Finn, Roediger & Rosenzweig, 2012). Barcroft (2015) argues that contrary to coping in which the focus is only on form, and contrary to synonym writing in which the focus is only on meaning, retrieval tasks provide learners with an opportunity to focus both on meaning and on form. He puts forward the claim that reading and receiving a pile of input will not result in effective vocabulary learning. Learners need to pay intentional attention in order to learn the target words. Barcroft (2015) develops the claim that attempting to retrieve words, whether successfully or unsuccessfully, will result in writing the form with a focus on meaning. This will lead to stabilized learning of vocabulary. On the basis of the evidence currently available (Ellis, 2006; 2016), it seems fair to suggest that focus solely on meaning will not result in effective learning, and, if any, it will be a long term process. Barcroft (2015) shows how teachers can teach new vocabulary through focus on form and without interrupting the focus on meaning while reading the text. The retrieval method that Barcroft has used helps students to achieve efficiency more quickly. Meanwhile the focus will be on form, not forms.

Nakata (2016) defines retrieval as a task in which "learners are required to access information about an L2 word from memory" (p. 257). He properly categorizes retrieval tasks into two different dichotomies: Recognition versus recall, and receptive versus productive. Accordingly, four types of retrieval can be showed in table 1 below:

Table 1: Four Different Types of Retrieval

	Recognition	recall
Receptive	1. Receptive recognition (e.g. choosing the Persian equivalent for an English word among some options available)	2. Receptive recall (e.g. writing the Persian equivalent for an English word from memory)
Productive	3. Productive recognition (e.g. choosing the English	4. Productive recall (e.g. writing the English equivalent

	equivalent for a Persian word among some options available)	for a Persian word from memory)

One issue with the method of Barcroft's (2015) study was the neglecting of these four different types of retrieval and the direction of translation during treatment and in posttest. Participants in experimental group were provided with L1 translation of the target words and were supposed to write the English equivalent words (that is, according to Webb (2007), a productive task). In post-test, both directions (productive and receptive skills) were assessed. The results indicated that participants performed better in the receptive skill (translation from L2 to L1) than in the productive one (translation from L1 to L2). This needs to be studied deeply. One group of participants should be provided by receptive tasks and another group with productive tasks, and then the groups should be assessed in both directions. So that more valid and reliable results are obtained regarding students performance in different directions of post-test according to the treatment they have received.

LITERATURE REVIEW

In the literature, the positive effect of retrieval on learning English vocabulary has been widely researched (e.g. Barcroft, 2007; Barcroft 2015; Fritz, Morris, Acton, Voelkel & Etkind, 2007; Nakata; 2016). Barcroft (2015) assessed the effect of providing learners with opportunities to retrieve novel words during reading a passage. Seventy-four Spanish-speaking participants in intermediate level of English as a second language participated in the study. They were instructed to read an English text for meaning. Five target words were used in the treatment, and each appeared three times in the text. In the control group, Spanish translations of the target words appeared next to the target word all three times. However, the experimental group viewed the Spanish translation for each target word only the first time and then attempted to retrieve and write the word on their own for the other two appearances. The results of this study revealed that the experimental group significantly outperformed in vocabulary learning over the control group. Contrary to many studies on vocabulary learning that have used non-words as target words during their treatment, Barcroft used actual words to help students benefit form learning the words. Obviously, the concern should not only be on immediate performance after

the treatment, as Barcroft's was, but on delayed performance as well. Consequently, another post-test is needed to find out the time effect on participants' performance.

Recently, several authors (e.g. Bjork, 1994; Pyc & Rawson, 2009; Rowland, 2014) have proposed the efficacy of recall retrieval over recognition retrieval. One of the most interesting approach to this issue has been proposed by Bjork (1994). He put forth the *retrieval effort hypothesis*, indicating that difficult recall enhances learning more than recognition. Accordingly, in the present study, only recall rather than recognition has been used during treatment. In another example, Rowland (2014), through meta-analyzing 60 related studies, examined the effect of retrieval tests of previously studied information versus simply restudying them. The results showed that recall retrieval tests brought about larger benefits than recognition tests. He proposed that future studies benefit from considering both episodic and contextually derived contribution to retrieval. The latter is what the present study aimed to investigate.

