

Impact of Interaction and Output Modality on the Vocabulary Learning and Retention of Iranian EFL Learners

Bahman Motaei, Ph.D. Candidate, Department of English, Tabriz Branch, Islamic Azad University, Tabriz, Iran
bahman5257@gmail.com

Saeideh Ahanghari*, Assistant Professor, Department of English, Tabriz Branch, Islamic Azad University Tabriz, Iran
s_ahangari@yahoo.com

Nasrin Hadidi Tamjid, Assistant Professor, Department of English, Tabriz Branch, Islamic Azad University Tabriz, Iran
nhadidi@iaut.ac.ir

Abstract

This study investigated the impact of interaction and output modality on vocabulary learning and retention of EFL learners. To investigate the impact of Interaction, solitary (n =69) and collaborative (n =62) groups served as experimental and No Interaction No Output (n =26) as control group. To address the effect of modality, spoken (n =39) and written (n =31) modalities served as experimental and No output modality (n =26) as control group. The study was done in 8 weeks. The groups read passages with target words highlighted. Solitary and collaborative groups reconstructed the passages individually or in dyads. The spoken and written modality groups reconstructed them in the respective modality. Then, pretest, immediate, and delayed posttests were administered. The ANOVA results showed that the collaborative group outperformed the other groups and spoken modality outperformed the written modality group focusing on interaction and modality separately. The 2×2×2 ANOVA results showed significant main effects for time, interaction, and output modality. An ‘interaction’ effect was found between time and interaction, time and modality, and modality and Interaction. The ‘interaction’ between time, Interaction, and modality was insignificant. The findings have implications for language teachers, syllabus designers, and language testing experts.

Keywords: collaborative output, interaction, output modality, solitary output, vocabulary learning and retention

Introduction

The oft-cited Input Hypothesis (Krashen, 1985) is considered a turning point in second language acquisition (SLA) research. It claims that the prerequisite condition for L2 learning is comprehensible input (i+1). Krashen ignores the role language output plays on the grounds that “only comprehensible input is effective in increasing proficiency” (p.56) Long (1981) attached more importance to output created as a result of meaning negotiation through requests for clarification or confirmation checks. In later versions of his hypothesis, Long (1996) claims that the meaning negotiation can make the learners to modify their own output. Swain (1985, 1995) proposed that comprehensible input may not be sufficient in all aspects of L2 acquisition. According to Swain (1985), learners need opportunities to develop their production by manipulating their interlanguage system in order to clarify message that their interlocutor does not understand. She claims that output serves some functions in L2 acquisition: a metalinguistic function, a hypothesis-testing function, and a noticing function.

Comprehensible input could be guaranteed through different kinds of modalities. They include writing, reading, auditory channel, picture, video, and a combination of all of these in the form of multimodal input. Modality of presentation is claimed to be beneficial for the general learning mechanism. This was justified on theoretical grounds. The first green light was given by Sadoski and Paivio (2001) in dual coding theory. According to this theory, human cognition processes the objects in two coding systems: verbal and nonverbal modality. Presenting information in both modes can improve recall. It also implies that exposing learners to contexts containing language items in different modes can increase vocabulary acquisition because it enriches the volume of verbal and nonverbal connections (Sadoski, 2005). Another justification for the differential role of different modalities was provided by Penny (1989). The separate-streams hypothesis (Penney, 1989) suggests that visually and aurally presented verbal materials are processed and retained independently. Penney (1989, p. 399) states that “the processing of auditory and visually presented verbal items is carried out separately in short-term memory.”

The shortcomings of comprehensible input theory were claimed to be resolved by Long’s Interaction hypothesis. Getting impetus from this theory, the SLA researchers felt a need for classroom activities that promote communicative interaction in second language classrooms (Doughty & Williams, 1998; Ellis, 2003, 2005; Lightbown, 1998; Nassaji & Fotos, 2007; Pica, 2007). The aspiration of these researchers were claimed to be achieved through pedagogical tasks that require learners to work together and produce output collaboratively (Kowal & Swain, 1997; Swain, 1995; 1998; Swain & Lapkin, 1995, 1998, 2000, 2001).

The conclusion to be drawn from the studies on the impact of interaction on vocabulary learning is that interaction and collaboration through pair work and small group is beneficial theoretically and pedagogically. Most of the studies compared different types of tasks completed individually or collaboratively. The research on the effect of pair work (collaborative) vs. individual (solitary) on vocabulary and retention of Iranian EFL learners, to the best of our knowledge, is scant.

Due to the fact that most of the studies investigated the impact of input modality, the present study addresses the question of the possible effect of output modality on the initial learning and later retention of the words. The studies done on the issue were mostly conducted in ESL contexts leading us to infer that the research in EFL milieu is scant. The conclusion to be drawn from the abovementioned studies is that input modality has different effects on language acquisition. An ample body of research can be found in SLA. Further research is needed to shed light on output modality in second language acquisition. The purpose of the current study is to investigate whether the way output is produced affects the immediate learning and delayed retention of the vocabulary.

Some studies investigated the role of collaborative tasks in Iran (e.g. Dehqan & Mohamadi, 2017; Soleimani & Mahmoudabadi, 2014; Minaei & Rezaie, 2014; Tajeddin & Jabbarpoor, 2014). To our best knowledge, no study was found to work on the impact of output modality and interaction in the same study. This study is claimed to bridge input, interaction, output, and more importantly modality. What’s more, most of the studies are done in the psychological milieu. The researchers felt a need to investigate the issue in EFL context of Iran.

To the best of our knowledge, two studies addressed the issue of output modality. Tian et al. (2015) investigated this variable in psycholinguistics. They examined the ability to follow instructions within working memory under varying input and output modalities. The findings showed that input and output modality affected the remembering of instruction in working memory. Niu & Park (2014) investigated the effect of collaborative output, output modality, and word engagement in vocabulary learning and retention in Chinese context. The results indicated

that oral output and written output outperformed the reading- only group in terms of receptive and productive vocabulary in all the posttests. The oral and written output groups difference was not significant.

Based on the above mentioned studies, this study seeks to answer the following research questions:

Q1.Does interaction significantly affect Iranian EFL learners' vocabulary learning?

Q2.Does output modality significantly affect Iranian EFL learners' vocabulary learning?

Q3.Do Interaction and output modality significantly affect Iranian EFL learners' retention of the vocabularies over time?

Literature Review

Modality and Output

Various empirical studies have highlighted the facilitative role of oral output, either collaborative or solitary, in vocabulary learning. Joe (1998) found that text-based solitary oral retelling promoted incidental vocabulary acquisition better than reading did because of the need for word generation in text retelling. Ellis and He (1999) and de la Fuente (2002) demonstrated that negotiated oral output played a more central role in lexical learning than modified aural input did. The researchers suggested that negotiated oral output engenders pushed output, word noticing, and discourse that scaffolds learning.

The superiority of written output over reading comprehension in promoting lexical learning has been confirmed by studies involving the writing of sentences (e.g. Laufer, 2003; Webb, 2005) and of compositions (Hulstijn & Laufer, 2001; Hulstijn & Trompetter, 1998; Kim, 2008; Laufer, 2003). The same has been true of some sentence or text gap-filling tasks; however, gap-filling writing has not been consistently superior to reading comprehension, especially when both receptive and productive lexical acquisition and retention have been examined (e.g. Hulstijn & Laufer, 2001; Laufer, 2003).

