Writing as a Learning Activity

Edited by

Perry D. Klein
Pietro Boscolo
Lori C. Kirkpatrick
Carmen Gelati

BRILL

LEIDEN | BOSTON

Contents

Writing as a Vocabulary Learning Tool

Kenan Dikilitaş and Jerome C. Bush

1   Introduction

In the early to mid-1970s, researchers and authors began to examine the impact of writing on learning (Britton, 1972; Emig, 1977). Since that time, the research on writing-to-learn has shown that, although writing may produce positive learning results, the relationship of writing to learning is complex (Tynjälä, 2001b). Recent research has shown that writing collaboratively leads to construction of knowledge (Nykopp, Marttunen, & Laurinen, this volume) and that writing is a multi-faceted activity that assists in the development of cognitive abilities as well as the understanding and memorization of content knowledge (Bazerman, Simon, & Pieng, this volume). However, writing in a second language adds an additional level of complexity as students are developing linguistic skills in addition to satisfying communicative and educational goals.

The bulk of the research on writing in a second language has been produced fairly recently, since the 1980s (Polio, 2003). Most of this research falls into four categories: research about the texts being produced, the process (or processes) being used, the participants, or the context. The research about writing in a second language has been focused mostly on developing the ability to produce acceptable written products, not on the impact writing has on language learning in general. However, there is a growing body of research on the impact of writing on other linguistic competencies, such as vocabulary (Grabe & Kaplan, 1996; Frodeson & Holten, 2003).

Learning vocabulary is a challenge for most students. Vocabulary acquisition competes for time with other aspects of language learning. Additionally, a large amount of vocabulary needs to be acquired for students to be effective communicators, and vocabulary knowledge is interrelated in complicated ways (Zimmerman, 2009). Moreover, teachers have a limited time in which to provide instruction and many other responsibilities such as lesson planning and evaluation (Nation, 2008). Because both teachers and students have so much to accomplish in a limited amount of time, the need to be strategic and efficient in learning second language vocabulary is clear.

Oxford (1990) identified six broad categories of learner strategies: cognitive, meta-cognitive, memory-related, compensatory, affective, and social (Ehrman & Oxford, 1990). N. Schmitt & McCarthy (1998) adapted Oxford's general categories to the specific area of vocabulary to create a taxonomy of strategies for L2 vocabulary acquisition. While a full discussion of vocabulary learning strategies is beyond the scope of this paper, it is important to establish that research has indicated a strategic orientation towards vocabulary acquisition is beneficial. This paper will describe a study examining vocabulary from a strategic perspective. Particularly, the impact of writing exercises on the acquisition of second language vocabulary. Additionally, the differences between receptive and productive vocabulary learning will be explored. This is related to implicit and explicit instruction. Implicit instruction was popularized by Krashen's (1985) Input Hypothesis, which was rebutted by Swain (1985). The results of the current study suggest that teachers and students make use of both of these orientations.

1.1 *Second Language Writing and Vocabulary Acquisition*

A number of studies have confirmed the positive impact of compositional writing on vocabulary learning. Muncie (2002) argued that L2 writing in certain contexts can be seen as a tool for general language improvement, and it could be especially beneficial for vocabulary development. Among the reasons for this are that students take more time in writing, which allows for greater experimentation with new words than speaking does. Writing also allows for greater use of resources such as dictionaries, the internet, and peer reviews. This may help students to activate less frequent but more appropriate words which may be in their passive vocabulary, but not yet fully part of their active vocabulary (Corson, 1997). Corson contrasts "active" and "passive" vocabulary and says that full vocabulary learning has not taken place until words are available for active use. According to Corson, if students are not able to experiment with low-frequency, academic words in activities such as writing, they will not be able learn the rules for proper usage of such words.

Coomber et al. (1986) listed three factors that may account for the positive influence of writing on vocabulary learning: the use of the words in meaningful contexts, the students' utilization of their higher level cognitive functions, and slow nature of writing that increases time for elaboration on lexical knowledge. Hulstijn and Laufer (2001) assert that the reason writing assists in vocabulary acquisition is that writing calls for deeper processing than other forms of practice. They have developed the Involvement Load Hypothesis, and ranked the level of involvement required of students to complete various activities. In short, their contention is that the deeper the involvement, the better the retention. In their study, those students who engaged in a letter writing

activity had roughly double the vocabulary retention of students who studied vocabulary through reading and reading with a fill-in task. The letter writing activity required the students to become more involved with the target vocabulary, which led to learning gains over the other two groups.