Contrary to Barcroft (2015), who examined the effect of retrieval in the context of reading passage, some other researchers investigated the effect of retrieval practice using word pairs or word lists (e.g. Liu, Rosburg, Gao, Weber & Guo, 2017; Toppino & Cohen, 2009). Nakata (2016), for example, studied vocabulary retrieval in paired associate. Nakata aimed to investigate the optimal retrieval format. Barcroft (2015) administered the posttest in both directions of translation (L2 to L1 that is productive and L1 to L2 that is receptive retrieval), but the treatment was only one directional (L2 to L1). Nakata (2016) put into account the retrieval direction both in treatment and in posttest. He taught 60 Swahili-English word pairs to 64 English-speaking college students. The treatment duration was less than 45 minutes, and students had no previous experience of Swahili. The results of his study indicated that productive recall formats are more effective than recognition.

Productive and receptive retrievals have attracted much attention from research teams (e.g. Campos, Rodriguez-Pinal & Perez-Fabello, 2014; Mondria & Wiersma 2004; Webb, 2009; Zhong, 2016). Webb (2009), for example, investigated the efficacy of receptive and productive learning of words and the effect of it on five aspects of vocabulary knowledge. 62 Japanese native speakers took part in that study. One of the experimental

groups studied 10 word pairs receptively and the other group studied the word pairs productively. The results, obtained by Webb, suggest that productive learning of word pairs is more effective than receptive learning. Similarly, the present study focused on both productive and receptive retrieval but contrary to Webb's study, learning and retrieval happened in the context of reading passage, not in word pairs.

PURPOSE OF THE STUDY

The purpose of this paper is to investigate the effect of retrieval on vocabulary learning during reading. In order to identify the optimal direction of retrieval during reading, the present study compared the effect of productive recall with receptive recall on students' performance in productive and receptive posttests. The time effect between immediate and delayed posttests was investigated to show the extent to which second language learners are able to recall learned words via productive and receptive recalls.

The following research questions are addressed in this study:

1. What are the main effects of productive and receptive retrieval during reading on students' productive and receptive vocabulary knowledge?
2. Is there any difference between vocabulary learning scores of the participants in the immediate and the delayed posttests?

METHOD

Participants

There were three independent variables in this study which were the treatment conditions including L1 to L2 (productive) retrieval, L2 to L1 (receptive) retrieval, and no retrieval condition. Participants' performance in immediate productive posttest, immediate receptive posttest, delayed productive posttest, and delayed receptive posttest were the four dependent variables. The study was carried out at an English teaching institute in Karaj, Iran. The age of the participants ranged from 14 to 16. The number of participants in this study consisted of 48 students who voluntarily took part in the research. All the participants were at the intermediate level in EFL who had received more than three years of formal English instruction.

Table 2 gives a visual representation of the participants' characteristics. Since randomization of individuals was not feasible, the engaged three intact classes were randomly assigned into one control and two experimental groups.

Table 2: Students' Background Information

Characteristics	Group A (n=16)	Group B (n=16)	Group C (n=16)
Age	14-16	14-16	14-16
Gender	Female	Female	Female
Proficiency level	Intermediate	Intermediate	Intermediate
Native language	Persian	Persian	Persian
Target language	English	English	English
Years of language learning	3-4	3-4	3-4
Place of language learning	Institutes and school	Institutes and school	Institutes and school

Instrumentation

The following instruments were used in the study:

Instrument 1. Proficiency Test

In order to ensure the homogeneity of the participants, Key English Test (KET) was employed to confirm that there was no significant difference between the language knowledge levels of the selected participants. The KR-21 reliability of the test was found to be 0.89.

Instrument 2. Knowledge of the Target Words Questionnaire

To ensure that the target words were unfamiliar to the participants, a questionnaire was designed including 3 questions. In the first question, participants were provided by 30 English vocabulary including the 5 target words. They were asked to check the words they know. The two others were grammar questions that functioned as distracting questions.

Instrument 3. Three Versions of the Reading Passage

The reading passage (see the Appendix I and II), which contained 456 words, was titled "The Prince and the Pauper". The passage was chosen from students' course book *Family and Friends 3* (Thompson & Simmons,

2010). Some sentences was added to the passage so that the experimental words be repeated three times. It was a short story and easy to comprehend but contained words that were not known by the participants. Participants could benefit from learning the target words. The six experimental words were *clever, gates, guards, swap, twins,* and *armor*. All three groups were provided with the Persian equivalent of the words for the first time that they appeared in the passage. For the two other appearances, Group A (first experimental group) was provided with Persian words and was asked to retrieve and write the English equivalent word. Group B (second experimental group) was provided with English words and was asked to retrieve and write the Persian equivalent. Group C (control group) had no retrieval opportunity and was provided with both English and Persian words all three times.