Few studies have ever investigated whether oral output and written output influence lexical learning differently. A notable exception is de la Fuente (2003), in which oral output and written output was executed within face-to-face interactions and computer mediated interactions, respectively. The study concluded that the two types of interactions were equally effective in promoting written receptive and productive lexical acquisition and retention because the negotiation process generated in interactions involved noticing the item and processing it in depth. Yet, the study discovered that face-to-face interactions facilitated oral productive lexical learning better than computer-mediated interactions did because the former was in oral mode and the latter in written mode.

Some studies investigated the effect of one modality auditory or visual versus a combination of two modalities. Kelly (1992) investigated whether reading a text containing the words and reading and listening to the text lead to the same results. In a pilot study, she had found a slightly stronger effect for the reading-only group on an immediate visual test, but a higher retention performance for the dual-modality group on delayed aural and visual tests. The researcher found a better and significant immediate effect for the reading group on an immediate visual test but no better performance on a delayed visual test. On the delayed aural test, the dual modality condition participants outperformed the reading- only condition. The conclusion was that "the ear does assist the eye in the long-term retention of words". The memory for foreign words was deemed to be dependent on both modality of presentation and the type of memory

measures. The findings of Baltova (1999) show that exposing the learners to words through a combination of sound, visual information, and text improves the learners' vocabulary learning.

While some research suggests that simultaneous presentation of information through different modalities may be helpful (see also Leahy, Chandler & Sweller, 2003; Mayer & Anderson, 1992), other studies suggest that adding to the modality of presentation may not assist learning (Solman, Singh & Kehoe, 1992; Wu & Solman, 1993). Some studies with L1 children have demonstrated that presenting new words with pictures does not help them in learning new words (Solman et al., 1992; Wu & Solman, 1993). It has been suggested that associating a word with pictures would divert the learner's attention from the printed word, hence leading to poor processing of the word for learning (Wu & Solman, 1993).

Pichette (2002) investigated whether exposing the learners to pictures and their pronunciation enhanced the learners' recall of concrete words in Spanish. Testing the learning of the words in four modality conditions (word alone, word plus sound, word plus picture, and word plus sound and picture), Pichette found that adding to the modality did not improve the memory and sometimes negative results were obtained. He found a higher performance for the word-only condition than the three other conditions, and a lower recall performance for the sound/ picture/ word condition than for the word/picture condition.

Interaction and Output

Inspired by 'Output Hypothesis' and 'Interaction Hypothesis', several studies investigated the role of collaborative output tasks. In the recent years, an unprecedented attention was given to pair work and small group activities on the ground that they would help learners in learning and assessment contexts. They are claimed to help the learners use their receptive and productive language skills in both pedagogy and testing.

Several studies have empirically examined the role of collaborative output tasks in L2 learning (e.g. García Mayo, 2002a, 2002b; García Mayo, 2017; Kim, 2008; Kowal & Swain, 1994; Kuiken & Vedder, 2002; Nabei, 1996; Nassaji & Tian 2014; Storch, 1998, 1999, 2005, 2007). García Mayo (2002a) compared the dictogloss with a text reconstruction task. García Mayo (2017) strived to answer the question whether interaction in pairs or small groups contributed to the frequency and outcome of Language Related Episodes (LREs). The results indicated that there was no significant difference between groups and pairs. Kowal and Swain (1994) investigated dictogloss as a particular type of collaborative output task. Kim (2008) compared the effect of pair and individual task on the acquisition of vocabulary items included in a dictogloss task. Those working in pairs performed significantly better on both an immediate and a delayed vocabulary posttest. Nassaji & Tian (2014) investigated language coproduction (collaborative task). Their results showed that output group outperformed the group input- only group. Their results also indicated that collaborative group's production were more correct compared to individuals. Storch (1999) analyzed the impact of interaction on grammatical accuracy across three different tasks: a cloze exercise, a text reconstruction task, and a composition task. The students who worked in pairs and had an opportunity to discuss their grammatical choices took longer to complete the tasks, but produced more accurate written texts than those working alone. Storch (2005) compared dyadic and individual performance on a short composition task based on a graphic prompt. The analysis of the oral interactions between the dyads confirmed that pair work provided the learners with an "opportunity to collaborate on the writing process, pool their ideas and knowledge, and provide each other with immediate feedback on language" (p. 154). Storch and Wigglesworth (2007) compared the performance of 24 pairs and 24 individual learners on two writing tasks, a report and an argumentative essay.

Wigglesworth and Storch (2009) compared 24 pairs and 48 individual learners writing an argumentative essay. In both studies pairs were assigned more time to complete the task than individual learners. The two studies obtained similar results: No differences were found in terms of fluency and complexity, but the texts written in pairs were significantly more accurate than those written individually.

In a series of studies, Storch (2001a; 2001b; 2002a; 2002b) investigated the nature of pair interaction in an adult ESL classroom. Based on her pair talk data, she identified four patterns of interaction amongst pairs. In the collaborative pattern, both learners work together throughout the task completion process and assist each other. Dominant/dominant pairs, on the other hand, show an unwillingness or incapability to engage with each other's contribution. Dominant/passive pairs involve a dominant participant who takes control of the task with an authoritarian stance, and a passive peer who maintains a subservient role. Finally, in expert/novice pairs, the more knowledgeable learner (expert) actively encourages the less knowledgeable learner (novice) to engage in the task.

One conclusion that may be drawn from the above studies is that collaborative pair work may facilitate learners' interaction and attention to the target forms, but it may not necessarily lead to superior learning in comparison to individual work.

These studies show that collaborative tasks may create more language learning opportunities, and in particular more vocabulary learning opportunities, than individual tasks. However, Nassaji and Tian (2010) failed to obtain clear evidence of the benefits of collaborative tasks over individual ones for L2 vocabulary learning. In this study, 26 English L2 learners completed two cloze tasks and two text-editing tasks focused on the use of English phrasal verbs. One of the two versions of each task was performed in pairs and the other one individually. All learners increased their knowledge of English phrasal verbs, as measured by a vocabulary pretest and posttest. In fact, learners working collaboratively improved more than those working individually, but the differences observed were not statistically significant.

Method

Design of the Study

This study is a quasi-experimental pretest-posttest control group design. The independent variables are output modality and Interaction. To measure output modality, NOM, SOM, and WOM groups were compared. To investigate the impact of Interaction, the NINO, SOPT, and COPT groups were compared.

The dependent variable is vocabulary learning with two levels: immediate vocabulary learning and delayed retention. Lexical learning is operationally defined here as the participants' scores on the immediate posttests. Lexical retention is defined as the participants' scores on the delayed posttests. The probable gains in the immediate posttest were considered lexical learning. The gains in delayed posttests is an indication of lexical retention.

