The Effects of Receptive and Productive Testing on EFL Active Vocabulary Knowledge: A Mixed Method Study

Dariush Aliakbari[1], Morteza Aslrasouli[2*], Davud Kuhi[3]

[1]Ph.D. Candidate, English Language Department, Maragheh Branch, Islamic Azad University, Maragheh, Iran, *dariush.aliakbari22@gmail.com*

[2*]Ph.D., English Language Department, Maragheh Branch, Islamic Azad University, Maragheh, Iran, *mortezarasuli@gmail.com*

[3]Ph.D., English Language Department, Maragheh Branch, Islamic Azad University, Maragheh, Iran, *davudkuhi@iau-maragheh.ac.ir*

Article Info	ABSTRACT
Article Type: **Research Article** **Received:** **31/05/2023** **Accepted:** **26/07/2023**	The development of new exam formats to aid the development of L2 vocabulary improvement has been the subject of extensive research. Nevertheless, it has rarely been discussed in prior literature whether different forms of assessments (receptive vs. productive) result in the thorough mastering of active vocabulary knowledge in in an EFL context like Iran, hence, this research examined the impact of receptive (multiple-choice tests) and productive (C-tests) vocabulary testing on EFL active vocabulary learning. To this end, a sample of 60 English learners from one of the private language institutes in Maragheh, Iran, were randomly assigned into the receptive vocabulary group, productive vocabulary group and traditional or control group. Each class had an equal number of students (N = 20). The data gathering tools were the Oxford placement test, a pretest in vocabulary to check the number of unfamiliar words, a posttest in vocabulary, and a semi-structured interview. The participants in the receptive vocabulary as the first experimental group took a series of multiple-choice tests, whereas the second group used C-tests during the treatment. The findings of Analysis of variance and subsequently, the post hoc results showed that while both C-tests and multiple-choice tests were effective in enhancing EFL active vocabulary knowledge, C-tests were more effective than multiple-choice tests. The findings of the interview with the foreign language students in the productive vocabulary class reinforced the findings of data analysis, as the students expressed positive attitudes toward the employment of C-tests as effective means for improving their active vocabulary recall. The results of this research provide new insights for language instructors and curriculum designers to apply C-tests in EFL courses. ***Keywords:*** Productive testing, receptive testing, active vocabulary knowledge, EFL learners

1. Introduction

The concept of lexicon has was described as being multidimensional or multicomponent (Milton & Fitzpatrick, 2014; Karafkan et al., 2022; Schmitt, 2014). It has long been recognized that vocabulary acquisition and knowledge can predict language use and ability (Milton, 2013; Nation, 2013).Wilkins (1972, p. 111) focused on the importance of vocabulary and stated "without grammar, very little can be conveyed; without vocabulary, nothing can be conveyed". According to Nation (2001), understanding a word entails understanding its grammatical functions, collocations, and usage restrictions like register and frequency, as well as its form, meaning, and use. The first refers to word parts, written, and spoken. The second refers to concepts and referents, form and meaning, and relationships, and the last refers to the use of words in context. All of the aforementioned elements of knowing a word means what are covered by receptive and productive knowledge, according to Nation (2001). For example, the receptive section of the spoken language is understanding the manner in which a term sounds, while the productive aspect is knowing how the word is pronounced. Hence, all three of a vocabulary's main parts include both productive and receptive vocabulary knowledge.

Scholars decided to categorize knowledge in vocabulary in accordance to how it is used in speaking, listening, writing, and reading. As a result, they distinguished between productive and receptive vocabulary (Durongbhandhu & Suwanasilp, 2023; Masrai, 2022; Schmitt, 2014). Receptive vocabulary knowledge refers to the understanding that students have of a word's meaning when reading text or listening to an audio file. Learners understand the meaning of vocabulary that is not used in speech or writing but that helps them understand the text they have read (Maskor & Baharudin, 2016). Learning words that are used by the students is considered productive vocabulary knowledge (Faraj, 2015). Because students may create new words to express their minds and emotions in a manner that the interlocutors can comprehend, productive vocabulary can be conceived of as a process of activating vocabulary (Webb, 2005). The active vocabulary that students can call orally is knowledge of vocabulary, according to Harmer (2001), whilst students can call and produce passive words through writing, they cannot create and generate a passive vocabulary via recognizing.

Receptive vocabulary development is the cornerstone of productive vocabulary growth, and word comprehension comes before word production (López Otero & Jimenez, 2022). The ability to identify and recognize writings may come before the capacity to communicate intention and do so in the appropriate vocabularies, or the capacity to identify and recognize written form may come before the capacity to spell words correctly. According to San Mateo-Valdehíta and Diego (2021), because students should first learn vocabulary through receptive learning before using it in productive ways, there is no need to divide receptive from the productive vocabulary.

Both the areas of instruction and evaluation deal with receptive vocabulary and productive vocabulary knowledge. As assessing complicated skills, both receptive and productive, are irrefutable components in monitoring the operability of instructional goals, assessment in language teaching is a crucial multidimensional endeavor (Polat, 2020). According to the conventional wisdom, vocabulary testing should focus on the separate abilities of reading, listening, speaking, and writing (together referred to as "receptive skills") (Fitzpatrick & Clenton, 2017). Tests of vocabulary knowledge are now employed for placement purposes due to the close relationship of vocabulary knowledge and total competency of language. Test items could also be divided into receptive and productive formats. For instance, matching and multiplechoice items are receptive (Evans & Morrison, 2011), while C-test and openended items are productive (productive article). The washback effects of these testing procedures could be different such that receptive formats stimulate passive teaching and learning, whereas productive formats encourage more active instruction that could lead to a more engaging learning process (Read & Chapelle, 2001). There are several studies in the field of receptive knowledge as well as productive vocabulary one (Durongbhandhu & Suwanasilp, 2023) that most of them are correlational in nature (e.g., Zhong, 2018); however, there are a few that studied these two strategies on active vocabulary performance of EFL learners. Therefore, this research's goal was to determine the effect of receptive (multiple-choice items) and productive vocabulary testing (C-test items) on active vocabulary recall and the EFL learners' points of view about these types of vocabulary tests.

