THE IMPACT OF CRITICAL THINKING ON LEARNERS' EFL VOCABULARY RETENTION: THE ARAB CONTEXT

Tariq Elyas
European Language Department, Faculty of Arts & Humanities,
King Abdulaziz University, Saudi Arabia
E-mail: telyas@kau.edu.sa

Budor S. Al-Zahrani
English Language Centre, University of Jeddah, Jeddah, Saudi Arabia
E-mail: telyas@kau.edu.sa

Abstract: The aim of this experimental study was to investigate the impact of a critical thinking based pedagogical approach on learners' retention ability of EFL vocabulary. There were two groups in the study: an experimental and a control group. The experimental group received vocabulary instruction with a critical thinking-based approach whereas the control group was taught in the conventional methods of vocabulary teaching. The treatment consisted of seven lessons delivered over 12 weeks. At the end of the treatment, the immediate post-test was administered and two weeks after the immediate post-test, the delayed post-test was administered. The resulting data was subjected to descriptive and inferential statistical analysis using SPSS software. The results showed that the experimental group outperformed the control group in their vocabulary retention. Based on these results, the researchers concluded that the adopted critical thinking-oriented language instruction to EFL vocabulary has a positive effect on vocabulary retention, with implications for policy-makers, teacher trainers and teachers.
Keywords: *Arab context; English as a Foreign Language; critical thinking; memory; retention; vocabulary.*

INTRODUCTION

Defining critical thinking is by no means an easy task. Despite many definitions and conceptualizations of critical thinking, "there is still no universal consensus on a definition of critical thinking amongst educators, philosophers and psychologists in the field" (Ab Kadir, 2007, p. 1). One can claim, "critical thinking is a liberating force in education and a powerful resource in one's personal and civic life. While not synonymous with good thinking, critical thinking is a pervasive and self-rectifying human phenomenon" (Facione, 1990, p. 3). Scriven and Paul (2003) define critical thinking as "the intellectually disciplined process of actively and skill- fully conceptualizing, applying, analyzing, synthesizing, and/or evaluating information gathered from, or generated by, observation, experience, reflection, reasoning, or communication, as a guide to belief and action" (p. 1).

Powerful strategies to teach, learn and think critically is learning "how to ask and answer questions of analysis, synthesis and evaluation" (Paul, 1992, p. 521). Strategies adopted for data analysis in this paper is based to develop critical thinking skills include Socratic questioning, which is "at the heart of critical teaching" (Paul, 1992, p.360). Developing curricula to include critical thinking is also necessary (Al-Gabrey, 2007). Some views also assume critical thinking as innate to all humans, developed through the "natural developmental process" (Judith *et al.*, 1985, p. 492).

For many L2 learners, learning vocabulary is tedious because it means to them memorizing lists of words and using their bilingual dictionaries to resolve their

lexical gap (Read, 2000). However, vocabulary learning is more than just learning individual words and their meanings in isolation (Cook, 2008). It is a multifaceted process which involves many factors in addition to memorizing words, such as being able to recall them, and, most importantly, using them appropriately (Nation, 2004; Verhallen & Schoonen, 1993; Elyas & Shah, 2018). According to Webb & Chang (2012), certain activities, which are designed around using target words, are:

> Activities designed around using target words are also necessary because they help learners to develop productive knowledge of the items. Critical to the effectiveness of vocabulary development through productive activities is ensuring that the target vocabulary will be used rather than ignored or replaced with a more frequent synonym (p. 658).

Just as the size of vocabulary knowledge matters, similarly, the quality and the depth of this knowledge is equal if not far more important. Initial encoding of new lexical items, by means of memorization for instance, and repeated encounters are not sufficient to support fluent language use; students must be able to reliably store the new items in long-term memory and successfully retrieve them when needed (Golonka, et al., 2012). For instance, Webb and Chang (2015) examined English as a Foreign Language (EFL) learners studying two levels of texts (level 1 and level 2) and found that students' learning gains were quite similar. Theoretically, reading an easier text should enable a student to acquire more words than reading a more difficult one because easier texts tend to lead to higher level of comprehension.

Research has shown that considerable learning can occur within a short period. For example, Webb (1962) found that participants were able to learn from 33 to 166 word pairs per hour over a four hour period. Cobb and Horst (2001) found that 140-180 words were learned in two months through one hour a week of concordance-based learning. However, few studies have investigated L2 vocabulary growth over longer periods of time. Milton (2006) examined the difference in vocabulary size between French foreign language learners at seven grade levels. He found that about 170 words were learned per year when students received between 58.5-78 hours of tuition and approximately 530 words were learned per year when students received 117-175.5 hours of tuition. Clark and Ishida (2005) found that there was no significant difference on the Vocabulary Levels Test (VLT) scores after one semester of study for English as a second language students (see also Nation, 1983, 1990; Schmitt, Schmitt, & Clapham, 2001). Nation (2006) found that the most frequent 1,000 word families in the British National Corpus (BNC) accounted for just over 77.86% of the words in the LOB corpus, the second most frequent 1,000 word families made up 8.23%, and the third most frequent 1,000 word families accounted for 3.70% of the corpus. According to Houston (2001), retention is intertwined with acquisition (as cited in Sanatullova-Allison, 2014). Vocabulary retention, which is the ability to recall the meanings of words and actively use the words days or weeks after they have been processed (Schuetze, 2014), is a perquisite to their acquisition process. Successful vocabulary learning requires an efficient encoding and storage of words with chances of retrieval when needed as a result of an extended process of practice (Golonka, *et al.,* 2012).

Our first research question seeks to shed light on the number of words that are learned by EFL learners each year over four years of instruction. One view on vocabulary learning and teaching is that retention depends "on the amount of mental and emotional energy used in processing a word" (Hedge, 2000, p. 118). Given that attention, practice, deeper processing, and linking to existing knowledge are all necessary to promote chances of vocabulary retention, the purpose of this research is to investigate the effectiveness of a critical thinking (hereafter referred to as CT) based

approach in developing the retention of vocabulary ability of EFL Saudi university students. This paper aims to answer the research question: "What is the effect of a CT-based pedagogical approach on vocabulary retention of EFL Saudi university students?"

Finding "a concrete definition of CT on which most researchers can agree remains elusive" (Lai, 2011, p. 41). However, a plethora of definitions have been proposed by experts and researchers in the fields of philosophy, cognitive psychology and in education. One of the most famous definitions of CT argues that CT is a "purposeful, self-regulatory judgment that results in interpretation, analysis, evaluation, and inference, as well as an explanation of the evidential, conceptual, methodological, criteriological, or contextual considerations upon which that judgment is based" (Facione, 1990, p. 3). This study adopted the definition that "CT is the intellectually disciplined process of actively and skillfully conceptualizing, applying, analyzing, synthesizing, and/or evaluating information gathered from, or generated by, observation, experience, reflection, reasoning, or communication, as a guide to belief and action" (Scriven & Paul, 2004, p.1).