Participants

The participants in this study were 157 university students with different majors studying in Kermanshah Islamic Azad University. They were divided into six groups. The groups included students taking General English. The group samples were taken from six intact general English classes. The first group (n =26) was called No Interaction No Output (NINO). The second group was also named No Output Modality (NOM) to focus on modality to the neglect of Interaction. The third group (n =38) was called Spoken Output Modality (SOM). The fourth group (n =31) was classified as Written Output Modality (WOM). The SOM and WOM and NOM were used to

answer the second research question. The fifth group (n =30) was classified as Solitary Output (SOPT). The sixth group (n =32) was called Collaborative Output (COPT). The NINO, SOPT, and COPT groups were used to answer the first research question. The age range of participants was 18 - 24 because they were studying general English at BA level. Their proficiency level was intermediate. Both genders were present in the groups. They were majoring in humanities. They were selected due to the higher proficiency level of engineering and science students.

Instrumentation

The following instruments were used in the current study.

Reading input passage

The first instrument used in this study was the students' textbook. The groups' main textbook was ACTIVE Skills for Reading: Book 2 (Anderson, 2007). Three passages were selected from the book as input passage. The peer groups were asked to circle unknown words for the purpose of selecting target words and distractors. Almost 40 words were selected as cued words for reconstruction (13 target words each passage). All the words were glossed with Persian meanings for the participants. This book was selected because it was the students' textbook. The passages were about different topics. There were words in the textbook qualifying as target words for the present study.

Treatment tasks

The three tasks were carried out in three stages. Each stage was allotted the same time length. To facilitate learning and retention, the cued words were provided on a sheet of paper. The main purpose of the present paper was investigating learning and retention. That's why, they were provided with the words on a sheet of paper. In the first stage, all the groups were supposed to read the first input passage in nearly 10 minutes. In the second stage, the experimental groups were supposed to reconstruct the passages. For the control groups i.e. no interaction no output (NINO) and no output modality (NOM), the students answered the reading comprehension questions. These two groups were not supposed to construct the passages. The solitary output (SOPT) group was supposed to read the passages and then reconstruct the passages individually without looking at the passage. The collaborative output (COPT) was required to reconstruct the passage in collaboration with one of the peers. The SOM and WOM groups reconstructed the passages via spoken and written modalities, respectively. The participants in SOM and WOM worked on the task without getting help from peers to remove the effect of interaction. In the third stage, all the groups consulted the passages again and checked the correctness or falsehood of their answers. The same process was followed for all passages.

Homogeneity test, pretest and posttests

In order to check the homogeneity of the groups, the reading and writing sections of PET test was administered. The rationale behind such a decision was that the reconstruction was done in written and spoken modalities. What's more, the main building blocks of reading and writing is vocabulary.

A vocabulary pretest was administered to understand the participants' baseline knowledge of target words. The Vocabulary Knowledge Scale (VKS) developed by Wesche and Paribakht (1996) was adapted to the vocabulary level of the participants.

Based on the participants answer to the pretest, those answering Item V correctly can be removed from the study because they use the word productively. Item I is the ideal one. It means

they completely lack that specific word' knowledge. Items I and II test the recognition and items III and IV test the production level of vocabulary knowledge. The VKS based pretest results could assure us of the homogeneity of the groups in terms of vocabulary knowledge. In addition to PET, VKS could tell the participants' level of baseline vocabulary, hence a second instrument for homogeneity.

The number of items in the pretest or self-report VKS was 40, one item for each target word).e.g. irrigation

I. I don't remember having seen 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 _____ provide Farsi meaning;

IV. I know this word. It means _____ (provide Farsi meaning or English synonym)

V. I know this word. I can use it in a sentence. -----

The posttests were a mix of recognition and production items. The rationale behind this is that an ample knowledge of vocabulary requires both recognition and production. Some individuals may recognize the words with knowing how to produce them in context. This combination of recognition and production was followed in both posttests. The recognition type included multiple choice and matching items. The production format was composed of items made based on the format of Productive Vocabulary Levels Test. This test is a diagnostic test developed by Laufer and Nation (1995). It required students to produce the words rather than to simply recognize them. The number of items for immediate and delayed posttests was 40. The allotted time for answering the posttests was one hour.

Cronbach's Alpha for internal consistency or reliability was used in order to determine if the parts of the pretest and posttests were consistent internally. The reliability index of VKS and immediate and delayed posttests were significant. The vocabulary knowledge scale had a high reliability index ($r = .91$). The posttests indexes for immediate and delayed posttests were .73 and .71, respectively.

Data Collection Procedure

157 university students majoring in Engineering and Humanities were selected. The classes were intact but the reading and writing part of PET (Preliminary English Test) was used to be ensured of homogeneity of groups regarding their vocabulary, reading, and writing knowledge.

The data was collected in an eight-week period. In the first week, all six groups (NINO, NOM, SOM, WOM, COPT and SOPT) completed the vocabulary pretest within 10 minutes. The allotted time was enough because the scales adapted from VKS were the same for all the words so the participants only focused on the words.

In week 2, all groups performed their tasks. In this week, the first passage was taught to all groups by one of the researchers. The NINO and NOM groups did the reading and were asked to answer multiple choice items based on the passage. The SOM group was asked to reconstruct the passage orally. The WOM group was needed to reconstruct the passage in written modality. The COPT group was asked to reconstruct the passage in pairs either in written or oral modality. The SOPT group reconstructed the passage in written or spoken mode. They were not allowed to get help from their peers or the teacher. To check if the groups focus on the target words, they were asked to reconstruct the passage using the highlighted target words. The second passage was provided in week3. In week 4, the third passage was given. The treatment procedures for all the passages were repeated. The reconstruction task was completed in three consecutive weeks (Week 2 Passage NO1, Week3 Passage NO2, Week4 Passage NO3).

In the next week (week5), Immediate posttest was administered to all the groups. This posttest measured the recognition and production of all groups regarding all 40 target words.

In Week 6, all groups were administered the immediate posttest2 was administered to all the groups. This posttest measured the production of all groups regarding all 40 target words.

The delayed posttests were administered 1 month later. In week7, the delayed recognition Test was administered. In week8 (the last session of the study) the delayed production test was administered. These post-tests measured the retention of the target words selected from the three passages. The distractors were selected from the same passage or passages covered in the same textbook.

Scoring procedure

According to Paribakht and Wesche (1996), the VKS scoring accepts self-reported answers of categories I and II for scores of 1 and 2, respectively. Wrong responses in categories III, IV or V are scored 2. A score of 3 indicates that an appropriate meaning was provided for categories III or IV. A score of 4 is given if the word is used in a sentence demonstrating the learners' knowledge of its meaning in that context but with a wrong grammatical category. A score of 5 is awarded if the target word is used semantically and grammatically correctly in a sentence context.

In this study, the scoring procedure for the pretest is changed due to the research methodology. The score for the pretest or self-report is 0-4. This study considers the first level as the most optimal condition for inclusion into the study. The purpose is to select the target words completely unknown to the participants. So, the answer to the first scale is awarded 0. Level 2 is still optimal for the study because the participants don't meet the requirement for the knowingness. The answer to this level, was awarded the score of 1. The level 3 is scored 2 if the provided meaning by the participants is correct. If the answer is incorrect, the score of level 2 namely 1 is awarded. The answer to the level 4 means that they know the word at the production level. The score for this item is 3 if a correct or partially correct answer is given. If the answer is not correct, the score for item 3 is awarded, namely 2. The least optimal condition for inclusion into the study is level 5 which means the participants know the word at the production level. The participants who answer this item correctly are awarded the score 4. This means that they should be removed from the sample or the target word should be removed from the study.