2. Literature Review

For L2 English language learners, acquiring vocabulary is a major difficulty, especially in classroom settings when there is insufficient exposure due to time constraints (Evans & Morrison, 2011). There is no such thing as a fully mastered vocabulary; instead, vocabulary grows and develops over time (Li & Hafner, 2022). Making the form-meaning connection the most crucial sub-category of vocabulary knowledge is one strategy for differentiating the various vocabulary forms (Teng, 2019). Establishing a link between form and meaning, which is described as the attribution of the orthographical expression to the meaning of the lexicon, may also be challenging for students. According to Hulstijn (2013), learning a new word or expression without deliberately memorizing it can be challenging for EFL learners. Poor language processing throughout language input access (Hulstijn, 2001) may prevent students from deriving meaning and creating a form-meaning connection. Additionally, in the field of vocabulary knowledge, intensional and rational relationships with other semantically related words as well as linkages between ideas and referents, are crucial and help one understand a term in considerable detail (Vasiljevic, 2014).

The receptive-productive aspect, another subtype of knowledge in vocabulary, can be used to distinguish the level of understanding and lexis production proficiency (Garnier & Schmitt, 2016; Schmitt, 2014; Stoeckel et al., 2021). Henriksen (1999) suggests different aspects to examine vocabulary acquisition: a "partial-precise knowledge" aspect, a "depth of knowledge" aspect, and a "receptive-productive" aspect (p.340). He further notes that most scholars agree that vocabulary in the form of receptive and productive are separated. While knowledge in the form of production is the linguistic proficiency needed to create vocabularies in speaking and writing, receptive knowledge is sometimes characterized as the capacity to understand terms in reading and listening (Schmitt, 2010). As Melks Teichrow (1982) stated, a student acquires more vocabulary when receptive knowledge gradually transforms into productive knowledge (cited in Pignot-Shahov, 2012). Many scholars concur on a receptive-productive spectrum, generally suggesting a string of new words moving from recognition that is passive to recall that is active (Schmitt, 2014; Teng & Xu, 2022).

Different receptive and productive vocabulary tests have been developed and offered by academics to measure vocabulary knowledge. These exams allow for the evaluation of various talents and sub-skills. The Vocabulary Levels Test (VLT) and the Vocabulary Size Test (VST) (Nation & Beglar, 2007) are two of the most popular of these assessments that allow students to select a meaning from a list of alternatives after being given the target word. Additionally, multiple-choice exams are the most often utilized examinations to gauge the receptive vocabulary knowledge of EFL/ESL students (Stewart et al., 2021). Most multiple-choice exams typically have four or five options for each item, and students must choose the right one. Several benefits come with multiple-choice questions. Exam papers are easy for scorers to review at first. It could be measured mechanically, ensuring the accuracy of the scorers, because it typically only has one accurate answer or a small number of correct answers. In addition to having a high degree of scoring reliability, multiplechoice items can save time, effort, and cost when administering extensive exams. According to Alderson (2000), multiple-choice tests are a common method of determining a student's level of understanding. They give testers some degree of control over the possible answers and the learners' thought processes during the testing procedure.

However, the usefulness of multiple-choice tests has frequently been disputed. Such test items, for instance, are vulnerable to the guessing effect, which could cause test scores to be overestimated (Chiu & Camilli, 2013). Additionally, multiple-choice exams' validity is questioned because they could lead to a "test of passive recognition of words that does not approximate the experience of readers of authentic English texts" (Stewart, 2014, p. 271). This implies that this test cannot assess language as a communication tool. In everyday life, appropriate replies are produced rather than selected from various possibilities (Luo & Zhang, 2011). The researchers have claimed that multiple-choice assessments may overstate estimates of vocabulary quantity

and levels (e.g., Kremmel & Schmitt, 2016; Nation & Webb, 2011; Webb, 2007). To estimate students' understanding of the relationship of form and meaning, it has been recommended that meaning-recall vocabulary tests, in which meanings are conveyed orally by students rather than chosen are so influential.

On the other hand, effective exam formats might result in a positive washback because they do not promote decontextualized memorization of new terms. One of the best language examinations for determining vocabulary proficiency is the C-test. The cloze tests and the C-test method are theoretically related (Gogolin et al., 2021). In contrast to cloze test approaches, it uses different deletion strategies and other criteria. The test in the form of C-test is an integrated written test of overall English knowledge with decreased redundancy. It includes of 5 to 6 succinct, thorough, and correct texts. In writings such as these ones, the first and last statements are left alone (Karimi, 2011). Using the "rule-of-two" that entails removing the second half of each other term starting with the 2th word of the 2th sentence, the C-test damages certain portions of the text by eliminating half of the term (Daller, Müller, & Wang-Taylor, 2021). This test style requires knowledge of the context and lexical cues, such as word connections, and is productive and context-dependent (Wang-Taylor & Milton, 2019).