Instrument 4. Distracting Math Tasks

A sheet of math problems (e.g. 89+79 = ?) was prepared to distract the participants before performing the posttests and in the middle of the two productive and receptive posttests.

Instrument 5. Immediate and Delayed L1 to L2 and L2 to L1 Post-Tests

Productive (e.g. عوض کردن: ……..) and receptive (e.g. swap: ……..) recall of the target words was conducted for all three groups, once immediately after the treatment and once after two weeks of delay.

Data Collection Procedure

The following steps were carried out:

1. To ensure the proficiency level of the participants, the Key English Test (KET) was administered and 48 students were chosen out of 52.
2. Participants were randomly assigned into two experimental groups and one control group.
3. Students were instructed to work individually and not talk with other students until the experiment was completed.
4. The pretest was conducted for all participants to make sure that they do not already know the experimental words. Participants were given 5 minutes to complete the pretest.

5. Instructions for the reading passage, written in Persian, were given to the participants. They were given as much as time to read the instructions carefully. In the instructions, participants were asked to read for comprehension and do not use dictionaries.
6. One of the experimental groups (Group A), containing 16 participants, was provided with the productive retrieval version of the reading text (Appendix I). The other experimental group (Group B), containing 16 participants, was provided with the receptive retrieval version of the reading text (Appendix II). The control group (Group C), containing 16 participants, was provided with English words next to the Persian equivalents all three times that the words appeared in the passage without any retrieval opportunity. Participants were given 10 minutes to read the passage and fill the blanks with retrieved words associated with their learning condition (receptive retrieval or productive retrieval).
7. Reading passages were gathered and participants were given a sheet of math problems to answer as many as they can in five minutes for distracting purposes. It was not announced that the math problems were for distracting purposes because it would affect students' performance.
8. Then, the immediate posttests were conducted. The participants rest assured that responses will not be identified by the individual and all the responses will be compiled together and analyzed as a class. All the three groups first answered the productive (L1 to L2) vocabulary test. After that, they solved distracting math problems for five minutes. Then answered the receptive (L2 to L1) vocabulary test. They were given two minutes for each L1 to L2 and L2 to L1 vocabulary tests.
9. Delayed posttests were administered two weeks after the treatment without prior notice to evaluate the time effect on participants' performance in each group.

Scoring

For productive (L1 to L2) translation, one score was given for each completely produced English word. Half a score was assigned for responses in which at least half of the target word was correctly produced. The same

scoring process was applied for receptive (L2 to L1) translation. However, score of half was not needed for the receptive posttest, because the participants could write in their L1.

Data Analysis

The row data of the immediate and delayed posttests were collected. In order to analyze the data, it was fed into SPSS (version 23). Prior to any analyses, the distributions of all variables were checked for normality. The Shapiro-Wilk's test, which is appropriate for the sample size of less than 50 (Rovai, Baker & Ponton, 2013), was used to test the normality assumption. Then, descriptive statistics of the study was represented, followed by a multivariate analysis of variance (MANOVA) at the .05 level of significance. Through using MANOVA, there is a better chance of discovering important factors. Moreover, it protects against Type I errors (Gamage, Mathew & Weerahandi, 2004). Finally, four independent sample t-tests were conducted at the .05 level of significance to determine if there was any statistically significant difference between participants' performances in the four (immediate productive, immediate receptive, delayed productive, and delayed receptive) posttests.

RESULTS

In the following, the results regarding each research question are presented.

Investigating Research Question 1

The first research question deals with the main effects of productive and receptive retrieval on students' productive and receptive vocabulary knowledge during reading. To this end, the distributions of all variables were first checked for normality. Review of the Shapiro-Wilk test for normality, skewness, and kurtosis statistics suggested that normality was a reasonable assumption for all the three groups in immediate and delayed posttests (table 3). The values for skewness and kurtosis more than -2 and less than +2 are considered acceptable in order to prove normal distribution (George & Mallery, 2010).