In addition to compositional writing, other writing types have been examined in regard to their impact on vocabulary acquisition. Many of these studies favor repetition. For example, Folse (2006) used three different activities in his study: a fill-in-the-blank exercise performed once, fill-in-the-blank exercises performed three times, and writing original sentences. The study found that students who did the fill-in-the-blank exercises three times outperformed the other two groups. In another study, a repetitive writing homework assignment was found to have a positive effect on beginning Iranian EFL learners' vocabularies (Dehghani, Motamadi, & Mahbudi, 2011). Likewise, Maftoon (2006) supports the idea that vocabulary learning is made more effective by seeing the written form of a word and repetitively using it in context.

However, Barcroft (2004, 2006) found that repetitive writing exercises inhibit vocabulary learning by forcing students to consider other aspects of word knowledge and language. Barcroft's rationale is that the students have limited processing capacities. By using their processing capacity to produce sentences with new words, they are not focusing their full attention on learning the target vocabulary. Writing sentences requires working with syntax and semantic knowledge, not simply lexical knowledge. Barcroft noted, however, that while the semantic elaboration involved in sentence production can have an inhibitory effect during the initial stages of learning a word (the "form learning" stage), it could facilitate memory for known words (the "semantic learning" stage). In addition, Barcroft made a critical evaluation of the diverse findings regarding writing and vocabulary acquisition and attributed the contrasting findings to learners' exposure to competing conditions, use of receptive versus productive measures and methodological limitations.

Although writing has been found to function as a tool for learning vocabulary, it is also important to realize that vocabulary skill plays a key role in constructing a text (Leki and Carson, 1994). Vocabulary is a key component of language, and writing in a foreign language is difficult when there is insufficient vocabulary. Therefore, many students are interested in developing their vocabulary so that they can be better writers. This suggests a bi-directional relationship between vocabulary and writing.

1.2   *Lexical Knowledge: Productive vs. Receptive*

A distinction between productive and receptive vocabulary is often made in the literature. Receptive vocabulary is the words which are understood when read-

ing or listening. Productive vocabulary is the words one can use while speaking or writing. The tasks teachers use while teaching vocabulary tend to be receptive rather than productive (Nagy, Herman, & Anderson, 1985; Nagy, Anderson, & Herman, 1987). Common teaching practices include providing a word along with a definition, matching words with their meanings, giving an example sentence that uses the word, guessing from context, learning from word pairs, or reading from a dictionary. In addition, some teachers provide other information such as synonyms and antonyms of the words, collocations, spelling, grammatical features, morphological features, or pronunciation. Receptive tasks are more commonly used and more popular than productive tasks mainly due to the simplicity they provide in designing, grading and completing assignments. Productive tasks, such as cloze exercises or sentence creation, are much less common.

It has been argued in a number of empirical studies that L2 learners have more difficulty with developing productive than receptive vocabulary (Laufer and Paribakht, 1998; Nation 2001; Webb, 2008; Zheng, 2009). Studies on learning from word pairs indicate that teaching style (receptive or productive) has a bearing on the type and amount of knowledge gained (G. Griffin & Harley, 1996; Waring, 1997b). In other words, receptive learning tends to lead to significantly more receptive knowledge of words, just as productive learning may lead to more productive knowledge.

Due to the simplicity of instruction and design, teachers use receptive tasks and test learners' vocabulary knowledge through receptive exercises. This may account for the relatively large receptive vocabulary learners develop compared to their productive vocabulary (Laufer & Paribakht, 1998; Waring, 1997a). Convenience for the teacher may come at the expense of the learners, who are generally trying to learn how to communicate in a second language.

1.3   *Lexical Knowledge: Breadth vs. Depth*
Learning vocabulary is not a simple act of memorizing a meaning. Nation (1990) lists a number of aspects that need to be acquired by a learner and that make word knowledge: meaning, associations, collocations, grammatical patterns where a word is to be used, frequency of use and orthography. Zimmerman also wrote about the complex and incremental nature of vocabulary teaching and learning (Zimmerman, 2009). Knowing the various aspects of a word, and how to use that word in multiple contexts, is considered deep knowledge. Reciting a definition for a word, without being able to use that word accurately, is an example of shallow knowledge.