Results

Interaction and vocabulary learning

The first research question sheds light on the impact of collaborative output task on delayed learning of vocabulary. To this end, the performance of three groups is compared. The descriptive statistics for NINO as control group and SOPT and COPT as experimental groups are shown in Table 1.

Table 1. *Descriptive statistics of interaction groups in immediate vocabulary posttest*

	NINO (n = 26)		SOPT (n = 69)		COPT (n = 62)	
	M	SD	M	SD	M	SD
Vocabulary Learning	2.88	1.47	4.22	1.72	4.31	2.01

To understand if the mean differences is significant, a one-way ANOVA was conducted. The results are shown in Table 2 below.

Table 2. ANOVA for vocabulary post-test for NINO, SOPT, and COPT

	Sum of Squares	Df	Mean Square	F	Sig.	η^2
Between Groups	41.69	2	20.84	6.36	.002	0.07
Within Groups	504.07	155	3.27			
Total	545.77	157				

The ANOVA results indicated that the mean differences are significant as far as the vocabulary is concerned ($F(2, 155) = 6.36, p = .002$). The estimated significance value is smaller than the preset significance level sig ($0.002 < 0.05$). We can be sure that the mean differences are significant. This, however, should not be considered as an indication of equal differences among the means. The effect size was not large ($\eta^2 = .07$).

The Scheffe post-hoc test is presented in Table 3 below.

Table 3. Scheffe post hoc test for vocabulary learning for interaction

(I) Interaction	(J) Interaction	Mean Difference (I-J)	Std. Error	Sig.
NINO	SOPT	-1.34396*	.41552	.00
	COPT	-1.42686*	.42374	.00
Solitary	NINO	1.34396*	.41552	.00
	COPT	-.08290	.31689	.96
Collaborative	NINO	1.42686*	.42374	.00
	SOPT	.08290	.31689	.96

The post- hoc results show that the difference between Solitary Output (SOPT) and Collaborative Output (COPT) groups is not significant at sig $0.96 > 0.05$. As expected by the researchers, the experimental groups outperformed the control group. The comparison of means shows that the SOPT and COPT had surprisingly similar results. Contrary to our expectations, the interaction (individual vs. pair work) did not lead to better results.

Output modality and vocabulary learning

The second question sheds light on the impact of output modality on vocabulary retention of the groups. The results of the all the groups in the post tests were compared. Table 4 reports on mean and SD of the performance of each group on this posttest.

Table 4. *Descriptive statistics of modality groups in the immediate vocabulary posttest*

Group	N	Mean	SD
No Output Modality	26	17.80	0.63
Spoken Output Modality	38	21.10	1.18
Written Output Modality	31	25.73	2.24

The experimental groups outperformed the control group. The written output modality group had the best results.

To show the significance of results in vocabulary posttests, the ANOVA test was applied. The results of ANOVA are displayed in Table 5.

Table 5. *ANOVA for ANOVA for vocabulary post-test for NOM, SOM, and WOM*

Sources of change	Sum of Squares	df	Mean Square	F	Sig.
	895.726	2	447.863	196.680	.000
Between Groups	209.495	92	2.277		
Within Groups	1105.221	94			
Total					

The ANOVA table shows that the results are significant at $0.000 < 0.05$ significance level, showing that the results of the groups in vocabulary posttest are significant ($F_{2, 92} = 196.680, p < .00$).

The post-hoc test for the immediate vocabulary post-test is shown in Table 6 below.

Table 6. *Scheffe post hoc test for vocabulary learning for output modality*

(I) output modality	(J) output modality	Mean Difference (I-J)	Std. Error	Sig.
SOM	WOM	-.02394	.31890	.54
	NOM	1.37214*	.41255	.00
WOM	SOM	.02394	.31890	.54
	NOM	1.39609*	.42824	.00
NOM	SOM	-1.37214*	.41255	.005
	WOM	-1.39609*	.42824	.00

The post hoc analysis shows that in the immediate vocabulary test there is not a significant mean difference between spoken output modality and Written Output modality groups ($\text{sig} 0.54 > 0.05$).

Output modality, Interaction and vocabulary retention

Do Interaction and output modality significantly affect Iranian EFL learners' vocabulary retention over time as tested by Vocabulary Performance posttests?

This research question sheds light on the performance gains of groups from immediate vocabulary post-test (IVP) to delayed vocabulary post-test (DVP). Time is considered a within group variable.

The descriptive statistics for interaction and output modality is presented in Table 7

Table 7. Descriptive statistics for interaction and output modality in IVP and DVP

	Interaction	Output modality	Mean	Std. Deviation	N
Immediate Vocabulary Post-test	Solitary	SOM	32.92	2.822	38
		WOM	24.58	4.395	31
		Total	29.17	5.506	69
	Collaborative	SOM	39.13	4.523	36
		WOM	28.38	4.11	26
		Total	34.62	6.878	62
	Total	SOM	35.94	4.862	74
		WOM	26.31	4.645	57
		Total	31.75	6.748	131
Delayed Vocabulary post-test	Solitary	SOM	10.84	3.522	38
		WOM	12.29	3.278	31
		Total	11.49	3.466	69
	Collaborative	SOM	11.69	3.437	36
		WOM	11.57	4.234	26
		Total	11.64	3.759	62
	Total	SOM	11.25	3.484	74
		WOM	11.96	3.727	57
		Total	11.56	3.595	131

The homogeneity of variances is calculated by Leven's test the results of which is presented in Table 8.

Table 8. Levene's Test of Equality of Error Variances for IVP and DVP

	F	df1	df2	Sig.
IVP	4.20	3	127	.37
DVP	.88	3	127	.45

The Levene's test results confirmed the homogeneity of variances for both IVP and DVP. The significance levels are .37 and .45, respectively. Since the values reach the significance level ($p > .05$), we can be sure that the groups are homogeneous.

A $2 \times 2 \times 2$ mixed design repeated measure ANOVA was conducted to investigate the impact of Interaction (2 levels: solitary and collaborative) and output modality (2 levels: spoken output modality and written output modality) and the timing of testing (2 time points: immediate posttest and delayed posttest). The ANOVA results are presented in Table 9.

Table 9. ANOVA results related to IVP and DVP scores

Source	Sum of Squares	Df	Mean Square	F	Sig.	Partial Eta
--------	----------------	----	-------------	---	------	-------------

						Squared
Within Group Results						
Time	24762.31	1	24762.31	1842.26	.00	.93
Time * Interaction	391.26	1	391.26	29.11	.00	.18
Time * output modal	1671.28	1	1671.28	124.34	.00	.49
Time * Interactio output modality	2.88	1	2.88	.21	.64	.002
Between-Group Results						
Interaction	413.58	1	413.58	26.92	.00	.17
output modality	1264.12	1	1264.12	82.28	.00	.39
Interaction * output Modality	63.44	1	63.44	4.13	.04	.03

As shown in Table 7, significant main effects were found for time, $F(1, 125) = 1842.26$, $Sig. p < .05$, Interaction, $F(1, 127) = 26.92$, $Sig. p < .05$, and output modality, $F(1, 127) = 82.28$, $Sig. p < .05$. An ‘interaction’ effect was found between time and Interaction, $F(3, 142) = 29.11$, $Sig. p < .05$, time and output modality, $F(3, 131) = 124.34$, $Sig. p < .05$ and output modality and Interaction was also significant, $F(3, 125) = 4.13$, $Sig. p < .05$. However, the three-way ‘interaction’ between time, Interaction, and output modality was not significant.