Table 3: Tests of Normality

Posttest (DVs)	Methods of retrieval (IVs)	Skewness	Kurtosis	Shapiro-Wilk		
				Statistic	df	Sig.
ProductiveIPT	Group A	.000	-1.168	.910	16	.116
	Group B	-.241	-.597	.938	16	.327
	Group C	.638	-.381	.929	16	.235
ReceptiveIPT	Group A	-1.065	.131	.917	16	.149
	Group B	-.590	-.626	.891	16	.059
	Group C	.511	-.052	.950	16	.489
ProductiveDPT	Group A	-.174	.135	.972	16	.877
	Group B	.472	-1.158	.887	16	.051
	Group C	.474	-.597	.947	16	.445
ReceptiveDPT	Group A	-.929	.680	.903	16	.090
	Group B	-.632	-.197	.942	16	.377
	Group C	.491	-.813	.925	16	.200

Table 4 provides the descriptive statistics for the immediate and delayed productive and receptive posttests.

Table 4: Descriptive Statistics

Posttest (DV)	Methods of retrieval (IV)	Mean	Std. Deviation	N
ProductiveIPT	Group A	5.0000	.70711	16
	Group B	3.2813	1.11009	16
	Group C	2.5625	1.13835	16
	Total	3.6146	1.42635	48
ReceptiveIPT	Group A	4.9688	.64469	16
	Group B	5.0313	.78462	16
	Group C	3.8750	1.13284	16
	Total	4.6250	1.01321	48
ProductiveDPT	Group A	3.7812	1.34125	16
	Group B	2.1250	1.51107	16
	Group C	1.8125	1.36473	16
	Total	2.5729	1.63079	48
ReceptiveDPT	Group A	4.6250	1.16190	16
	Group B	4.4687	1.14701	16
	Group C	3.4688	1.08733	16

Total	4.1875	1.22312	48

The analysis of multivariate tests (table 5) shows that there was a statistically significant difference in participants' performance in immediate posttest based on the method of retrieval, F (8, 84) = 7.50, $p < .0005$; Wilk's $\Lambda = 0.34$, and partial $\eta^2 = .42$.

Table 5: Multivariate Tests[a]

	Effect	Value	F	Hypothesis df	Error df	Sig	Partial Eta Squared	Noncent. Parameter	Observed Power[d]
Intercept	Pillai's Trace	.98	419.092[b]	4.00	42.00	.00	.98	1676.37	1.00
	Wilks' Lambda	.02	419.092[b]	4.00	42.00	.00	.98	1676.37	1.00
	Hotelling's Trace	39.91	419.092[b]	4.00	42.00	.00	.98	1676.37	1.00
	Roy's Largest Root	39.91	419.092[b]	4.00	42.00	.00	.98	1676.37	1.00
method	Pillai's Trace	.79	7.038	8.00	86.00	.00	.40	56.31	1.00
	Wilks' Lambda	.34	7.495[b]	8.00	84.00	.00	.42	59.96	1.00
	Hotelling's Trace	1.55	7.944	8.00	82.00	.00	.44	63.55	1.00
	Roy's Largest Root	1.24	13.299[c]	4.00	43.00	.00	.55	53.20	1.00

a. Design: Intercept + method
b. Exact statistic
c. The statistic is an upper bound on F that yields a lower bound on the significance level.
d. Computed using alpha = ,05

Analyzing between subject effects shows that, in the case of immediate posttests, method of retrieval has a statistically significant effect on both productive (F (2, 45) = 25.10; $p < .0005$; partial $\eta^2 = .53$) and receptive posttest scores (F (2, 45) = 6.77; $p < .005$; partial $\eta^2 = .28$). Method of retrieval also has an effect on delayed productive (F (2, 45) = 17.91; $p < .005$; partial $\eta^2 = .29$) and delayed receptive posttest scores (F (2, 45) = 6.29; $p < .05$; partial $\eta^2 = .18$). In the case of delayed posttest scores, the results are significant at the .05 level of significance but not at the .01 level of significance ($p = .01$).

Tukey's HSD was conducted so that Group multiple comparisons could be made. Table 6, which represents the results of Tukey's HSD, shows that mean scores of productive immediate posttest were statistically significantly different between Group A and Group B ($p < .0005$), and Group A and Group C ($p < .0005$), but not between Group B and Group C ($p = .19$). Mean receptive immediate posttest scores were statistically significantly different between Group A and Group C ($p < .005$), and Group

B and Group C (p = .005), but not between Group A and Group B (p = .97).

Mean scores of productive delayed posttest were statistically significantly different between Group A and Group B (p =.005), and Group A and Group C (p < .005), but not between Group B and Group C (p = .81). Mean receptive delayed posttest scores were significantly different between Group A and Group C (p = .01), and Group B and Group C (p = .04), but not between Group A and Group B (p = .92).