The majority of the literature concentrates on vocabulary size considering the effectiveness of on receptive and productive knowledge (Ozturk, 2015; Webb & Chang, 2012; Zhong, 2011; Zhong & Hirsh, 2009), and it consistently finds that L2 students have larger sizes of receptive vocabulary than productive vocabulary sizes, as well as various growth rates for the 2 categories of vocabularies. Since that knowledge in vocabulary is a nuanced construct, size alone does not entirely demonstrate a word's movement and improvment from receptive use to productive one (Nation, 2001; Schmitt, 2010). The connection between two knowledge and English skills and subskills is being researched in the other area of receptive and productive vocabulary investigation. While EFL knowledge in vocabulary is necessary and a requirement for comprehending other English skills and sub-skills, its documented relationship to reading ability and writing is probably maybe the most popular explanation for the evaluation of EFL vocabulary knowledge (Hartono & Prima, 2021; Uchihara et al., 2020; Zhong, 2018). It is still not apparent which type of vocabulary knowledge—receptive or productive—is a better forecaster of other skills, including reading skill (McLean et al., 2020). Scientists assert that because reading is fundamentally a receptive one, a vocabulary exam that emphasizes recognition of vocabulary in the form of receptive knowledge is likely to be the most accurate indicator of the ability in reading. They claim that the large percentage of vocabulary tests employed to estimate reading ability are in fact receptive tests of vocabulary size or levels (Ha, 2021).

Two types of vocabulary knowledge were the focus of some experimental and correlational studies, for example, Maskor and Baharudin (2016) studied

the effective roles of these two strategies on writing skill and concluded that integration of two-vocabulary knowledge can be influential in increasing learners' interest to write essays. Moreover, Torabian et al. (2014) examined the connection of Iranian students' collocational skill and vocabulary knowledge. The findings supported the existence of a large gap between the undergraduate learners' receptive and productive lexical knowledge.

Although a bulk of research exists on receptive and productive vocabulary knowledge, there is a paucity of research investigating the effect of receptive and productive testing on active vocabulary recall and the EFL learners' points of view about these types of vocabulary tests. To this end, this study aimed to niche the gap and answer the following research question:

RQ1: Do various test types (multiple choice and C-test) affect Iranian EFL students' active vocabulary knowledge?

RQ2: What are EFL students' perceptions toward using tests to improve active vocabulary knowledge?

3. Method

3.1. Participants

The participants of this research were intermediate-level EFL learners of a private language center in Maragheh, Iran that were chose in terms of the convenience sampling. A total of 60 learners (36 female and 24 male) were randomly assigned into the receptive vocabulary group, productive vocabulary group, and traditional or control group. Each class had an equal number of students (N = 20). The participants were homogeneous in age (between 20 and 27 years old) and language proficiency at data gathering time. In addition, they share the same mother tongue, Turkish, but they are also fluent in Persian.

3.2. Materials and Instruments

3.2.1 Oxford Placement Test

A standardized test called the Oxford Placement Test (OPT) was selected to gauge language proficiency. The average administration duration for this 60-item test is 70 minutes. Those scores falling between X and one SD were chosen since the acquired scores did not significantly deviate from the normal distribution. The students in the current research were the intermediate-level ones, in accordance to the institute's report, however to ensure more, OPT test was used to study whether the learners were the same in their general English level. The test's reliability was estimated and reported to be.81, showing the test items' high level of internal consistency.

3.2.2. Target Words

Before the study, a practice vocabulary test with 80 common English terms was created. The researcher tried to find the topics that the subjects had little familiarity. This examination was created by an updated Test of Academic Lexicon (Scarcella & Zimmerman, 1998), which asks test takers to categorize their word knowledge into one of four levels as expressed below:

(a) I comprehend the new lexicon

(b) I do not understand the term

(c) I have seen it previously, but I am unsure of its meaning

(d) I comprehend the term whenever I see or hear it in a statement; however, I am confused of using it in speaking or writing.

The items provided in the exam were obtained from the vocabulary list of the 3rd volume of the book "4000 Essential English Vocabulary" by Nation (2009). The test showed acceptable reliability for testing depth of vocabulary knowledge and Cronbach's alpha reported to be .95. After the results, 60 lexicons with an average score below 3 were retained from the study since it was doubtful that participants had receptive vocabulary knowledge or productive vocabulary knowledge. Six sets of 10 things each were created from these 60 pieces. Each session, the participants were given a different assortment of these things to investigate.

3.2.3 Multiple-Choice Test

Based on the words unfamiliar to the participants, a series of 6 multiplechoice tests consisting of 20 items were designed to see EFL learners' active vocabulary development. The test was designed to see if the multiple-choice test affects the first experimental group and how receptive testing could affect active vocabulary knowledge over time.

3.2.4. C-test

Based on the words unfamiliar to the participants, a series of 6 C-tests consisting of 20 items was designed for each productive vocabulary learning group session. To make the C-test, the rule-of-two was applied to construct items using passages based on the target words instructed to the groups. The exam was used to estimate the development of EFL students' active vocabulary knowledge in the second experimental group. The C-test has completion items formed by mutilating half of every other word starting from the second word of the second sentence. The first and last sentences were kept intact. The tests consisted of passages with five gaps that required active vocabulary knowledge of the students.

3.2.5. Post-test in Vocabulary

For measuring the vocabulary level of the learners after the treatment, a posttest of vocabulary with 40 items in the form of both recognition (20 items) and production (20 items) tests were designed. In the recognition test, the participants were needed to recognize the suitable words through multiplechoice items. According to Bachman and Palmer (2010), "recognition form items only required the examinees to recognize the correct response from among the alternatives provided for each stem" (p.32). However, in this study, the production test was described as the learners' ability to produce the correct vocabulary in the correct context by answering the items in the form of fill-inthe-blanks. According to Bachman and Palmer (2010), a production test refers to a suppletion test and "suppletion or completion from items required to supply the missing part(s) of the stem or complete an incomplete item" (P.32). The test piloted with 15 EFL learners from the same population of the institute and its reliability reported being .73. Furthermore, the validity of the test validated by three university instructors of Maragheh Azad University.