Defining CT based on Bloom et al.'s taxonomy, as in the latter definition, is recognized as 'the educational approach' (Lai, 2011). This definition combines higher-order thinking skills and the often missing component of CT, reflection; which is a "meaning-making process" (Rodgers, 2001, p. 843). In CT-based language instruction, learners are required to put forth more cognitive and linguistic effort by requiring them to do problem solving tasks, where they compose pieces of writings, or induced to participate in discussions, make opinions or provide justifications and so forth. Such challenging activities are better described as 'desirable difficulties'. Desirable difficulties effect is a learning principle that activities that increase the level of learning difficulty initially can result

in better long-term term retention of information (Golonka, *et al.,* 2012).

When examining the literature of L2 vocabulary teaching and learning, few studies are found compared to other skills of language. This apparent gap has been noted by Saudi scholars who claimed that linguists rarely addressed this issue (Al-Zahrani & Elyas, 2017). Similarly, not enough studies have tackled the issue of vocabulary learning and retention in EFL classrooms in the Saudi context. For example, Al-Masrai and Milton (2012) discussed the aspect of vocabulary learning among EFL Saudi learners. Al-Hadlaq (2003), investigated the influence of different teaching activities on EFL Saudi vocabulary retention (as cited in Golonka, *et al.,* 2012).

Lately, one of the proposed questions is: "What strategies do learners use to acquire new words or to retain them?" (Hedge, 2000, p. 111). Similarly, taking into consideration the lack of a sufficient exposure to English outside classrooms, another question can be added which relates to the pedagogy is: What strategies can EFL teachers implement in order to facilitate vocabulary acquisition and retention? Amen's (2008) study analyzed critical thinking and achievement and retention among second year high school students in history through role switching. Her findings showed that it is necessary for curriculum planners to devise courses that will develop critical thinking in students. Furthermore, the study also revealed in order for teachers to develop critical thinking skills in students, they need to undertake courses in teaching strategies for this purpose. Rafi (2011) states that, "learners may become proficient in English language if they are motivated and taught how to display CT in English language usage" (p. 107). However, the relationship between engaging the students' CT and L2 vocabulary learning has not yet been described clearly. This study aims at implementing a CT-based approach to teach EFL vocabulary in order to enhance students' retention abilities. Thus, the present study can be classified as unique to

some extent because it is the first study, to the best of the researchers' knowledge, that relates CT to vocabulary retention locally, and among the very few that do so globally.

Several studies explored the impact of some of CT techniques or strategies on the development of an EFL learner's lexicon. Some of these studies focused on techniques including discussions and dialogues (Webb, & Chang, 2012; 2015). For instance, Shore, Ray, and Gooklasian (2015) investigated the relationship between science vocabulary and the use of some cognitive science principles, particularly, drawing pictures and talking about the definitions of the lexical items. The study compared outcomes of such practices to those of the common practice of copying the definitions with no interaction. Post-test results showed significant improvement in vocabulary retention.

This research intends to form a preliminary statement about the relationship between CT and vocabulary retention in foreign language teaching and learning, both of which are areas of cognitive psychology research. As cognitive psychology involves the study of such mental processes, its findings as well as findings of other fields such as linguistics are applied to obtain a deeper understanding of the ideas under investigation in the current study. The relevant theories selected for this purpose, clarified respectively, are: the *Multi-Store Model of Memory*, the *Levels of Processing Theory* and the *Generative Learning Theory*.

Memory is one construct that is greatly involved with learning vocabulary. It refers to the mental processes of retaining information for later use and retrieving. Atkinson and Shiffrin's (1968) Multi-Store Model is among the most influential in literature on memory. It has a three-scale processing model of memory: sensory memory, short-term memory (later becomes working memory), and long-term memory, as demonstrated in Figure 1. According to this model, it is suggested that we receive the information through our senses to be held briefly in our sensory memory. With attention and selection, the information is held temporarily in our short-term memory. This process of information entering into the short-term memory is called 'working memory'. Retention of vocabulary occurs when they are transferred from the working memory to the long-term memory in a process known as encoding.
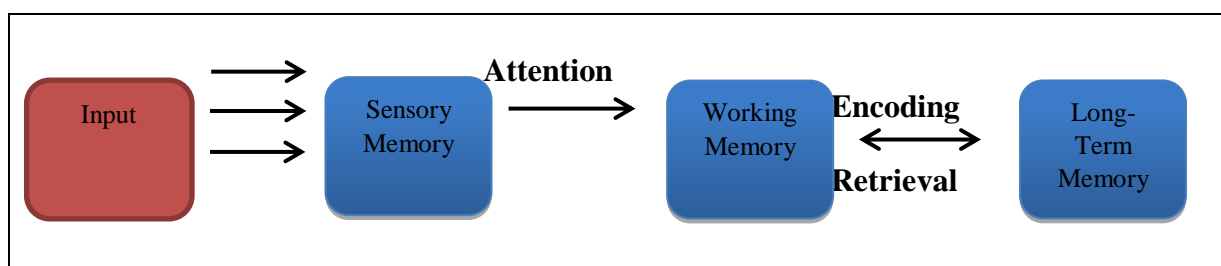

Figure 1. *The memory multi-store model*

Craik and Lockhart (1972) suggested a framework that attempts to explain what alters an input to become encoded and transferred to the long-term memory. According to their Levels of Processing Framework, a word is instantly processed at different levels, and its possibility to remain in the long-term memory depends on the "depth of processing". In other words, a learner needs to consider the word while an elaborative rehearsal of that word occurs. A learner shall spend more time and cognitive effort on absorbing that information. As deeper processing of a word takes place, the word is internalized and stored in the long-term memory for future retrieval. Nation (n.d.) while indicating the significance of this theory states that, "in learning, the amount of efforts is not that important; what is important is the quality of activity in the brain" (as cited in Schmitt, 1995, p. 5). Knowledge without any deep processing, for

example, learning words from lists, will be easily forgotten" (Yu, 2011).

Another cognitively oriented hypothesis is the Generative Theory which was proposed by Wittrock (1974) to promote long-term vocabulary retention through merging new knowledge with old. Using target words to formulate sentences where the semantic connection is established between the new words and other words or linguistic knowledge is suggested as to permit longer retention. Wittrock stated that "although a student may not understand sentences spoken to him by his teacher, it is highly likely that a student understands sentences that he generates himself" (as stated by Lee, Lim, & Grabowski, 2008, p. 112). Joe (1998) also reported that generative tasks promoted the retention of target L2 vocabulary words. Similarly, according to Yu (2011), productive tasks (as sentence writing) are more likely to lead to a long term memory than rote repetition and learning words from lists by L1 translation.