Table 6: Tukey HSD for Multiple Comparisons

Dependent Variable	(I) method	(J) method	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
ProductiveIPT	Group A	Group B	1.72*	.36	.00	.86	2.58
		Group C	2.44*	.36	.00	1.58	3.30
	Group B	Group A	-1.72*	.36	.00	-2.58	-.86
		Group C	.72	.36	.12	-.14	1.58
	Group C	Group A	-2.44*	.36	.00	-3.30	-1.58
		Group B	-.72	.36	.12	-1.58	.14
ReceptiveIPT	Group A	Group B	-.06	.31	.98	-.82	.69
		Group C	1.09*	.31	.00	.34	1.85
	Group B	Group A	.06	.31	.98	-.69	.82
		Group C	1.16*	.31	.00	.40	1.91
	Group C	Group A	-1.09*	.31	.00	-1.85	-.34
		Group B	-1.16*	.31	.00	-1.91	-.40
ProductiveDPT	Group A	Group B	1.66*	.50	.01	.45	2.86
		Group C	1.97*	.50	.00	.76	3.18
	Group B	Group A	-1.66*	.50	.01	-2.86	-.45
		Group C	.31	.50	.81	-.89	1.52
	Group C	Group A	-1.97*	.50	.00	-3.18	-.76
		Group B	-.31	.50	.81	-1.52	.89
ReceptiveDPT	Group A	Group B	.16	.40	.92	-.81	1.13
		Group C	1.16*	.40	.02	.19	2.13
	Group B	Group A	-.16	.40	.92	-1.13	.81
		Group C	1.00*	.40	.04	.03	1.97
	Group C	Group A	-1.16*	.40	.02	-2.12]3	-.19
		Group B	-1.00*	.40	.04	-1.97	-.03

Based on observed means.

The error term is Mean Square (Error) = 1.283.

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

Investigating Research Question 2

The second research question deals with the vocabulary learning scores of the participants in the immediate and the delayed posttests in both receptive and productive directions. To answer this question, four independent sample t-tests were conducted at the .05 level of significance. The analysis of the gathered data showed that at the .05 level of significance, in Group A, there was a significant difference in the scores for productive IPT (M= 5, SD=.71) and productive DPT (M=3.78, SD=1.34); t(30)=3.22, p < .005" but not a significant difference in the scores for receptive IPT (M= 4.96, SD=.64) and receptive DPT (M=4.62, SD=1.16); t(30)=1.03, p = .31". For Group B, there was a significant difference in the scores for productive IPT (M= 3.28, SD=1.11) and productive DPT (M=2.12, SD=1.51); t(30)=2.47, p = .01" but not a significant difference in the scores for receptive IPT (M= 5.03, SD=0.78) and receptive DPT (M=4.46, SD=1.14); t(30)=1.65, p = .11. For Group C, there was not any significant difference in the scores for productive IPT (M= 2.56, SD=1.14) and productive DPT (M=1.81, SD=1.36); t(30)=1.69, p = .10 as well as in the scores for receptive IPT (M= 3.87, SD=1.13) and receptive DPT (M=3.46, SD=1.09); t(30)=1.04, p = .30"

DISCUSSION

In the present paper, the focus of attention has been on finding the optimal retrieval direction for effective learning of words. We have also considered the consequences of time interval on participants' performance in each direction. To our knowledge, this is the first study to deal with vocabulary retrieval in productive and receptive direction in the context of reading passage.

Totally, both experimental groups significantly outperformed the control group in both productive and receptive posttests that is consistent with Barcroft's (2015), Carpenter et al.'s (2008), and Karpicke and Smith's (2012) findings regarding the positive effect of retrieval on vocabulary learning.

The first research question in this study asked about the effect of productive and receptive retrieval on students' productive and receptive vocabulary knowledge. Statistical analysis of the posttest scores showed a