3.2.6. Semi-structured Interview

A semi-structured interview was the last tool utilized in the current research to measure the participants' perspectives toward the employment of the vocabulary testing method that outperformed the other group's vocabulary knowledge. To this end, an interview was conducted with the participants of productive vocabulary testing group. The reason for conducting interview sessions with this group was the fact that the productive vocabulary group improved the other groups and the researcher conducted one by one interview with the 20 participants. As it is clear, the interview is a daunting and timeconsuming task and for this reason, productive vocabulary group was of prime importance for running interview due to the high performance. For running interview, the researcher invited the students of the group to reflect on their overall perception of the testing strategy used in the classes, such as whether they liked the class and recommended it to others. They were also required to give reasons for their answers. Besides, the participants were needed to say their ideas about the testing strategy's strengths and weaknesses and suggest ways to improve it. Each student was interviewed individually while recording their voices for further analysis and presentation. The interviewees provided their responses in Farsi, which were translated into English and presented as part of the study's qualitative data. To estimate the credibility of the interview questions, the information gathered from the interviewees were presented and they were required to focus on each question, and the answers were given themselves to see whether there was any problem with or differences in the answers. The participants confirmed the accuracy of the data, so the interview's credibility was approved. For dependability, 30% of the interview results were re-checked by two of the researcher's colleagues who were familiar with the data analysis section. The inter-rater reliability result was reported to be .87.

3.3. Procedure

At the onset of the study and following a general ethics code for research, the Academic Affairs of the university's language department approved the consensus forms to start the research. Thus, the learners were informed on the study's goals and the confidentiality of personal information. It was crucial to ensure that both the research institute and the participants understood their right to anonymity. In the first phase and two weeks before the onset of the treatment, OPT was used for 75 students to be sure the EFL students' homogeneity in language proficiency. The test was used to examine whether the students were about the same regarding their language proficiency. After analyzing the scores of the proficiency test, 15 learners were discarded from the study due to their high/low scores on the test, and 60 learners were divided among three groups: receptive vocabulary (N = 20), productive vocabulary (N = 20), and control group (N = 20). It is worthwhile stating that the criteria for homogenizing the students on general English was 1SD below and above the mean. After this phase, the vocabulary pretest was run to the learners in the groups one week before the study, and the allotted time was 60 minutes. Then, the treatment was started. The treatment lasted for ten sessions, and the unfamiliar words were instructed to the learners by the researcher.

In both experimental groups, the instructors presented a topic based on the selected new words such as 'disappear', 'abandon', 'wisdom', and did warmup activities to introduce the new topic. Teaching new words using synonyms and antonyms, making sentences, giving feedback, and encouraging the learners to take part in the process of instruction were the other responsibilities of the teachers. The primary variation of two experimental groups was that the receptive testing strategy group got multiple-choice tests after teaching the unknown new words in each session. In contrast, the productive vocabulary testing strategy group received C-tests as a part of their vocabulary testing strategy. A vocabulary test related to the study material was presented to the learners at the final minutes of each session. These tests were designed in the multiple-choice tests and C-tests. The situation was different for the learners in the control group as they needed to memorize the list of the same new words, make sentences, and get feedback from the teacher or peers.

Moreover, there was no test after the instruction during each session. After finishing the treatment, the researcher administered a posttest in vocabulary in the form of recognition (multiple-choice format) and a productive one (fill in the blanks format), and the allotted time was 60 minutes. To give more credit to the findings of this study, a semi-structured interview was also administered to know how learners evaluate the effectiveness of the most effective method in enhancing their active vocabulary knowledge compared with the other groups. The researcher conducted one-by-one interviews with each student while recording their voices for further analysis and presentation. The

interviewees provided their responses in Farsi, which were translated into English and presented as part of the study's qualitative data.

3.4. Data Analysis

The data were statistically analyzed by SPPS software (Version 22.0) to compare the scores obtained from the experimental and control groups to assess the progress in EFL learners' active vocabulary knowledge. To this end, independent samples ANOVA was run to probe the differences of the groups and check how receptive and productive testing types could affect the active vocabulary knowledge of the students. The key constructs of interview were analyzed via thematic analysis and reported as extracts. The analysis of the qualitative data in the form of the interviews were conducted based on the guidance presented by Schmidt (2004). The interview transcripts were analyzed using his semi-structured interview analysis method. The students' complete responses to each interview question served as the unit of analysis for the interview data. When coding the interview data, only the codes related to the question of the interview were presented in the form of sub-themes and the extracts.

4. Results and Discussion

4.1. Results

4.1.1. Quantitative Data Analysis: Descriptive and Reliability Statistics

The research question number one dealt with the effectiveness of receptive vs. productive vocabulary testing strategy on EFL learners' vocabulary knowledge. Prior to comparing the increases across the groups, the data's normality was verified using Shapiro-Wilk, which produced a normal distribution. Thus, parametric tests were employed to compare the effectiveness of the groups individually as well as between them. The results of the posttest in vocabulary among groups are displayed in Table 1.

Table 1.

Descriptive Statistics of Post-test in Vocabulary

Groups		Post-test	
	N	M	SD
Receptive	20	13.88	2.49
Productive	20	17.66	2.53
Control	20	11.55	1.78

To find out if there are meaningful differences among the treatment groups, their posttest scores were compared using the ANOVA procedure. Table 1 reveals that the mean score of the learners in the receptive vocabulary group is 13.88 with an SD of 2.49 and the mean score of learners receiving productive testing strategy is 17.66 with an SD of 2.53. In addition, the mean of the conventional vocabulary testing strategy group as a control group with an SD of 1.78 is 11.55. This table shows that the means of groups are roughly different in the posttest. However, the differences among groups needed to be tested statistically; thus, the assumption of a parametric test needed to be tested. Prior to the use of the ANOVA, the assumption of normality and equality of variances were checked. The result of the Levene's test showed that the assumption of Homogeneity of Variances was not violated (Levene Statistic: 0.62; Sig. = .187 > .05); consequently, it can be concluded that the data is normally distributed in the posttest of vocabulary. The result of the ANOVA is presented in Table 2.