METHOD

The aim of this research is to measure the effectiveness of CT on L2 vocabulary retention. An experimental study was designed to examine the effect of a vocabulary teaching intervention using a CT-based approach on EFL students' retention of newly-learned vocabulary. This experiment adopts a two-group pre and post-tests design where the independent variable was intervention *versus* no intervention (experimental *versus* control, respectively), while the dependent variable was student performance on assessments of vocabulary retention immediately after the intervention and also after an interval of delay. As the study sample was relatively homogenous in terms of factors such as proficiency level, age, and gender, the effect of such extraneous factors on the study's outcome may be expected to be minimal.

The study was conducted in a foundation year English language program at a university in Saudi Arabia. The aim of this program is to provide first year undergraduate students with a one-year, preparatory, intensive course in English language. The program has a quarter system and four levels, the last of which, level 104, brings them to an intermediate proficiency level. Students must pass this level as a prerequisite for continuing with their university studies. Developing the use of vocabulary is one of the main objectives of the language course and it is placed in clear contexts in the language curriculum. The mid and final tests and progress tests have complete sections that are devoted to vocabulary testing.

At the time of the study, the participants in this research were Saudi female students between the ages of 18-20 years in the preparatory year of the baccalaureate degree program at a Saudi university. A convenience sample was assembled for the study. Participants were drawn from four different classrooms at the foundation year at an intermediate level (104), which is equivalent to language proficiency level B1+ according to CEFR (Common European Framework of Reference for Languages). Students belonging to two classrooms constituted the experimental group while students in the other two classrooms constituted the control group. At the beginning, there were a total number of 96 students, but after removing those who missed a/ the post-test/s, 67 students remained (of which 34 were in the experimental group). Their ages ranged from 18 to 25 years old. The selection of the four classrooms was a function of instructor willingness to participate in the study and the availability of class time in which to run the study. The assignment of the classrooms to experimental or control groups was random.

This study made use of three different tests for different purposes. They were used at different stages of the research. One of these tests was administered before the intervention. The other two were administered on different days after the intervention. A detailed explanation of each

of these tests' design and purposes is provided below.

Prior to the intervention, a test was administered to all study participants in both the experimental and the control groups. This was a diagnostic test and its main purpose was to measure the participants' baseline familiarity with the targeted words. The pre-test was yet another mean to ensure homogeneity of experimental and control groups with respect to vocabulary knowledge.

The instrument used for generating this test was the Vocabulary Knowledge Scale (VKS). It was developed by Paribakht and Wesche (1993, 1997). It collects participant self-reports of vocabulary knowledge and assesses the depth or the quality of second language vocabulary knowledge. It can be used with any set of words (Read, 2000). It includes five categories or levels of knowledge and test-takers have to select the category that best reflects their knowledge about a particular word. This study adopted the version of Alghamdi (2004) which includes only three preliminary categories or levels of vocabulary knowledge. These levels, escalating in depth, are as follows:

I. I do not remember seeing this word before.
II. I have seen this word before, but I don't know what it means.
III. I have seen this word before, and I think it means.

The other two categories: Level IV (recognize the word and be able to give the meaning of the word), and Level V (be able to use the word in a sentence) were not employed for this study because they demanded higher levels of word knowledge than was required for this study. The target was to assess the participants' preliminary knowledge and familiarity with words that were new to them prior to the experiment. Participant self-reports of any minimal level of familiarity with a certain word was sufficient for excluding that word from the study. Thus, revealing more complex or deeper lexical knowledge was not necessary. A translation of the test was included in the participants' native language, Arabic, to ensure more valid reporting of their real vocabulary knowledge. A sample of the VKS is presented in Figure 4 in the English language version.

Immediately after all the seven sessions were completed and without further notice, the immediate post-test was administered to evaluate the participants' retention ability of the target vocabulary. After two weeks from the end of the intervention and also without further notice, the delayed post-test was administered. Both tests were objective tests that consisted of two sections and ten items. The first section includes multiple-choice tasks with three distractors in the response option for each item. The second section includes gap-filling tasks. Each post-test includes a different set of vocabulary. The immediate post-test is presented in Figure 5 as a sample. One of them was administered immediately after the intervention, and the other was administered after a delay of two weeks post-intervention. A comparison between the groups' results determined whether the intervention affected the vocabulary retention of the experimental group's participants at different times. Also, a comparison between the same group's results on both tests would indicate the change in the participants' retention ability over time. Administration of two post-tests instead of just one is advised when one has the goal of building a better and more complete idea of the impact of a certain treatment. The study aims to measure vocabulary learning, and specialists agree that it is integral to utilize a delayed post-test since immediate post-tests often overestimate learning (Waring & Takaki, 2003; as cited in Golonka, *et al.,* 2012). Sometimes, test results taken too soon after an intervention can trivialize or maximize an effect, while in a later test, the effect may have faded. Also, the researcher preferred to use achievement tests instead of using self-report tests similar to the pre-test. This was to avoid repetition.

Figure 2. *The immediate post-test*

The times, types and purposes of the tests used in the experiment are shown below in Table 1.

Table 1. *Details of test administration and purposes*

Test Time	Type	Purpose
Before the intervention	Diagnostic Pre-test	To remove words familiar to the participants from the experiment's targeted vocabulary.
After the intervention	Immediate post-test	To compare vocabulary retention abilities of the experimental and the control groups immediately after the intervention.
	Delayed post-test	To compare delayed vocabulary retention abilities of the experimental and control groups, four weeks post-intervention.
		To compare, within group, each group's immediate and delayed vocabulary retention ability.

The researcher ensured that the control group teachers did not implement teaching methods that might (clearly) promote their participants' CT abilities and that they adhered to the syllabus material. Conventional methods of vocabulary teaching were used, such as, synonym and definition, contrasts and opposites, translation and examples. Often the use of such methods was aided by an application that includes a gap-filling tasks. No type of technology was included in teaching the experimental group, in order to avoid any possible influences of extraneous factors on the outcome. The researcher taught seven sessions over the period of 12 weeks to each class. The allocated time for each session was around 50 minutes.