significant effect of retrieval method on participants' performance in productive and receptive posttests. Participants who read the passage with productive retrieval opportunity outperformed in both immediate and delayed productive posttests. However, there was not a significant difference in the performance of the group with receptive retrieval opportunity and the group with no-retrieval opportunity in immediate and delayed productive posttests. This is partially consistent with the findings of Mondria and Wiersma (2004) who state that productive retrieval promotes larger gains in productive learning and consistent with the findings of Nakata (2016) who found that recall formats were more effective than recognition for gaining the productive knowledge of words. The results are also in good agreement with the findings of Webb (2009) who showed that productive learning of word pairs was more effective than receptive learning. Therefore, it seems fair to suggest that productive retrieval of vocabulary during reading will result in better performance in productive vocabulary test. The results of this study revealed that participants who were provided with receptive retrieval opportunity during reading outperformed the control group in both immediate and delayed receptive posttests. More specifically, the results indicated that productive retrieval during reading will result in better both productive and receptive performance, but receptive retrieval during reading will result in only better receptive performance. This finding is partially in contrast with Steinel, Hulstin, and Steinel's (2007) study who found that receptive retrieval resulted in better performance in receptive vocabulary knowledge while productive retrieval brought about better performance in productive vocabulary knowledge. As stated earlier, the results of the current study showed that productive learning of words resulted in better performance in immediate posttest in both productive and receptive modes. This difference in findings can be due to the effect of contextualization, as Steinel et al. (2007) investigated the direction of testing in a paired associate paradigm but in the current study, retrieval direction is examined in the context of reading passage.

With regard to the second research question about the differences in vocabulary learning scores between the immediate and the delayed posttests, the findings were quite unexpected and provided confirmatory evidence that the participants' performance in experimental groups declined

significantly in delayed productive posttest compared with immediate productive posttest. However, there was not any significant difference in immediate and delayed receptive posttests for the experimental groups. Time effect was not significant for the control group in productive nor receptive delayed and immediate posttests. An important implication of these findings is that none of the retrieval directions, productive or receptive, can be considered as a guarantee for long-term productive learning of vocabulary.

CONCLUSION AND IMPLICATIONS

The main purpose of this paper was to give a comprehensive account of the effect of the productive and receptive retrieval directions in the context of reading passage on productive and receptive word knowledge. The present study also demonstrated the time effect on performance in each direction. Summing up the results, five important conclusions can be drawn from this study. First, participants who retrieve productively during reading benefit from both productive and receptive learning in immediate posttest. Second, performance of the control group is only slightly and insignificantly different in delayed and immediate posttests. Third, both experimental groups outperformed the control group in all four conditions of immediate receptive, delayed receptive, immediate productive, and delayed productive vocabulary tests. Fourth, none of the retrieval directions, productive or receptive, can be considered as a guarantee for long-term productive learning of vocabulary. And finally, receptive retrieval during reading will result in only better receptive performance.

The main limitation of the present study is the number of target words that were investigated. The question is whether it is reliable to conclude the obtained conclusions just relying on the result of six words learnt. Thus, it seems that the study needs to be repeated by different and more words. There are other limitations that, based on them, future research suggestions are provided. Clearly, further research will be required to investigate the effect of retrieval on different aspects of word knowledge. In the current study, the focus was only on receptive and productive aspects. Researchers have presented different aspects of word knowledge. Warren (2013), for example, states four different aspect for vocabulary knowledge including "semantic, syntactic, phonological and orthographic" (p. 39). Webb (2007),

in his study, focused on five aspects including "grammatical functions, syntagmatic association, paradigmatic association, orthography, and meaning and form" (p. 63). Another suggestion is putting under test all four retrieval directions. As it is mentioned earlier in this paper, four different retrieval directions have been presented in the literature. The present study has investigated the effect of only two directions (receptive recall and productive recall). Consequently, more experiments will be needed to verify the effect of two other retrieval directions (receptive recognition and productive recognition).

The proposed method can be readily used in practice for teaching new words during actual reading activity. The results are applicable to curriculum so that, according to the instructional goals, appropriate retrieval direction can be designed.

References

Barcroft, J. (2007). Effects of opportunities for word retrieval during second language vocabulary learning. *Language Learning*, *57*(1), 35-56.

Barcroft, J. (2015). Can retrieval opportunities increase vocabulary learning during reading? *Foreign Language Annals*, *48*(2), 236-249.

Bjork, R. A. (1994). Memory and metamemory considerations in the training of human beings. In J. Metcalfe, & A. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 185–205). Cambridge: MIT Press.

Campos, A., Rodríguez-Pinal, M. D., & Pérez-Fabello, M. J. (2014). Receptive and productive recall with the keyword mnemonics in bilingual students. *Current Psychology*, *33*(1), 64-72.

Carpenter, S. K. (2012). Testing enhances the transfer of learning. *Current directions in psychological science*, *21*(5), 279-283.

Carpenter, S. K., Pashler, H., Wixted, J. T., & Vul, E. (2008). The effects of tests on learning and forgetting. *Memory & Cognition*, *36*, 438–448. doi: 10.3758/MC.36.2.438

Ellis, R. (2006). Researching the effects of form-focused instruction on L2 acquisition. *AILA Review*, *19*(1), 18-41.