For the first independent variable (time), there was a significant main effect. In fact, in Time 1 (IVP), the performance of groups was better. The first between-groups factor (Interaction) had a significant main effect. The collaborative group outperformed the solitary group. The third factor (output modality) had also a significant main effect. As for this between-groups factor, the spoken output modality group had a better performance than Written output modality. As the results show interaction * time had a significant effect. In fact, the solitary and collaborative groups had a better performance in time 1 than time 2. Time * output modality main effect was also significant. Both SOM and WOM groups performed better in delayed vocabulary post- test (DVP). The effect of Interaction*output modality was statistically significant. In fact, collaborative group outperformed the solitary group. This effect is noticeable in spoken modality than the written modality. The three-way time*interaction*output modality effect was not statistically significant.

The full results are displayed in Table 10.

Table 10. *Main and 'Interaction' effects for IVP and DVP*

			Mean	Std. Error
Interaction	Output modality	Time		
Solitary	SOM	1	32.92	.64
		2	10.84	.58
	WOM	1	24.58	.71
		2	12.29	.64
Collaborative	SOM	1	39.13	.66
		2	11.69	.60
	WOM	1	28.38	.78
		2	11.57	.70
Output modality		Time		
	SOM	1	36.03	.46
		2	11.26	.41
	WOM	1	26.48	.53
		2	11.93	.47
Interaction		Time		
Solitary		1	28.75	.48
		2	11.56	.43
Collaborative		1	33.76	.51
		2	11.63	.46
Interaction		output modality		
Solitary	SOM		21.88	.45
	WOM		18.43	.49
Collaborative	SOM		25.41	.46
	WOM		19.98	.54
		Time		
		1	31.256	.352
		2	11.601	.318

As illustrated in Table 10, in all conditions the mean scores are lower at time 2, that is in delayed vocabulary post- test.

To better elucidate the point, Figures 1 and 2 should be considered.

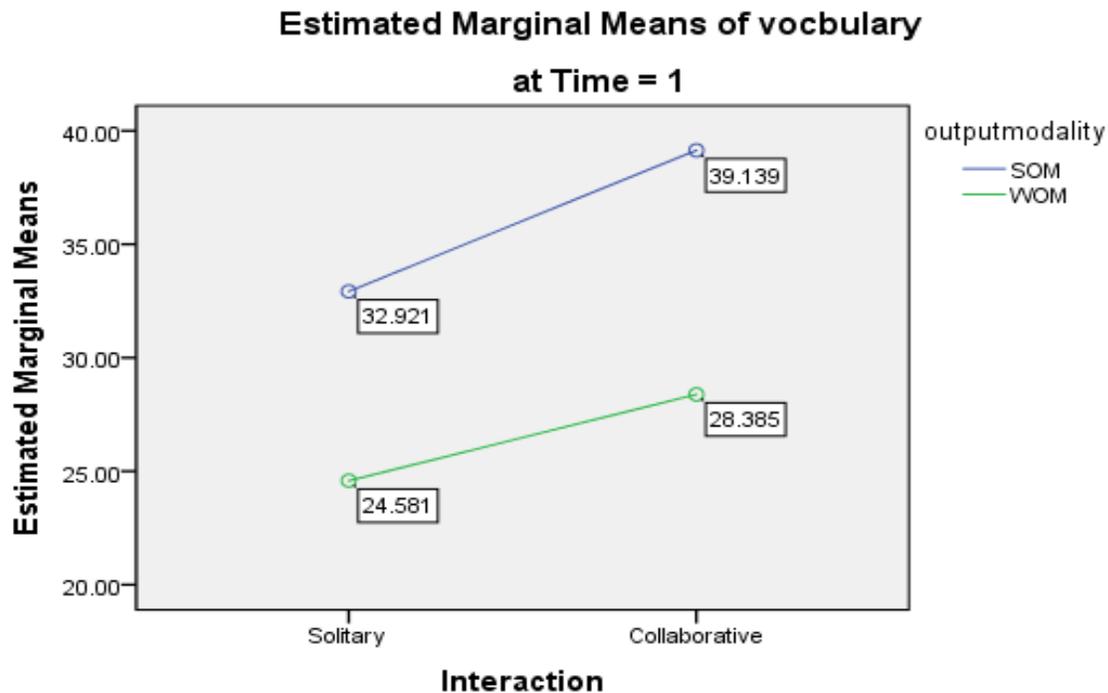


Figure 1. Interaction and output modality effects in time 1 (IVP)

As illustrated in Figure 1, at time 1 (IVP) the collaborative group outperformed the solitary group. It also shows that the SOM group performed better than the WOM.

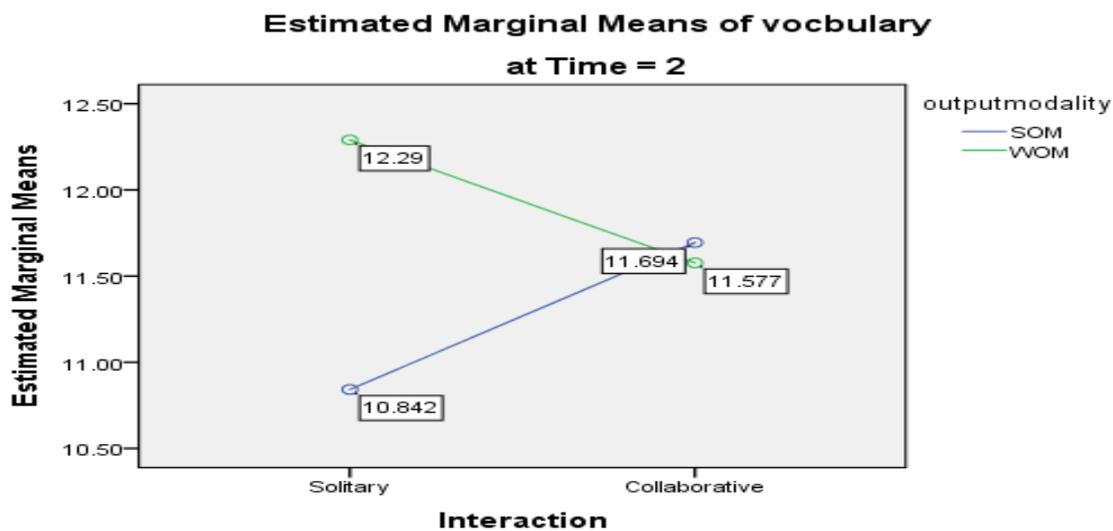


Figure 2. Interaction and output modality effects in time 2 (DVP)

As Figure 2 shows, at time 2 (DVP) the collaborative group was better than the solitary group in SOM not WOM indicating that interaction is effective when the output modality is spoken not written.