One of the researchers taught the experimental group. The experimental group experienced a teaching intervention for vocabulary involved use of a CT-based approach. The target vocabulary was derived from the English language curriculum course book following the pacing guide at the time the experiment. The target vocabulary includes 28 words and fixed expressions. The vocabulary teaching intervention materials were designed in accord with Scriven and Paul's (2004)

definition of CT that combines 'the educational approach' with the component of reflection. Bloom et al.'s taxonomy represents an efficient framework for incorporating CT in the educational system since it is more applicable to education (Abudawood, 2015). The general learning objectives were designed to tap the cognitive process involved in higher-order thinking skills drawn from the revised Bloom et al.'s taxonomy; these processes are, from the simplest to the most complex, to analyze, evaluate and create. The learning outcomes of each session's tasks targeted all the three higher-order thinking levels aided by a visual description of CT skills or process involved. A sample of the vocabulary teaching intervention materials designed followed a visual description of CT skills or process involved is presented in Figure 2 and 3.

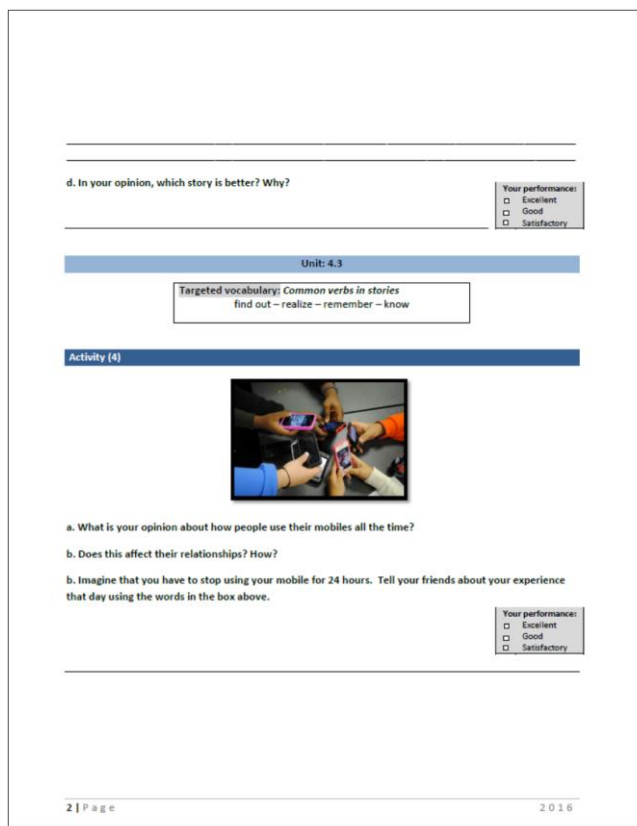Figure 3. *The Vocabulary teaching intervention materials*

Figure 4. *Vocabulary teaching intervention materials (Continued)*

The teaching materials contain challenging activities that are better described as 'desirable difficulties'. Students were required to put forth more cognitive and linguistic effort in doing problem solving tasks, or participate in discussions and debates, make opinions, provide justifications and so forth. In Figure 4, the sample contains an example of such activities. In some tasks they had to compose pieces of writings.
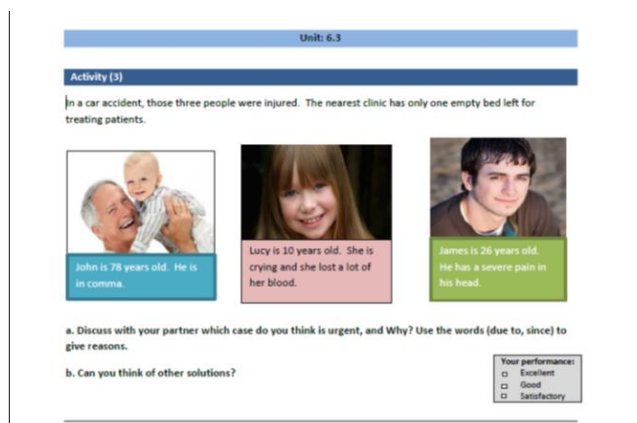

Figure 5. *The vocabulary teaching intervention materials*

The manner of vocabulary instruction was explicit instruction and the researchers posed challenging questions, invited various opinions, cast doubts or sought justification while facilitating the learning process. In addition, the tasks were designed in order to target the participants' reflective thinking. Students had to practice reflective thinking through the requirement that they relate the themes and ideas in the material to their

interests, experiences and social contexts or backgrounds. Multiple opportunities were created for them to review and apply their knowledge. Also, the students had to frequently monitor and evaluate their own progress or performance. Examples of opportunities where students were stimulated to practice such reflective thinking in relation to the activities are:

- Visualizing their day without using mobiles (Figure 1).

- Revising and assessing previous judgments based on newly presented information or facts.
- Assessing their attitude towards elderly people around them in Saudi contexts.
- Every time they have to evaluate their performance in the self-evaluation box after every activity; they have to self-assess and judge their own performance.

Figure 6. *The vocabulary knowledge scale*

RESULTS AND DISCUSSION
In this section, the data collected for this research are presented and analyzed. The aim is to identify the relationship between CT and L2 vocabulary retention. In order to analyze the data, statistical analyses using IBM SPSS statistics ver. 20.0 (Statistical Package for Social Sciences) was conducted.

The pre-test
The VKS was used as a pre-test. It was based on a one-point scale where the participants would attain a score of zero if they report total unfamiliarity with the target word or if they declared familiarity with the word but fail to provide its meaning correctly. In case of reporting the correct meaning of the word, they would attain a

score of one. The total possible score is 35 points (one X 35). This pre-test had two objectives:

1. The first objective was to delete familiar words from the target list of vocabulary that was based on their institute's pacing guide.
2. The second objective of administering the pre-test was to assess homogeneity of the experimental group and the control group.

To achieve the first objective, all participants' scores were coded into SPSS for locating words which 30% of participants declared as familiar. After calculating the students' scores, words which received 30% familiarity by the students were deleted from the study's tests. In order to calculate 30%

familiarity words, the following equation was run manually based on the number of the pre-test takers' in both groups (96 students) and their scores per each word:

$$The\ test\ takers\ \ total\ number \times 30 \div 100$$

$$96 \times 30 \div 100 = 28.8 \cong 29$$

Any word that received 29 points (statistically calculated as 28.8 but resided at 29) was deleted from the study's targeted vocabulary. The original number of the study's targeted words was 35 words, and the number of words that were left out after this procedure was 28 words (after deleting 7 words).

To achieve the second objective, which is assuring that both groups are homogenous, a comparison between both groups' pre-test scores using statistical tests was applied. As VKS serves as an indicator of the learners' depth of knowledge about words, it can also be an indicator of the test takers' language proficiency. However, before the comparison

was conducted, a numerical testing was run to assess the normality of the pre-test data using a Shapiro-Wilk Test. The significance value for this study was set at 0.05. The significance values of the experimental and the control groups scores were Sig. = 0.01, Sig. = 0.027 respectively, where p < 0.05. This rejects the null hypothesis of assumption of normality and indicates that the pre-test data was not approximately normally distributed which is expected in a test of untaught materials.