Table 2.

Results of ANOVA on Posttest

Sum of Squares	df	Mean Square	F	Sig.
Between Groups 347.72	2	1713.31	338.49	.000
Within Groups 247.22				
	54	41.455		
Total 621.94				
	57			

Since the P-value (.000) is lower than.05, it can be concluded that the groups are not the same in vocabulary learning after treatment, $F_{(2, 54)}$ =338.4, p = .000. Having found a statistically meanigful difference among the groups in the posttest, the researcher needed to compute a post hoc test (Tukey or scheme post hoc test) to see where the differences existed. Table 3 indicates the findings of the post hoc test.

Table 3.

Results of Post Hoc Test in Posttest of Vocabular

(I) Group1	(J) Group1	Mean Difference (I-J)	Std. Error	Sig.
productive	Receptive	146.00000[*]	2.27342	.000
	Control	183.05556[*]	2.27342	.000
Receptive	Productive	-145.00000[*]	2.27342	.02
	Control	49.05556[*]	2.27342	.07
Control	Productive	-183.05556[*]	2.27342	.000
	Receptive	-49.05556[*]	2.27342	.07

As Table 3 shows, there was a meaningful difference between productive and receptive tests and between productive testing strategy and the control group. Based on the first raw, it can be concluded that the productive vocabulary testing strategy was better than the receptive vocabulary testing strategy group and control group in vocabulary learning. In the second raw, there is a statistically meaningful difference between receptive and productive vocabulary testing strategy but not between receptive vocabulary testing strategy and the control group (Sig. = .07). Finally, there was a significant difference between the control group and productive vocabulary testing strategy in which the effect of productive vocabulary testing strategy in vocabulary knowledge was significant. However, there was no meaningful difference between the control group and receptive vocabulary testing strategy.

4.1.2. Qualitative Data Analysis

According to the interview results, most students in the productive vocabulary testing strategy group preferred this type of testing strategy to the traditional classroom (18 out of 20 students). They suggested this strategy be used in other English-related courses. The students expressed their points of view in the sentences; however, the principal codes, including the merits and demerits of the strategy under study represented in the following two tables (Table 4 and Table 5).

Table 4.

Positive Attitudes of Learners towards Productive Vocabulary Testing Strategy

Extracts	Sub-themes
I can better retrieve the active vocabulary knowledge I learned from C-tests compared to the multiple-choice tests I already experienced. High retrieval helped me to adapt to my To adopt stress stress.	High retrieve
I enjoyed learning vocabulary by using C-tests; in fact, they were like playing games that caused fun.	Enjoyable Fun as games
Productive vocabulary instruction that I did not experience so far was amazing and motivating since t hey were like photos that could be remembered whenever needed.	Attractive Motivating Acted like photos
The teaching procedures were new and different from what Different context we had learned before since the class time was devoted to detailed instruction of new words in the form of missing words. detailed instruction of	new words in the form of missing words
Active participation and happiness in class increased my motivation and decreased my anxiety level.	Active participation Happiness increased motivation decreased anxiety
Most of the new words were practiced and practiced for several times. Also, feedback from teachers and peers when there were mistakes should not be ignored since they were helpful.	Practice feedback

The C-test encourages students to engage in more active and in-depth vocabulary learning since they are acted as awarenessraising tasks.	In-depth instruction awareness raising tasks

Notwithstanding the positive points, the course had some demerits. However, they were fewer in numbers. Table 5 summarizes the negative attitudes that some of the interviewees in the treatment group mentioned to.

Table 5.

Negative Attitudes of Learners towards Productive Vocabulary Testing Strategy

Extracts	**Sub-themes**
There are many blanks in C-test that make everyone bored, and this can frustrate the students. This fact made me stressed.	Boring and frustrating
	Lots of blanks
	Increased anxiety
Before this course, I knew something about C-tests and the missing words, but the negative point with this type of instruction was that sometimes, much time was wasted discussing a new word	Time-wasting (on the new word)

Productive vocabulary instruction and C-tests were full of fun Time-wasting (on for me. Also, the strategies that we learned were new and methods) informative; however, the method and the procedures were timewasting for me.

As it is evident from the above table, the interviewees reported on the course's pros and cons. They reported on the time-wasting of the course and the boring and frustrating nature of the testing strategy in the form of C-test with lots of blanks in the new words as important downsides. However, the course motivated them as it was attractive and motivating. The productive vocabulary classroom had novelty to the learners as they resulted in high retrieval of the new words due to the high practice and low anxiety.

4.2. Discussion

The current research attempted to explore the impacts of receptive and productive types of testing on Iranian EFL learners' active vocabulary knowledge performance. The multiple-choice and C-tests were employed using a true experimental design. Additionally, the students' perceptions toward the use of such testing types were explored.

As the results for the first research question showed, both the productive and receptive groups' mean scores increased following the intervention. The experimental groups' C-test results (the productive and receptive groups) were significantly higher than those of the control group. In addition, the Iranian EFL students' active vocabulary knowledge was improved more by the productive style of testing than by the receptive type. The improvment of the two treatment groups on total scores confirms the results of past research that cloze tests had a more positive washback effect on active vocabulary production than the multiple-choice formats (Amini & Ibrahim-González, 2012).