To better clarify the point, the performance of groups in terms of output modality and interaction are compared in Time 1 and Time 2 the results of which are presented in Figures 3 and 4.

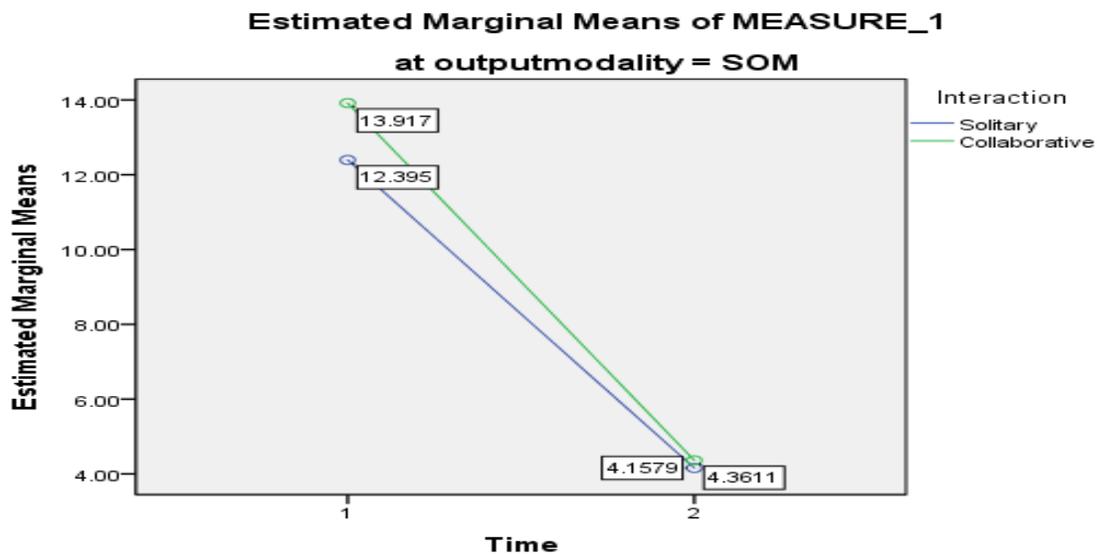


Figure 3. Interaction and SOM gain from time1 to time 2

As Figure 3 shows, the performance of SOM group drops from Time 1 to Time 2.

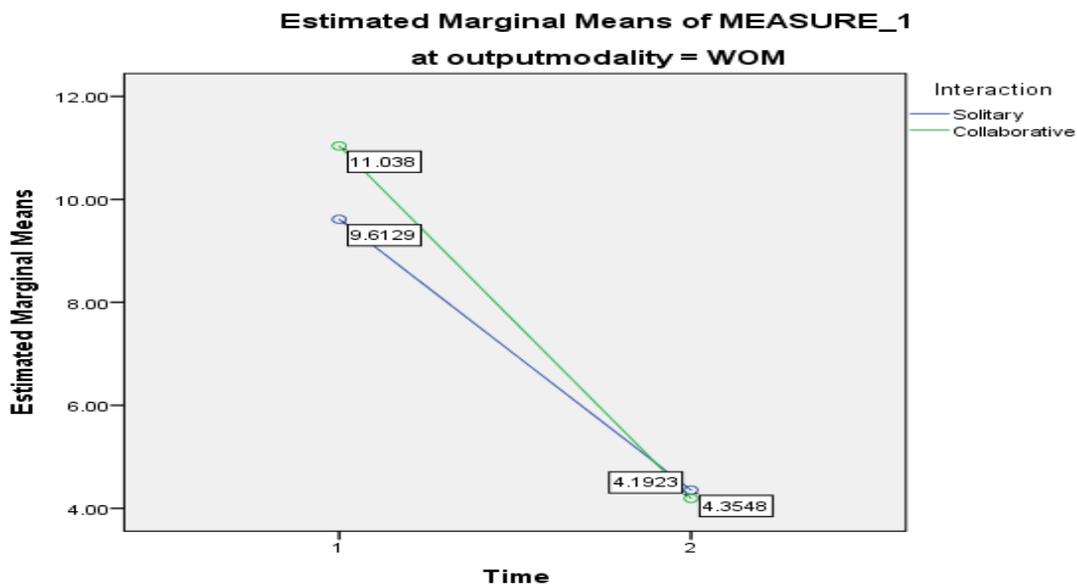


Figure 4. Interaction and WOM gain from time1 to time 2

As figures 3 and 4 show both solitary and collaborative spoken output modality (SOM) and Written Output Modality (WOM) groups experienced a drop in gain scores from time 1 to time 2.

Discussion

The present study aimed at investigating the possible impact of collaborative output on Iranian EFL learners' vocabulary learning and retention compared to the input-only and solitary output. The results showed that COPT group outperformed the other two groups in two post-tests: immediate and delayed posttests. The ANOVA results indicated that the difference among the control and experimental groups was significant in all post- tests. In delayed posttests, although the COPT group outperformed the SOPT, the comparison among the groups showed that the differences did not reach the significance level. The results of the study showed that input-only condition is not optimal for vocabulary learning and retention since the NINO or the first control group had the least mean score. This indicates that incidental vocabulary learning via input alone (reading passage, here) is not sufficient. So, there is a need to use complementary tasks such as collaborative output tasks to get better results in vocabulary learning and retention.

The results of this study regarding the impact of output was consistent with Swain's Output hypothesis. The presence of output significantly improved the performance of the participants in all posttest conditions. So, solitary and collaborative groups outperformed the control group. The collaborative group in all conditions outperformed the other two groups indicating that the impact of output task is dependent on the collaborative nature of the tasks.

The general finding regarding the outperformance of collaborative group in all conditions is in line with Dobao, 2012, 2014; Tajeddin & Jabbarpoor (2013); Kim (2008); Nassaji & Tian, 2010; Shehadeh, 2011; Storch, 1999, 2005, 2008; Storch & Wigglesworth (2007). The present study showed that collaborative output is effective in the short term.

In conclusion, the findings of this study reemphasized the effect of output practice in vocabulary learning and retention. It also supports the using of the collaborative tasks in language courses on vocabulary learning and retention. It, however, makes us to be cautious in this regard. As far as vocabulary learning is concerned, the tasks should be used as early as we expose the learners to the target vocabulary. To get more generalizable results, further research should be conducted using more standard vocabulary tests.

The results of this study regarding the impact of output modality was consistent with Swain's Output hypothesis because the output modality groups outperformed the control group in all conditions. The results confirmed the results of Niu and Park (2014) in immediate posttest. Unlike this study, the difference between written and spoken output modalities was significant. The findings of their study revealed that oral output led to significant performance in all the posttests. In our study, the same thing happened in oral and written output. The oral output means were better than control group and written output group. The reason may be that the written output requires speaking and writing at the same time. In fact, the participants need to negotiate the problems through speaking and then convert their ideas to the written form. That's why the written output groups were no better than the oral groups.

The inconsistent results, if any, could be attributed to the reasons such as, task type effect (cognitive demanding nature of text reconstruction), proficiency level of participants, the confounding effect of the inclusion of both recognition and production levels, the participants' individual differences (unwillingness to get help from the peers), the passiveness of some participants, and uncontrolled variables. Still, another reason might be the nature of reconstruction tasks. The oral and written reconstruction tasks require pooling all the knowledge available to the participants. It is also cognitively demanding. It depends on individual differences and motivational factors. Although one of the researchers observed the research process meticulously, the extent to which the participants contributed with the required modality was not optimal. Some learners were unwilling to participate well.