Accordingly, the non- parametric test, Mann-Whitney U Test, was used to analyze the pre-test scores of the two groups. As can be seen in Table 3, the analysis shows that Mann-Whitney U = 1028.000, while the two - tailed (Sig. = 0.361) is greater than 0.05 (p > 0.05). Hence, there is no statistically significant difference between the pre-test scores of the experimental group and the control group. Accordingly, it can be concluded that the two groups are homogenous.

Table 2. *Ranking for the mann-whitney U test of the pre-test scores*

Group	N	Mean Rank	Sum of Ranks
Experimental Group	47	51.13	2403.00
Control Group	49	45.98	2253.00
Total	96		

Table 3. *The mann-whitney u value and significance level of the pre-test scores*

Test Statistics	
	Pre-test
Mann-Whitney U	1028.000
Wilcoxon W	2253.000
Z	-.913
Asymp. Sig. (2-tailed)	.361

a. Grouping Variable: Group

The post-test

This study investigates the relationship between the two variables, CT and vocabulary retention. Thus, it attempts to test two hypotheses. The first hypothesis, which represents the null hypothesis (*H*0) of the present study, states the following:

- *The use of a CT-based approach to teach EFL vocabulary does not affect the EFL learners' vocabulary retention.*

The second hypothesis, which is the *alternative one* (*H*1), states the following:

- *The use of a CT-based approach to teach EFL vocabulary can enhance EFL learners' vocabulary retention.*

In order to accept or reject the study's null hypothesis, two post-tests were administered. One was an immediate post-test and the other was a delayed post-test (after two weeks). The calculation of post-tests scores was done as follows: for each correct answer,

participants received the score of one and for each incorrect answer, they received the score of zero. The overall score was 10 points (one-point x 10 words). Later, all participants' overall test scores were coded and were subjected to statistical analyses using SPSS.

The statistical analyses served two main purposes:

1. To compare between the results of the experimental group and the control group in both tests, the immediate post-test and the delayed post-test.

2. To explore any associations, if any existed, between the immediate and the delayed post-tests within each group, the experimental group and the control group.

To determine the type of the statistical test (parametric or non-parametric) to be used in order to compare the results of the experimental group and the control group in both post-tests, the normality of distribution

test was utilized twice, once for the immediate post-test data and again for the delayed post-test data. Later, to further understand and assess any changes in the same group's performance, a comparison between the same group's post-tests scores was conducted. Additionally, the Bivariate Pearson Correlation was applied in order to further explore any associations between both post-tests scores within each group.

The immediate post-test

When running the normality test, the significance values of the experimental and the control groups scores successively were sig. = 0.023, sig. = 0.050, ≤ 0.05. This indicates that the immediate post-test data of both groups were approximately normally distributed. Thus, the means of the scores represented a reliable measure and the parametric test, the independent t-test, was used to compare between the means of both group.

Table 4. *The independent t-test for the immediate post-test scores*

Group Statistics					
	Group	N	Mean	Std. Deviation	Std. Error Mean
Immediate test	Post-Experimental Group	34	7.00	2.015	.346
	Control Group	33	4.91	1.809	.315

The independent t-test (Table 4) revealed that the mean score of the experimental group (M = 7.00, SD = 2.015) is statistically higher (*t* = 4.472, *df* = 64.614), (Sig. 2 - tailed = 0.000) than the control group (M =

4.91, SD = 1.809). Thus, there is a highly significant difference between the immediate post-test means of both groups. The mean difference was (2.091) with [95% CI, 1.156 to 3.026] (see Table 5).

Table 5. *The Levene test for the equality of variance between the immediate post-test scores of the experimental and the control groups*

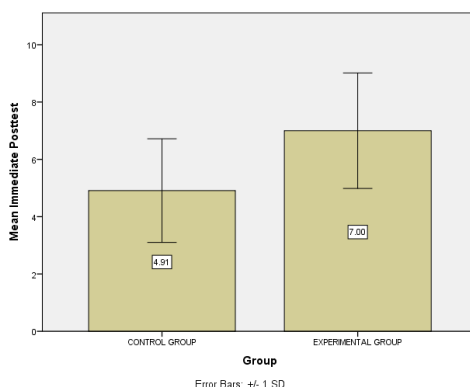| Independent Samples Test		Levene's Test for Equality of Variances		t-test for Equality of Means					95% Confidence Interval of the Difference	
		F	Sig.	T	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Immediate Post-test	Equal variances assumed	1.092	.300	-4.465-	65	.000	-2.091-	.468	-3.026-	-1.156-
	Equal variances not assumed			-4.472-	64.614	.000	-2.091-	.468	-3.025-	-1.157-

Figure 7. *The means of the results of the immediate post-test of the experimental and the control groups*

The delayed post-test

For the delayed post-test scores, the data was not approximately normally distributed. Numerically, according to the normality test, the significance level for the experimental group was p = 0.001 (p < 0.05). Graphically, the experimental group's test takers' scores created a positively skewed distribution (See Figure 7). All the experimental group scores were above the half score of the test (5), where all the scores range between (6) as the minimum score and the full score (10) as the maximum.
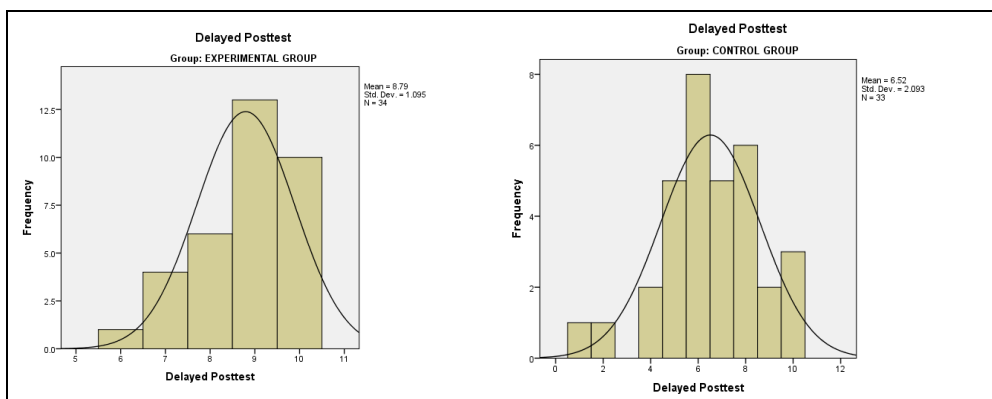

Figure 8. *The skewed distribution of the delayed post-test scores of the experimental and the control groups.*

A parametric test cannot be applied as it assumes the normality of the data. The mean constitutes an unreliable measure while the median score represents a more reliable alternative measure in this case. The non-parametric equivalent test for the independent t-test, Mann-Whitney U Test, was used to analyze the delayed post-test scores.