Hencefore, it can be stated that C-tests were more effective than multiplechoice tests in shifting students' passive vocabulary to active vocabulary knowledge because the students attempted to deduce the meaning using the context that was presented. Babaii and Jalali Moghadam (2006) found that Ctest taking entails macro-level processing, requiring test takers to look for contextual clues, such as lexical chains. Therefore, the positive effects of the C-test may be related to this type of processing that triggers a higher level of processing and, hence an active recall of vocabulary knowledge and strengthen overall comprehension (Chae & Shin, 2015). In other words, productive testing could lead to a better active recall of vocabulary knowledge, and this was confirmed by Enayat and Derakhshan (2021), who showed that productive measures of vocabulary knowledge were more associated with L2 speaking ability than receptive measures.

The results further support previous research by Chai et al. (2020) which studied the impact of applying C-tests in developing students' active vocabulary knowledge in EFL context. The findings of this research revealed that C-tests were influential in enhancing students' active vocabulary knowledge. In addition, Chai et al. (2020) suggested that cloze-based passages help measure and expand EFL learners' vocabulary development. Therefore, these findings might imply that using cloze tests to provide some passages can be useful in enhancing EFL learners' active vocabulary knowledge. Their study also concluded that cloze-based tasks like cloze-passages with multiplechoice items helped students improve their vocabulary knowledge.

This finding is consistent with Kılıçkaya (2019) study, indicating that multiple-choice items resulted in the learners' better employment of receptive knowledge rather than productive knowledge (Read, 2012). The participants may score better on the multiple-choice exam than on the c-test because,

according to Coxhead (2018), the participants' receptive vocabulary can be significantly greater than the productive one. The multiple-choice format does promote speculation. It is clear that choosing the key on a test paper by guessing is not always preferable to coming up with active vocabulary. According to Read (2000), multiple-choice tests prescribe examinees merely to have one choice from a range of three or four alternatives and inhibit examinees from expressing creativity and demonstrating original and imaginative thinking. Moreover, the findings indicated that EFL learners' mean scores in the vocabulary posttest in the control group were the lowest in comparison with the other two experimental groups. One justification for the low score in the control group can be rote learning in the form of memorization of a string of new words. Rote learning enables the learners to repeat new words from memory rather than to learn them to understand (Tayebi & Marefat, 2019). Therefore, repetition minus understanding resulted in shortterm rather than long-term learning, and this particular memory strategy may not suit the needs of all learners. Therefore, it is the teachers' responsibility to try to enhance, assist, and guide the learners to use different strategies in acquiring vocabulary to get better improvement and accurate results in the big field of vocabulary learning (Li & Cutting, 2011).

The results of this research further showed that EFL learners expressed positive attitudes towards using C-tests to develop their active vocabulary knowledge. The participants agreed that using C-tests was more effective than the receptive type of multiple-choice vocabulary tests in enhancing the students' motivation and interest to learn and recall vocabulary for productive use. The participants agreed that this type of activity was intended to assess their syntactical knowledge by asking them to show how well they understood the meaning in light of the context (Stopar, 2014). The participants tried to guess whether the word to be placed into the blank can be an adjective, a verb, or a noun without fully understanding what it meant. This is in line with a study that investigated the effects of adventure video games as a productive instrument for learning words on receptive and productive vocabulary recall. The authors found that using words to play the games could not only affect the active recall of the learners but also influence the participants' motivation, such that they expressed positive attitudes toward using productive tools for learning and developing receptive and productive vocabulary recall (Janebi Enayat & Haghighatpasand, 2019).

5. Conclusion and Implications

The findings of the current empirical research provided evidence for the positive effects of productive testing using C-tests on active vocabulary knowledge. The findings of the semi-structured interview also showed that students were more motivated by taking productive measures of vocabulary than taking the receptive type of testing like the multiple-choice tests. They preferred C-tests to the cloze tests due to the mentioned merits such as active

participation and happiness of the class context. It was concluded that C-tests could significantly affect the active vocabulary knowledge of intermediate Iranian EFL learners compared to multiple-choice tests.

This research has some significant educational and instructional implications. The findings of this study encourage teachers to employ various vocabulary testing techniques, such as C-test and multiple-choice tests, to assist learners improve their active vocabulary knowledge. The findings of this study can also assist teachers in determining which of these two strategies is the most effective. In addition, the test developers are recommended to use more productive vocabulary tests as they can have a positive washback effect. Materials writers could also use productive tasks more in textbooks for EFL students as they would be more motivated.

Despite the implications of this study, there were some limitations that future researchers should address. First, the present study delimited the investigation to just the intermediate students so that further studies could probe the impacts of receptive and productive testing on active vocabulary knowledge for other proficiency levels. Second, this study used a convenient sample from a language institute in Iran. Further studies could employ more representative samples to enhance the generalizability of the findings. Finally, the present research used C-test as a productive measure of vocabulary (Laufer & Nation, 1999), while further studies could use tests like the Productive Vocabulary Levels Test (PVLT).

References

Alderson, C. (2000). *Assessing reading*. Cambridge University Press.

Amini, M., & Ibrahim-González, N. (2012). The Washback Effect of Cloze and Multiple-Choice Tests on Vocabulary Acquisition. *Language in India*, *12*(7), 1-23.

Bachman, L. F., & Palmer, A. S. (2010). *Language testing in practice*. Oxford University Press.

Chae, E. Y., & Shin, J. A. (2015). A study of a timed cloze test for evaluating L2 proficiency. *English Teaching*, *70*(3), 117-135.

Chai, Z. F., Swanto, S., & Din, W. A. (2020). Variants of Cloze-Test Based Tasks and Vocabulary Achievement. *Universal Journal of Educational Research*, *8*(7), 2980-2989.