Another reason could be the impact of test task characteristics. The recognition and production posttests used in this study were adapted to the level of Iranian learners. The results could be different if we used more standard tests with more items.

The results of this study can be interpreted based on Craik and Lockhart's (1972) processing depth theory. They claimed that the chance of a new word being stored in long-term memory is determined by the depth at which that word is processed. In delayed test, this processing was low because the participants were waiting for the end of the term vacations, hence the low processing depth.

Conclusion

Vocabulary is considered one of the important components of second language acquisition and learning. Many complaints are often heard in language classrooms about the best way to learn the words and the way they should be retained whenever needed. Many vocabulary learning strategies have been suggested by researchers.

It goes without saying that output is a determining factor in EFL classrooms in general and reading classes in particular. The modality of this output can make differences in performance if some factors are taken into consideration: The output task should be congruent with the proficiency level of learners. The more cognitively demanding tasks should not be used for less proficient learners. The learners' gender and L1 background should be taken into consideration.

The focus of output modality and interaction in the language classroom can be a new choice for language teachers and learners. A combination of output modality and interaction in the form of small groups and dyads is an aid for learners and teachers. The results of this study support the determining role of interaction and output modality and time. The combined effects of time*interaction, time *output modality, interaction*output modality were confirmed. This showed that these two variables should be incorporated into language textbooks and syllabi.

Interaction or pair work practice while producing output significantly affects the vocabulary learning. The time of the task, the modality of output, and the collaboration pattern are determining factors in immediate vocabulary learning and delayed vocabulary retention.

Pair- work Interaction and output modality were effective in vocabulary learning i.e. in immediate posttest. To put it another way, interaction is helpful on the condition that the time of task be immediate. The spoken output group's performance was better in the short run and written output was conducive to retention when the participants were given sufficient time to practice.

Teachers should incorporate tasks into the curriculum that require language learners to produce output (fill-in-the-blank vocabulary exercises, sentence completion, or composition writing, reconstruction tasks, dictogloss, etc.) to facilitate second language vocabulary development. A combination of these tasks can be a determining factor to guarantee the learning and retention of vocabulary.

Input modality have been considered to make changes in the learning process. The studies suggested that adding to input modality leads to better results. The results of this study, however, showed that the output modality should be a priority in language classrooms. The results of this study suggested that oral modality is the best choice. The written modality is effective if enough time is provided to the learners. A mix of oral and written tasks at different stages of vocabulary leaning seems to be working.

References

- Alanen, R. (1995). Input enhancement and rule presentation in second language acquisition. In R.Schmidt (Ed.), *Attention and awareness in foreign language learning* (pp. 259–302). Honolulu: University of Hawai‘i Press.
- Baddeley, A. (1994). Working memory: The interface between memory and cognition. In D.L. Schachter and E. Tulving (eds.), *Memory systems*. Cambridge, MA: MIT Press, pp. 351–367.
- Baltova, I. (1999). Multisensory language teaching in a multidimensional curriculum: The use of authentic bimodal video in core French. *Canadian Modern Language Review*, 56, 31–48.
- Bisson, M., Van Heuven, W., Conklin, K. & Tunney, R. (2014). The role of repeated exposure to multi-modal input in incidental acquisition of foreign language vocabulary. *Language Learning*, 64(4), 855-877.
- Bygate, M., Skehan, P., & Swain, M. (Eds.). (2001). *Researching pedagogic tasks: Second language learning, teaching, and testing*. London: Pearson.
- Craik, F.I.M., & Lockhart, R.S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behaviour*, 11, 671–684.
- Day, R., O. & Hiramatsu, M. (1991). Incidental EFL vocabulary learning and reading. *Reading in a Foreign Language*, 7(2), 541-551.
- Day, R. R., & Swan, J. (1998). Incidental learning of foreign language spelling through targeted reading. *TESL Reporter*, 31(1), 1-9.
- Dehqan, M., & Mohamadi, M. (2017) Collaborative Output Tasks and their Effects on Learning English Comparative Adjectives. *Journal of Teaching Language Skills (JTLS)* 36(1), 2017, 1-26.
- DeKeyser, R., & Sokalski, K. (1996). The differential role of comprehension and production practice. *Language Learning*, 35, 613–642.
- De la Fuente, M. J. (2002). “Negotiation and oral acquisition of L2 vocabulary. The roles of input and output in the receptive and productive acquisition of words.” *Studies in Second Language Acquisition*, 24, 81–112.
- Dobao, A. (2012). Collaborative writing tasks in the L2 classroom: comparing group, pair, and individual work. *Journal of Second Language Writing*, 21, 40–58.
- Dobao, A. (2014). Vocabulary learning in collaborative tasks: A comparison of pair and small group work. *Language Teaching Research 2014*, 18(4) 497–520.
- Dobao, A., & Blum, A. (2013). Collaborative writing in pairs and small groups: learners’ attitudes and perceptions. *System*, 41, 365–378.
- Ellis, R. (1999). *Learning a second language through interaction*. Amsterdam/Philadelphia: John Benjamins.
- Ellis, R. (2003). *Task-based language learning and teaching*. Oxford: Oxford University Press.
- Ellis, R. & He, X. (1999). The roles of modified input and output in the incidental acquisition of word meanings. *Studies in Second Language Acquisition*, 21, 285–301.
- García Mayo, M.P. (2002a). The effectiveness of two form-focused tasks in advanced EFL pedagogy. *International Journal of Applied Linguistics*, 12, 156–75.
- García Mayo, M.P. (2002b). Interaction in advanced EFL pedagogy: A comparison of form-focused activities. *International Journal of Educational Research*, 37, 323–41.
- García Mayo, M.P. (2017). Lexical language -related episodes in pair and small group work. *IJES*, 17 (1), 2017, 61–82.

Hill, M., & Laufer, B. (2003). Type of task, time-on-task and electronic dictionaries in incidental vocabulary acquisition. *IRAL*, 41, 87-106.

Hulstijn, J. H. (2001). Intention and incidental second language vocabulary learning: a reappraisal of elaboration, rehearsal and automaticity. In P. Robinson (ed.), *Cognition and second language instruction* (pp. 258–86). Cambridge: CUP.

Hulstijn, J., & Trompetter, P. (1998). Incidental learning and second language vocabulary in computer-assisted reading and writing tasks. In Albrechtsen, D., Henriksen, B., Mees, I. & Poulsen, E. (Eds.), *Perspectives on foreign and second language pedagogy*, pp. 191-202. Denmark: Odense University Press.

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

Kelly, P. (1992). Does the ear assist the eye in the long-term retention of lexis? *IRAL*, 30, 137–145.

Kim, Y. (2008). The contribution of collaborative and individual tasks to the acquisition of L2 vocabulary. *The Modern Language Journal*, 92, 114–130.

Kowal, M. & Swain, M. (1994). Using collaborative language production tasks to promote students' language awareness. *Language Awareness*, 3(2), 73-93.