The test indicates that there is a highly statistically significant difference between the delayed post-test median scores of both groups, Mann-Whitney U = 189.00, two-tailed (Sig. = .000), (p < 0.05) (Table 8). As demonstrated in Tables 6 and 7, the experimental group's mean rank and median score are (44.94) (M = 9) respectively, which is statistically significantly higher than the control group's mean rank and median score (22.73) (M = 6).

Table 6. *NPar tests-mann-whitney U test for the delayed post-test scores*

Statistics			
Experimental Group	N	Valid	34
		Missing	0
	Median		9.00
	Percentiles	25	8.00
		50	9.00
		75	10.00
Control Group	N	Valid	33
		Missing	0
	Median		6.00
	Percentiles	25	5.00
		50	6.00
		75	8.00

Table 7. *Rankings for the mann-whitney U test of the delayed post-test scores of the experimental and the control groups*

Ranks	Group	N	Mean Rank	Sum of Ranks
Delayed Post-test	Experimental Group	34	44.94	1528.00
	Control Group	33	22.73	750.00
	Total	67		

Table 8. *The mann-whitney U value and significance level of the delayed post-test scores*

Test Statistics	Delayed Post-test
Mann-Whitney U	189.000
Wilcoxon W	750.000
Z	-4.735-
Asymp. Sig. (2-tailed)	.000

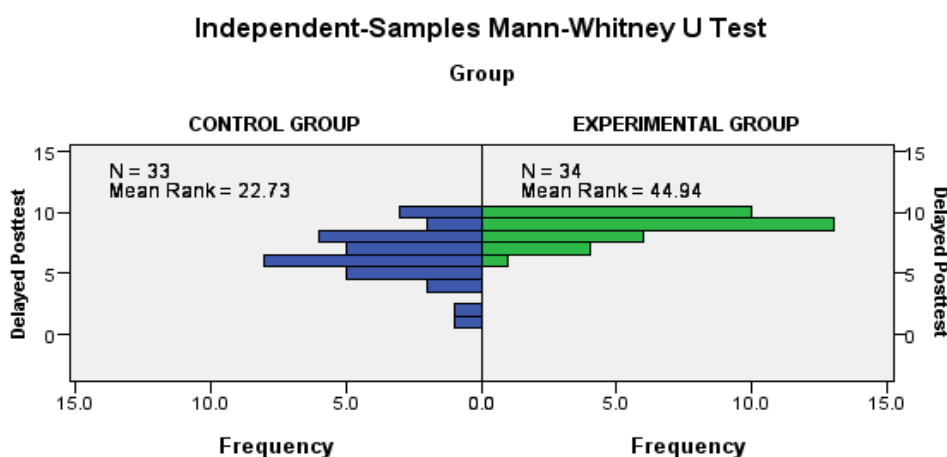a. Grouping Variable: Group

Figure 9. *The delayed post-test scores frequencies for the experimental and the control groups*

In summary, the experimental group's scores were significantly greater than the control group in two different times, in the immediate post-test and in the delayed post-test. Figure 9 displays a comparison between the distributions of the two post-tests data of both groups.
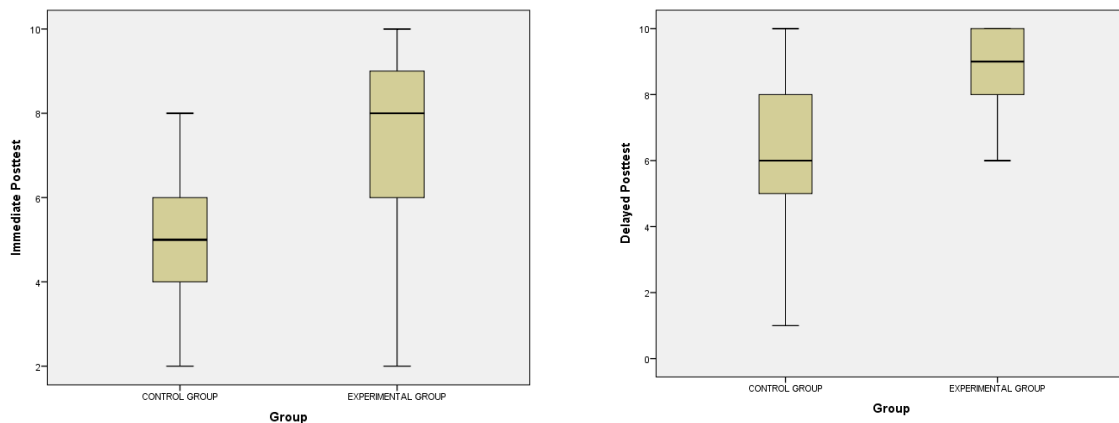
Figure 10. *The distributions of the scores of the immediate and the delayed post-test in the experimental and the control groups*

The immediate post-test versus the delayed post-test

The researchers aimed at comparing between the results of the immediate post-test and the delayed post-test within the same group. As mentioned in the previous section, the experimental group's delayed test scores do not constitute parametric data. Thus, a non-parametric test for a paired sample, Wilcoxon Signed-Ranks test, was applied to determine if there is any statistically significant difference in the performance of the experimental group in the two post-tests. The test indicated that a statistically significant difference was found between the results of immediate post-test and the delayed post-test. The immediate post-test ranks were statistically significantly higher than the ranks of the immediate post-test ($z = 4.137$, $p = 0.000$, $p < 0.05$). A similar procedure was conducted to the control group. However, the equivalent parametric test to Wilcoxon test, t-test for a paired sample, was applied to determine whether there is a statistically significant difference between the means of the immediate post-test and delayed post-test. A significant difference was evident where the mean score for the delayed post-test ($M = 6.52$, $SD = 2.093$) was statistically higher ($t = 3.620$, $df = 32$, two-tailed $p = 0.001$) from that of the immediate post-test ($M = 4.91$, $SD = 1.809$).

Additionally, the Bivariate Pearson Correlation (Table 9) was used to measure the strength and direction of any association that might exist between the post-tests data within each group. A statistically significant positive correlation was found between the experimental group's scores in both tests ($R = 0.385$, Sig. = 0.025). On the other hand, there was no significant correlation between the control group's scores in both post-tests ($R = 0.153$, Sig. = 0.395).