Chiu, T. W., & Camilli, G. (2013). Comment on 3PL IRT adjustment for guessing. *Applied Psychological Measurement*, *37*(1), 76–86.

Coxhead, A. (2018). Vocabulary assessment. In J. I. Liontas (Ed.), *The TESOL encyclopedia of English language teaching* (pp. 1-6). John Wiley & Sons, Inc.

Daller, M., Müller, A., & Wang-Taylor, Y. (2021). The C-test as predictor of the academic success of international students. *International Journal of Bilingual Education and Bilingualism*, *24*(10), 1502-1511.

Durongbhandhu, N., & Suwanasilp, D. (2023). Computer-assisted English lessons for logistics with task-based language teaching for enhancing receptive and productive vocabulary knowledge of EFL Thai learners. *Language Education and Acquisition Research Network*, *16*(1), 47-74.

Enayat, M. J., & Derakhshan, A. (2021). Vocabulary size and depth as predictors of sec-ond language speaking ability. *System*, *99*, 102521. https://doi.org/10.1016/ J.SYSTEM.2021.102521.

Evans, S., & Morrison, B. (2011). Meeting the challenges of English-medium higher education: The first-year experience in Hong Kong. *English for Specific Purposes*, *30*(3), 198-208.

Garnier, M., & Schmitt, N. (2016). Picking up polysemous phrasal verbs: How many do learners know and what facilitates this knowledge? *System, 59*, 29-44.

Gogolin, I., Schnoor, B., & Usanova, I. (2021). Crossing the bridge to literacy in foreign languages: C-test as a measure of language development. *Multilingua*, *40*(6), 771-790.

Faraj, A. K. A. (2015). Effective strategies for turning receptive vocabulary into productive vocabulary in EFL context. *Journal of Education and Practice*, *6*(27), 10-19.

Fitzpatrick, T., & Clenton, J. (2017). Making sense of learner performance on tests of productive vocabulary knowledge. *TESOL Quarterly*, *51*(4), 844-867.

Ha, H. T. (2021). Exploring the relationships between various dimensions of receptive vocabulary knowledge and L2 listening and reading comprehension. *Language Testing in Asia*, *11*(1), 1-20.

Harmer, J. (2001). *The practice of English language teaching.* New York.

Hartono, D. A., & Prima, S. A. B. (2021). The correlation between Indonesian university students' receptive vocabulary knowledge and their reading comprehension level. *Indonesian Journal of Applied Linguistics*, *11*(1), 21-29.

Henriksen, B. (1999). Three dimensions of vocabulary development. *Studies in Second Language Acquisition*, *21*(2), 303-317.

Hulstijn, J. H. (2001) Intentional and incidental second language vocabulary learning: A reappraisal of elaboration, rehearsal and automaticity. In Robinson, P. (ed.), *Cognition and second language instruction.* Cambridge University Press.

Hulstijn, J. H. (2013). Incidental learning in second language acquisition. In: C. A. Chapelle (Ed.). *The encyclopedia of applied linguistics*. WileyBlackwell. Online DOI: 10.1002/9781405198431.wbeal0530.

Janebi Enayat, M., & Haghighatpasand, M. (2019). Exploiting adventure video games for second language vocabulary recall: A mixed-methods study. *Innovation in Language Learning and Teaching*, *13*(1), 61-75.

Karafkan, M. A., Ansarin, A. A., & Hadidi, Y. (2022). Depth and breadth of vocabulary knowledge as predictors of narrative, descriptive and argumentative writing. *Journal of Modern Research in English Language Studies*, *9*(2), 27-50.

Karimi, N. (2011). C-test and vocabulary knowledge. *Language Testing in Asia*, *1*(4), 1-32.

Kılıçkaya, F. (2019). Assessing L2 vocabulary through multiple-choice, matching, gap-fill, and word formation items. *Lublin Studies in Modern Languages and Literature*, *43*(3), 155-166.

Kremmel, B., & Schmitt, N. (2016). Interpreting vocabulary test scores: What do various item formats tell us about learners' ability to employ words?. *Language Assessment Quarterly*, *13*(4), 377-392.

Laufer, B., & Nation, P. (1999). A vocabulary-size test of controlled productive ability. *Language Testing*, *16*(1), 33-51.

Li, Y., & Hafner, C. A. (2022). Mobile-assisted vocabulary learning: Investigating receptive and productive vocabulary knowledge of Chinese EFL learners. *ReCALL*, *34*(1), 66-80.

Li, X., & Cutting, J. (2011). Rote learning in Chinese culture: Reflecting active confucian-based memory strategies. In *Researching Chinese learners* (pp. 21-42). Palgrave Macmillan, London.

López Otero, J. C., & Jimenez, A. (2022). Productive vocabulary knowledge predicts acquisition of Spanish DOM in Brazilian Portuguese-speaking learners. *Languages*, *7*(4), 273-289.

Luo, S., & Zhang, X. (2011). Multiple-choice item and its backwash effect on language teaching in China. *Theory and Practice in Language Studies*, *1*(4), 423-425.

Masrai, A. (2022). The relationship between two measures of L2 phonological vocabulary knowledge and L2 listening comprehension. *TESOL Journal*, *13*(1), 2-12.

Maskor, Z. M., & Baharudin, H. (2016). Receptive vocabulary knowledge or productive vocabulary knowledge in writing skill, which one important. *International Journal of Academic Research in Business and Social Sciences*, *6*(11), 261-271.

McLean, S., Stewart, J., & Batty, A. O. (2020). Predicting L2 reading proficiency with modalities of vocabulary knowledge: A bootstrapping approach. *Language Testing*, *37*(3), 389-411.

Milton, J. (2013). L2 vocabulary acquisition, knowledge and use. *EUROSLA Monographs Series 2*, 57-78.