Kuiken, F., & Vedder, I. (2002). The effect of interaction in acquiring the grammar of a second language. *International Journal of Educational Research*, 37, 343–358.

Kuiken, F., & Vedder, I. (2011). Task performance in L2 writing and speaking: The effect of mode. In P. Robinson (Ed.), *Second language task complexity: Researching the Cognition Hypothesis of language learning and performance* (pp. 91–104). Amsterdam: Benjamins.

Laufer, B. (2003). Vocabulary acquisition in a second language: Do learners really acquire most vocabulary by reading? Some empirical evidence. *Canadian Modern Language Review*, 59, 565–585.

Laufer, B., & Hulstijn, J. (2001). Incidental vocabulary acquisition in a second language: The construct of task-induced involvement. *Applied Linguistics*, 22, 1–26.

Laufer, B., & Nation, P. (1995). Vocabulary Size and Use: Lexical Richness in L2 Written Production. *Applied Linguistics*, 16, 307-322.

Leahy, W., P. Chandler & J. Sweller. (2003). When auditory presentations should and should not be a component of multimedia. *Applied Cognitive Psychology*, 17, 401–418.

Long, M.H., 1983. Native speaker/non-native speaker conversation and the negotiation of comprehensible input. *Applied Linguistics*, 4 (2), 126-141.

Lund, R. (1991). A comparison of second language listening and reading comprehension. *The Modern Language Journal*, 75, 196–204.

Mayer, R. E., & Anderson, R. B. (1992). The instructive animation: Helping students build connections between words and pictures in multimedia learning. *Journal of Educational Psychology*, 84, 444-452.

Milton, J. (2009). *Measuring second language vocabulary acquisition*. N.Y: Multilingual Matters.

Murphy, V.A. (1997). The effect of modality on a grammaticality judgement task. *Second Language Research*, 13, 34–65.

Nabei, T. (1996). Dictogloss: Is it an effective language learning task? *Working Papers in Educational Linguistics*, 12, 59–74.

Nassaji, H., & Tian, J. (2010). Collaborative and individual output tasks and their effects on learning English phrasal verbs. *Language Teaching Research*, 14, 397–419.

Nassaji, H., & Tian, J. (2014). The Role Of Language Coproduction In Learning English Vocabulary. *Procedia - Social and Behavioral Sciences*, 143 794 – 798.

Nation, I.S.P. (2001). *Learning Vocabulary in Another Language*. Cambridge: Cambridge University Press.

Niu R. & R. H-Park. (2014). Interaction, modality, and word engagement as factors in lexical learning in a Chinese context. *Language Teaching Research*, 18(3) 345–372.

Nobuyoshi, J., & Ellis, R. (1993). Focused communication tasks and second language acquisition. *English Language Teaching*, 47, 203–210.

Paribakht, S. & Wesche, M. (1996). Enhancing vocabulary acquisition through reading: A hierarchy of text-related exercise types. *Canadian Modern Language Review*, 52, 155–78.

Penney, C. (1989). Modality effects and the structure of short-term verbal memory. *Memory and Cognition*, 17, 398–422.

Pichette, F. (2002). Second-language vocabulary learning and the additivity hypothesis. *Canadian Journal of Applied Linguistics*, 5, 117–130.

Rott, S., Williams J., & Cameron, R. (2002). The effect of multiple-choice L1 glosses and input- output cycles on lexical acquisition and retention. *Language Teaching Research*, 6 (3), 183-222.

Sadoski, M. (2005). A dual coding view of vocabulary learning. *Reading & Writing Quarterly*, 21(3), 221–238.

Sadoski, M., & Paivio, A. (2001). *Imagery and text. A Dual Coding Theory of Reading and Writing*. Mahwah, New Jersey, London: Lawrence Erlbaum Associates.

Schmitt, N. (2008). Instructed second language vocabulary learning. *Language Teaching Research*, 12, 329–363.

Shehadeh, A., & Coombe, C. (eds.) (2010). *Applications of task-based learning in TESOL*. Alexandria, VA: TESOL.

Shih, Y. & S.M. Alessi. (1996). Effects of text versus voice on learning in multimedia courseware. *Journal of Educational Multimedia and Hypermedia*, 5, 203–218.

Soleimani, H. & Mahmoudabadi Z. (2014). The Impact of Interactive Output Tasks on Developing Vocabulary. *Iranian Journal of Applied Linguistics (IJAL)*, 17(2), 93-113. Knowledge of Iranian EFL Learners

Solman, R. T., Singh, N. N. & Kehoe, E. J. (1992). Pictures block the learning of sightwords. *Educational Psychology*, 12, 143-153.

Storch, N. (1998). A classroom-based study: Insights from a collaborative text reconstruction task. *ELT Journal*, 52, 291–300.

Storch, N. (1999). Are two heads better than one? Pair work and grammatical accuracy. *System*, 27, 363–374.

Storch, N. (2001). How collaborative is pair work? ESL tertiary students composing in pairs. *Language Teaching Research*, 5, 29–53.

Storch, N. (2002a). Patterns of interaction in ESL pair work. *Language Learning*, 52, 119–158.

Storch, N. 2002b. Relationships formed in dyadic interaction and opportunity for learning, *International Journal of Educational Research*, 37, 305–22.

Storch, N. (2005). Collaborative writing: Product, process and students' reflections. *Journal of Second Language Writing*, 14, 153–173.

Storch, N. (2007). Investigating the merits of pair work on a text editing task in ESL classes. *Language Teaching Research*, 11, 143–159.

Storch N, & Aldosari A. (2013) Pairing learners in pair-work activity. *Language Teaching Research*, 17(1), 31–48.

Storch, N., & Wigglesworth, G. (2007). Writing tasks: Comparing individual and collaborative writing. In M. P. Garcı́a Mayo (Ed.), *Investigating tasks in formal language learning* (pp. 157–177). London: Multilingual Matters.

Swain, M. (1985). Communicative competence: Some roles of comprehensible input and comprehensible output in its development. In S. Gass & C. Madden (Eds.), *Input in second language acquisition* (pp. 235-256). New York: Newbury House.

Swain, M. (1995). Three functions of output in second language learning. In G. Cook & B. Seidlhofer (Eds.), *Principle and practice in applied linguistics: Studies in honour of H.G. Widdowson* (pp. 125-44). Oxford: Oxford University Press.

Tajeddin, Z. & Jabarpoor, SH. (2013). Individual and collaborative output tasks: effects on the acquisition of English inversion structures. *RALs*, 4(2), 16- 32.

Tian-X. Y., Richard J. A., Qi-J. Y. & Raymond C. K. C. (2015). The influence of input and output modality on following instructions in working memory. *Scientific Reports*, 5, 17657, DOI: 10.1038/srep17657.

Webb, S. (2005). Receptive and productive vocabulary learning: The Effects of Reading and Writing on Word Knowledge. *Studies in Second Language Acquisition*, 27, 33-52.

Wigglesworth, G., & Storch, N. (2009). Pair versus individual writing: Effects on fluency, complexity and accuracy. *Language Testing*, 26, 445–466.

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

Wu, H. M. & Solman, R.T. (1993). Effective use of pictures as extra-stimulus prompts. *British Journal of Educational Psychology*, 63, 144-160.