Table 9. *Correlations between the immediate post-test and the delayed post-test in both groups*

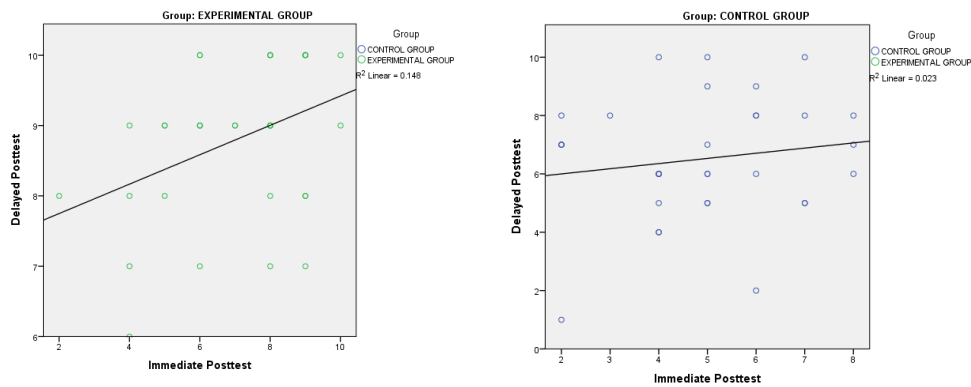| Paired Samples Correlations			N	Correlation	Sig.
Group					
Experimental Group	Pair 1	Immediate Post-test & Delayed Post-test	34	.385	.025
Control Group	Pair 1	Immediate Post-test & Delayed Post-test	33	.153	.395

Figure 11. *Correlations scatterplots between the immediate and the delayed post-test scores for both groups.*

The analysis of the results showed that the experimental group by far outperformed the control group in both post-tests. This result rejects the null hypothesis and supports the research hypothesis which states that the use of a CT-based approach can enhance EFL vocabulary retention. Also, the comparison between the results of the immediate post-test and the delayed post-test within each group showed that both groups performed better in the delayed post-test than in the immediate post-test; yet, the experimental group showed greater improvement. Moreover, when applying a correlation test to examine any changes within each group, the analysis showed a positive correlation between the results of the immediate post-test and the delayed post-test in the experimental group. This indicates a constant and increasing improvement in the students' ability of vocabulary retention. On the other hand, no significant correlation was found between the control group's post-test achievement. This conveyed that an almost arbitrary and non-constant achievement can be observed.

However, a possible explanation for the control group's better performance in the delayed post-test as opposed to the immediate post-test does not necessarily indicate an improvement in their retention ability as a result of instruction. Instead, such improvement (although arbitrary among students) can be the result of a 'testing effect'. To the researchers' surprise, a vocabulary quiz was scheduled after only one week from the end of the intervention assigned by their institution. The researchers tried to avoid or minimize its possible effect by postponing the second post-test for another week (it was originally planned to be administered after one week only from the end of the intervention). The 'Testing effect' is an improved retention of materials that results when they are retested at a later time (Delaney et al., 2010; Roediger & Karpicke, 2002; as cited in Shore, Ray, & Gooklasian, 2015). Thus, the fact that the students had studied and were tested for the institutions' official quiz might have interfered with the groups' achievements in the delayed post-test leading to a better performance. If the advantage of the 'testing effect' has occurred, it would have affected both groups equally. However, it still does not explain the far greater improvement in the experimental group which can only be the result of the CT-based approach implementation.

The literature lacks empirical studies that have adopted the same purpose or examined the variables of the current study in EFL contexts. However, this study's result aids and elaborates on the findings of other studies that observed a clear association between CT and vocabulary learning strategies (Nosratinia, Abbasi, & Zaker, 2015; Nosratinia & Zaker, 2015). It also confirms the existence of a positive relationship between CT and lexical knowledge reported in Faramarzi, Elekaei and Tabrizi's (2016) study. In addition, the

results of the current study of establishing a relationship between CT and vocabulary retention is consistent with the findings of other studies that have found a relationship between CT and the reading ability with unfamiliar words (Kamali & Fahim, 2011) and also between CT and the translation ability since lexical knowledge is an essential part of translation (Azin & Tabrizi, 2016).

Furthermore, the link found between the CT-based approach in general and the vocabulary retention in the current study supports the observed effectiveness of some of the CT strategies or techniques on EFL lexicon learning and retention; for instance, the positive effect of negotiation on vocabulary learning and retrieval (Nixon & Fishback 2009). The findings in this study also support other studies that have proved the effectiveness of some of CT strategies on EFL learners' vocabulary learning and retention such as collaborative dialogue.

CONCLUSION
The aim of this study was to investigate the impact of adopting a CT-based approach on learners' retention ability of EFL vocabulary. An experimental study was employed in a two-group pre and post-test design to measure such possible impact. The experimental group received vocabulary instruction with a CT approach whereas the control group was taught in the conventional methods of vocabulary teaching. The treatment consisted of seven lessons delivered over two weeks. The targeted vocabulary includes 28 words and fixed expressions that are drawn from the language curriculum course book. Two post-tests were administered after the intervention, one immediately after the treatment and another two weeks later. The findings showed that the experimental group outperformed the control group.

Several pedagogical implications can be drawn from this study. The first pedagogical implication is for EFL syllabus designers and material developers; CT should not be considered merely an additional thing that

teachers may or may not incorporate in their lessons but it should be treated as an integral part of EFL teaching materials. Many advantages have been attributed to the use of CT in EFL classrooms, for example, improving the learners' speaking proficiency (Alnofaie, 2013), and writing skills (Mall-Amiri & Sheikhy, 2014). Therefore, more serious efforts should be devoted to the infusion of CT in EFL syllabus and materials. Providing more intellectually challenging vocabulary tasks which provide plenty of opportunities for students to practice elaborate rehearsal is important for helping them achieve better retention of the learnt vocabulary. The second pedagogical implication concerns EFL instructors; it is highly recommended that they use CT strategies and problem-solving tasks to promote discussions, debates and students' reflections and to raise their level motivation and engage their cognitive abilities. Turner (1995) claims that complex and challenging tasks that promote CT and demand higher-order thinking skills can trigger students' motivation (as cited in Lai, 2011).

The current study was carried out in a Saudi context, where classroom dynamics and students' beliefs are shaped by the unique religious, societal, economic and educational variables of the region. Similar studies should be conducted in different contexts to gain a broader understanding of the role of CT in EFL learning and the extent of its impact. Furthermore, the participants in the current study were students in the intermediate level (B1+). Similar studies conducted with students in different levels of language proficiency might lead to more insights on how cognitive or intellectual involvement might be hindered (or not) by the learners' language proficiency. This could lead to a better understanding on how language and intellect interact. Consideration could also be given to the advisability of including a third follow-up post-test after the whole course to gain a broader understanding regarding vocabulary retention.

REFERENCES

Ab Kadir, M. (2007). Critical thinking: A family resemblance in conceptions. *Education and Human Development, 1*(2), 1-11.