Milton, J., & Fitzpatrick, T. (2014). *Dimensions of vocabulary knowledge*. Palgrave Macmillan.

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

Nation, P. (2009). Reading faster. *International Journal of English Studies*, *9*(2), 131-144.

Nation, I. S. P. (2013). *Learning vocabulary in another language*. Cambridge University Press.

Nation, I. S. P., & Beglar, D. (2007). A vocabulary size test. *The Language Teacher, 31*(7), 9–13.

Ozturk, M. (2015). Vocabulary growth of the advanced EFL learner. *The Language Learning Journal*, *43*(1), 94-109.

Pignot-Shahov, V. (2012). Measuring L2 receptive and productive vocabulary knowledge. *Language Studies Working Papers*, *4*(1), 37-45.

Polat, M. (2020). Analysis of multiple-choice versus open-ended questions in language tests according to different cognitive domain levels. *NovitasROYAL (Research on Youth and Language)*, *14*(2), 76-96.

Rahmani, N., & Maleki, A. (2023). Comparative effects of Semantic vs Structural Elaboration on Iranian EFL Learners' Vocabulary Learning Process. *Journal of Modern Research in English Language Studies 10*(2), 45-67.

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

Read, J. (2012). Assessing vocabulary. In C. Coombe, P. Davidson, B. O'Sullivan, & S. Stoynoff (Eds.), *The Cambridge guide to second language assessment* (pp. 257-264). Cambridge University Press.

Read, J., & Chapelle, C. A. (2001). A framework for second language vocabulary assessment. *Language Testing*, *18*(1), 1-32.

San Mateo-Valdehíta, A., & de Diego, C. C. (2021). Receptive and productive vocabulary acquisition: effectiveness of three types of tasks. Results from French students of Spanish as second language. *Onomázein*, *5*(1), 37-56.

Scarcella, R., & Zimmerman, C. (1998). Academic words and gender: ESL student performance on a test of academic lexicon. *Studies in Second Language Acquisition*, *20*, 27–49.

Schmitt, N. (2010). Key issues in teaching and learning vocabulary. In *Insights into non-native vocabulary teaching and learning* (pp. 28-40). Multilingual Matters.

Schmitt, N. (2014). Size and depth of vocabulary knowledge: What the research shows. *Language Learning*, *64*(4), 913-951.

Schmitt, N., Schmitt, D., & Clapham, C. (2001). Developing and exploring the behaviour of two new versions of the Vocabulary Levels Test. *Language Testing*, *18*(1), 55-88.

Stewart, J. (2014). Do multiple-choice options inflate estimates of vocabulary size on the VST?. *Language Assessment Quarterly*, *11*(3), 271-282.

Stewart, J., Stoeckel, T., McLean, S., Nation, P., & Pinchbeck, G. G. (2021). What the research shows about written receptive vocabulary testing: A reply to Webb. *Studies in Second Language Acquisition*, *43*(2), 462471.

Stoeckel, T., McLean, S., & Nation, P. (2021). Limitations of size and levels tests of written receptive vocabulary knowledge. *Studies in Second Language Acquisition*, *43*(1), 181-203.

Stopar, A. (2014). Testing and assessing English word-formation. *Romanian Journal of English Studies, 11*(1), 1-8.

Tayebi, G., & Marefat, S. (2019). The impact of rote learning on vocabulary learning: The case of Iranian EFL learners with visual and auditory learning styles. *Journal of Studies in Learning and Teaching English*, *8*(1), 129-145.

Teng, F. (2019). The effects of context and word exposure frequency on incidental vocabulary acquisition and retention through reading. *The Language Learning Journal*, *47*(2), 145-158.

Teng, M., & Xu, J. (2022). Pushing vocabulary knowledge from receptive to productive mastery: Effects of task type and repetition frequency. *Language Teaching Research*, *0*(0). https://doi.org/10.1177/13621688221077028.

Torabian, A. H., Maros, M., & Subakir, M. Y. M. (2014). Lexical collocational knowledge of Iranian undergraduate learners: implications for receptive & productive performance. *Procedia-Social and Behavioral Sciences*, *158*, 343-350.

Uchihara, T., Saito, K., & Clenton, J. (2020). Re-examining the relationship between productive vocabulary and second language oral ability. In *Vocabulary and the Four Skills* (pp. 146-165). Routledge.

Vasiljevic, Z. (2014). Teaching collocations in a second language: Why, what and how?. *Elta Journal*, *2*(2), 48-73.

Wang-Taylor, Y., & Milton, J. (2019). Predicting international students' academic success with vocabulary and intercultural communicative competence. *Journal of Applied Linguistics*, *32*, 83-101.

Webb, S. (2005). Receptive and productive vocabulary learning: The effects of reading and writing on word knowledge. *Studies in Second Language Acquisition*, *27*(1), 33-52.

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. (2021). A different perspective on the limitations of size and levels tests of written receptive vocabulary knowledge. *Studies in Second Language Acquisition*, *43*(2), 454-461.

Webb, S., & Chang, A. C. (2012). Vocabulary learning through assisted and unassisted repeated reading. *Canadian Modern Language Review*, *68*(3), 267-290.

Wilkins, D. A. (1972). *Linguistics in language teaching*. Edward Arnold.

Zhong, H. (2011). Learning a word: From receptive to productive vocabulary use. In *The Asian conference on language learning: Official conference proceedings*, 116-**126**.

Zhong, H. F. (2018). The relationship between receptive and productive vocabulary knowledge: A perspective from vocabulary use in sentence writing. *The Language Learning Journal*, *46*(4), 357-370.

Zhong, H., & Hirsh, D. (2009). Vocabulary growth in an English as a foreign language context. *University of Sydney Papers in TESOL, 4*(4), 85113.