Ellis, R. (2016). Focus on form: A critical review. *Language Teaching Research*, *20*(3), 405-428.

Endres, T., Carpenter, S., Martin, A., & Renkl, A. (2017). Enhancing learning by retrieval: Enriching free recall with elaborative prompting. *Learning and Instruction*, *49*, 13-20.

Finn, B., Roediger, H. L., & Rosenzweig, E. (2012). Reconsolidation from

negative emotional pictures: Is successful retrieval required? *Memory & Cognition*, *40*(7), 1031-1045.

Fritz, C. O., Morris, P. E., Acton, M., Voelkel, A. R., & Etkind, R. (2007). Comparing and combining retrieval practice and the keyword mnemonic for foreign vocabulary learning. *Applied Cognitive Psychology*, *21*(4), 499-526.

Gamage, J., Mathew, T., & Weerahandi, S. (2004). Generalized p-values and generalized confidence regions for the multivariate Behrens–Fisher problem and MANOVA. *Journal of Multivariate Analysis*, *88*(1), 177-189.

George, D., & Mallery, M. (2010). *SPSS for Windows step by step: A simple guide and reference*. Boston, MA: Pearson.

Goossens, N. A., Camp, G., Verkoeijen, P. P., & Tabbers, H. K. (2014). The effect of retrieval practice in primary school vocabulary learning. *Applied Cognitive Psychology*, *28*(1), 135-142.

Karpicke, J. D., & Smith, M. A. (2012). Separate mnemonic effects of retrieval practice and elaborative encoding. *Journal of Memory and Language, 67*, 17–29. doi:10.1016/j.jml.2012.02.004

Liu, Y., Rosburg, T., Gao, C., Weber, C., & Guo, C. (2017). Differentiation of subsequent memory effects between retrieval practice and elaborative study. *Biological Psychology*, *27*, 134-147.

Migueles, M., & García-Bajos, E. (2007). Selective retrieval and induced forgetting in eyewitness memory. *Applied Cognitive Psychology*, *21*(9), 1157-1172.

Mondria, J.A., & Wiersma, B. (2004). Receptive, productive, and receptive + productive L2 vocabulary learning: What difference does it make? In P. Bogaards & B. Laufer (Eds.), *Vocabulary in a second language: selection, acquisition and testing* (pp. 79-100). Amsterdam: Benjamins.

Nakata, T. (2016). Effects of retrieval formats on second language vocabulary learning. *International Review of Applied Linguistics in Language Teaching*, *54*(3), 257-289.

Pyc, M. A., & Rawson, K. A. (2009). Testing the retrieval effort hypothesis: Does greater difficulty correctly recalling information lead to higher levels of memory? *Journal of Memory and Language*, *60*(4), 437-447.

Rovai, A. P., Baker, J. D., & Ponton, M. K. (2013). *Social science research design and statistics: A practitioner's guide to research methods and IBM SPSS*. Chesapeake, VA: Watertree Press LLC.

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. http://dx.doi.org/10.1037/a0037559.

Steinel, M. P., Hulstijn, J. H., & Steinel, W. (2007). Second language idiom learning in a paired-associate paradigm: Effects of direction of learning,

direction of testing, idiom imageability, and idiom transparency. *Studies in Second Language Acquisition*, *29*(3), 449-484.
Thompson, T., & Simmons, N. (2010). *Family and friends 3*. Oxford: Oxford University Press.
Toppino, T. C., & Cohen, M. S. (2009). The testing effect and the retention interval: Questions and answers. *Experimental Psychology*, *56*(4), 252-257.
Warren, P. (2013). *Introducing psycholinguistics*. Cambridge: Cambridge University Press.
Webb, S. (2007). Learning word pairs and glossed sentences: The effects of a single context on vocabulary knowledge. *Language Teaching Research*, *11*(1), 63-81.
Webb, S. (2009). The effects of receptive and productive learning of word pairs on vocabulary knowledge. *RELC journal*, *40*(3), 360-376.
Zhong, H. F. (2016). The relationship between receptive and productive vocabulary knowledge: A perspective from vocabulary use in sentence writing. *The Language Learning Journal*, *6*(11), 1-17.