Ab Kadir, M. (2009). *Rethinking thinking schools, learning nation: Teachers' and students' perspectives of critical thinking in Singaporean education.* Unpublished Dissertation. Melbourne.

Abudawood, E. (2015). *Saudi Arabian students' perceptions of the importance of critical thinking in their learning experiences in institutions of higher education in the Philadelphia region in comparison to their Saudi Arabian educational experience.* Unpublished Dissertation. Saint Joseph's University. Retrieved from http://search.proquest.com/docview/1781948761?accountid=142908.

Al-Gabrey, W. (2007). *The impact of utilizing brain storming method in developing the critical thinking & academic achievement of the secondary school first grade in mathematics curriculum.* Unpublished Dissertation. Umm Al.qura University, Makkah.

Alghamdi, E. (2014). *An investigation of the efficacy of multimedia glosses in incidental EFL vocabulary learning and retention through playing an online hidden-object* game. Unpublished Dissertation. The Graduate School of Missouri State University.

Al-Masrai, A., & Milton, J. (2012). The vocabulary knowledge of university students in Saudi Arabia. *TESOL Arabia Perspectives, 19*(3), 13-19.

Al Zahrani, B. & Elyas, T. (2017). The implementation of critical thinking in a Saudi EFL context: Challenges and opportunities. *Indonesian Journal of English Language Teaching and Applied Linguistics, 1*(2), 133-141.

Atkinson, P., & Silverman, D. (1997). Kundera's immortality: The interview society and the invention of the self. *Qualitative Inquiry, 3*(3), 304-325.

Atkinson, R. C., & Shiffrin, R. M. (1968). Chapter: Human memory: A proposed system and its control processes. In K. W. Spence, J.T. Spence (Eds.), *The psychology of learning and motivation* (pp. 89-195). New York: Academic Press.

Chang, A. C. S. (2015). A Japanese word association database of English. *Vocabulary Learning and Instruction, 4*(1), 47-57. doi: 10.7820/vli.v04.1.chang.

Clark, M. K., & Ishida, S. (2005). Vocabulary knowledge differences between placed and promoted EAP students. *Journal of English for Academic Purposes, 4*(3), 225-38.

Cook, V. (2008). *Second language learning and language teaching* (4th ed.). London: Holder Education.

Craik, F., & Lockhart, R. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior, 11*(6), 671-684.

Elyas, T. & Shah, S. R. (2018). Teaching/Developing Vocabulary through Peer Engagement and Interactive Strategies. *The TESOL Encyclopedia of English Language Teaching.* 1–7. Wiley-Blackwell Publishing (Area Editor: Editor-in-chief: John I. Liontas) doi.org/10.1002/9781118784235.eelt0742.

Facione, P. A. (1990). *Critical thinking: A statement of expert consensus for purposes of educational assessment and instruction.* Retrieved on March 1, 2019 from https://www.scirp.org/(S(351jmbntvnsjt1aadkposzje))/reference/ReferencesPapers.aspx.

Golonka, E., Brugman, C., Bowles, A., Silbert, N., Kramasz, D., Richardson, D., Corbett, R., & Doughty, C. (2012). *Foreign language vocabulary: Effective practices for learning and teaching.* University of Maryland, Maryland: CASL. Retrieved from https://criticalelt.wordpress.com/2017/04/16/foreign-language-vocabulary-effective-practices-for-learning-and-teaching/.

Hedge, T. (2000). *Teaching and learning in the language classroom: A guide to current ideas about the theory and practice of English language teaching.* Oxford: Oxford University Press, USA.

Joe, A. (1998). What effects do text-based tasks promoting generation have on incidental vocabulary acquisition? *Applied Linguistics, 19*(3), 357-377.

Judith, W. S., Susan, F. C., & Robert, G. (1985). *Thinking and learning skills: Relating instruction to research.* New York and London: Routledge Taylor and Francis Group.

Lai, E. R. (2011). Critical thinking: A literature review. *Pearson's Research Reports, 6*, 40-41.

Lee, H. W., Lim, K. Y., & Grabowski, B. L. (2008). Generative learning: Principles and implications for making meaning. In M. J. Spector, D. M. Merrill, J. van Merrienboer, & M. P. Driscoll (Eds.), *Handbook of research and educational communications and technology* (3rd ed.). New York, NY: Taylor & Francis Group.

Mall-Amiri, B., & Sheikhy, F. (2014). The comparative impact of autonomy and critical thinking on EFL learners' writing achievement. *Theory and Practice in Language Studies, 4*(5), 903–916. Retrieved from http://doi.org/10.4304/tpls.4.5.903-916.

Milton, J. (2006) Language lite? Learning French vocabulary in school. *Journal of French Language Studies, 16*(2), 187-205.

Nation, ISP. (1983). Testing and teaching vocabulary. *Guidelines, 5*(1), 12-25.

Nation, ISP. (2006). How large a vocabulary is needed for reading and listening? *The Canadian Modern Language Review, 63*(1) 59-82.

Nixon, S., & Fishback, J. (2009). Enhancing comprehension and retention of vocabulary concepts through small-group discussion: Probing for connections among key terms. *Journal of College Science Teaching, 38*(5), 18-21. Retrieved from http://search.proquest.com/docview/200387909?accountid=142908.

Rafi, M. (2011). Promoting critical pedagogy in language education. *International Research Journal of Arts and Humanities, 39*(39), 63–73.

Read, J. (2000). *Assessing vocabulary*. Cambridge: Cambridge University Press.

Rodgers, C. (2002). Defining reflection: Another look at John Dewey and reflective thinking. *Teachers College Record, 104*(4), 842-866.

Schmitt, N., Schmitt, D., & Clapham, C. (2001). Developing and exploring the behavior of two

Webb, W. B. (1962). The effects of prolonged learning on learning. *Journal of Verbal Learning and Verbal Behavior, 1*, 173-82.

new versions of the vocabulary levels test. *Language Testing, 18*(1), 55-88.

Scriven, M., & Paul, R. (2004). *Defining critical thinking*. Retrieved on March 29, 2019 from https://www.criticalthinking.org/pages/defining-critical-thinking/766.

Shore, R., Ray, J., & Gooklasian, P. (2015). Applying cognitive science principles to improve the retention of science vocabulary. *Learning Environments Research, 18*(2), 233–248. Retrieved from http://doi.org/10.1007/s10984-015-9178-1.

Webb, S. A., & Chang, A. C. S. (2012). Second language vocabulary growth. *RELC Journal, 43*(1), 113-126. Retrieved from https://doi.org/10.1177/0033688212439367.

Webb, S., & Chang, A. C. S. (2015). How does prior word knowledge affect vocabulary learning progress in an extensive reading program? *Studies in Second Language Acquisition, 37*(4), 651-675. doi:10.1017/S0272263114000606.