APPENDICES

Appendix 1: Reading Passage with Productive Retrieval Opportunity

The Prince and the Pauper

Tom Canty was a boy from a very poor family. His clothes were old and dirty and he never had enough food. However, Tom believed that he could have a better life one day. He was **clever (باهوش)** and he knew how to read and write. He wanted to get a good job and earn enough money to have a house, clothes and good food.

One day, Tom was outside the palace **gates (دروازه ها)** when he saw Prince Edward. Tom stood and watched the prince for a while. The **guards (نگهبانان)** who had **armor (لباس رزم)** tried to send Tom away, but the Prince stopped them. The two boys were very surprised when they saw each other. They had very different lives, but they looked just the same. They had the same eyes, the same hair and the same faces. They were both **…….. (باهوش)** boys. They were like **twins (دوقلوها)**. The only thing that was different was their clothes. Prince Edward invited Tom into the palace and the two boys talked about their lives. Edward had lots of money, fine clothes, jewels and plenty of food, but he had to stay in the palace and be quite and good all the time. Tom had nothing, but he could play outside the

palace …….. (دروازه ها), in the street with other boys and meet lots of interesting people without the …… (نگهبانا).

"I want to be like you," Tom said to Edward, "you are rich and you can have everything you want." "I want to be like you," Edward said to Tom. "You are free and you can do everything you want." "Well, we look like each other like …….. (دوقلوها), and we are the same age," said Edward. "Let's **swap** (عوض کردن) clothes. You can stay here and I can live with your family for a while." Tom thought this was a wonderful idea. The two boys started to …….. (عوض کردن) their clothes.

"We are like …….. (دوقلوها)," Tom laughed. Edward was …….. (باهوش). Before leaving the palace, he hid a wax disc in a suit of ……. (لباس رزم), then he left the palace quickly, before the …… (نگهبانان) found the two boys together. Soon Edward was with Tom's family, but it was not much fun outside the palace …….. (دروازه ها). Tom's father was unkind. He was often angry and he shouted at Edward all the time. Edward wanted to leave. He ran away and met a soldier called Miles Hendon. He had not ……. (لباس رزم). He was a kind man and he looked after Edward. They had lots of adventures together. Edward didn't really like life outside the palace. He wished he did not …….. (عوض کردن) his clothes. He saw that life was very difficult for poor people. He decided to be a good king and help people of England.

Appendix 2: Reading Passage with Receptive Retrieval Opportunity

The Prince and the Pauper

Tom Canty was a boy from a very poor family. His clothes were old and dirty and he never had enough food. However, Tom believed that he could have a better life one day. He was (باهوش) **clever** and he knew how to read and write. He wanted to get a good job and earn enough money to have a house, clothes and good food.

One day, Tom was outside the palace (دروازه ها) **gates** when he saw Prince Edward. Tom stood and watched the prince for a while. The (نگهبانان) **guards** who had (لباس رزم) **armor** tried to send Tom away, but the Prince stopped them. The two boys were very surprised when they saw each other. They had very different lives, but they looked just the same. They had the same eyes, the same hair and the same faces. They were both

(……) clever boys. They were like **(دوقلوها) twins**. The only thing that was different was their clothes. Prince Edward invited Tom into the palace and the two boys talked about their lives. Edward had lots of money, fine clothes, jewels and plenty of food, but he had to stay in the palace and be quite and good all the time. Tom had nothing, but he could play outside the palace **(…….) gates**, in the street with other boys and meet lots of interesting people without the **(……..) guards**.

"I want to be like you," Tom said to Edward, "you are rich and you can have everything you want." "I want to be like you," Edward said to Tom. "You are free and you can do everything you want." "Well, we look like each other like (…….) **twins**, and we are the same age," said Edward. "Let's **(عوض کردن) swap** clothes. You can stay here and I can live with your family for a while." Tom thought this was a wonderful idea. The two boys started to **(……) swap** their clothes.

"We are like (……..) **twins**," Tom laughed. Edward was (……) **clever**. Before leaving the palace, he hid a wax disc in a suit of **(…….) armor**, then he left the palace quickly, before the **(……..) guards** found the two boys together. Soon Edward was with Tom's family, but it was not much fun outside the palace (……) **gates**. Tom's father was unkind. He was often angry and he shouted at Edward all the time. Edward wanted to leave. He ran away and met a soldier called Miles Hendon. He had not (……..) **armor**. He was a kind man and he looked after Edward. They had lots of adventures together. Edward didn't really like life outside the palace. He wished he did not **(…….) swap** his clothes. He saw that life was very difficult for poor people. He decided to be a good king and help people of England.