The breadth of vocabulary knowledge required by second language students to produce texts has also been discussed in the literature. It has been found that

the size of a vocabulary can have an impact on the quality of a learner's written work (Laufer & Nation, 1995). Laufer and Nation (1995) admit that vocabulary size is not the only determinant of the quality in second language writing. However, a positive correlation between vocabulary size and writing proficiency has been observed. Muncie (2002) found that Japanese students will use more (and more advanced) vocabulary in subsequent versions of a written project using the process writing approach. This shows again that the amount of vocabulary used is related to the quality of the text being produced. This underscores the importance of vocabulary learning. According to Nation (2001), "In general, high-frequency words are so important that anything that teachers and learners can do to make sure they are learned is worth doing." (p. 16). Proficiency in a language requires knowledge of a wide enough vocabulary at a deep enough level to productively communicate.

2 The Current Study

After extensive review of the available literature it became apparent that the impact of the cognitive and productive task of writing on second language vocabulary acquisition was not fully explored. Furthermore, enough evidence was found in the literature to suggest that cognitive and productive exercises, such as writing, could have a positive effect. Therefore, the current study was designed to measure the impact of writing personally created texts on vocabulary acquisition. The study follows a quasi-experimental design and may rightly be called an intervention study. True random assignment was not done. For the purposes of clarity, we will be referring to the two groups as the "writing group" and "control group". The following hypotheses were developed:

1. The writing group will demonstrate that they recognize the word definitions better by performing better on a matching test.
2. The writing group will demonstrate a deeper understanding of the vocabulary items by performing better on sentence writing exercises.
3. The writing group will experience greater overall vocabulary gains.

3 Method

3.1 Setting
The study was conducted at Gediz University in Izmir, Turkey. The university was established in 2009 and now offers courses in 4 faculties in which the

language of education is English. To prepare students for this academic journey, the preparatory school teaches academic English. Nearly 500–600 students attend the preparatory school each year. The students, with or without scholarship, are admitted to the university through a national placement test. Once they have enrolled in the school, they are required to graduate from the English language program at B2 level to be eligible to start their majors in the faculties. The preparatory program is composed of 4 quarters, each lasting 8 weeks. The students finish one level per quarter. In each quarter, the students are given 4 achievement tests, one reader quiz, weekly vocabulary and speaking quizzes as well as one exit test. The combined results of these tests are averaged to determine the grade of the students. The students go through an extensive exam-based process during the year. The program aims to develop students' ability to use written and oral language productively as they need to listen to their professors, communicate with them and submit weekly reports and assignments in English. A student starting the program at A1 level is supposed to graduate at B2 level in an 8-month period.

3.2   *Participants*

The study was conducted with 32 adult university students. Ages ranged from 18 to 20 in both groups. All the students were Turkish, except for two females who were from Mali. Students were planning on majoring in engineering, administration, or architecture. The control group had five females and eleven males, and the writing group had seven females and nine males. As the students were taking pre-sessional courses at the time of the study, they had no GPA's. The students were not randomly assigned to one of the groups formed for the current research. However, the classes had no special distinctions. In other words, although purposeful randomization did not occur, the two classes could be said to be somewhat similar because they had gone through the same placement process and were assigned to classes at the same level. Therefore, the sampling method was purposeful as well as convenient.

When the students entered the preparatory program at the university, they were given a placement test and placed as B1. At the beginning of the study, students in both groups had just begun level B2, in line with the common European framework. At Gediz University, this is called the "upper intermediate" level. They had been in the program for 2 and half months before the current class. The writing class consisted of 17 students. However, one had to be excluded because of frequent absence. Other students completed all the phases of study, attending the lessons and the activities, and completing the tasks given. The control class had 16 students, one of whom was absent on the day of the matching exam. However, it was decided not to exclude the student

from the study because the written data was deemed to be valid and important.

3.3   *Materials*

The words in this study came from Reading Explorer 2 (MacIntyre, 2009) and Reading Explorer 3 (Douglas, 2010). The materials used in the reading instruction consisted of 16 units that were covered in a period of eight weeks. The intervention was carried out in 128 hours completed in 32 days in total. In both groups the same books and units were used during the same amount of time. The words targeted for the current research were the ones that the book originally highlighted and aimed to teach. The vocabulary and activities were designed for the proficiency level of the participant students. Each unit in each book has 10 high-frequency target vocabulary words. The units start out with a pre-reading exercise that includes pictures, maps, and discussion questions. Then the books present reading passages that contain and highlight the target words accompanied by colorful pictures. After that is a page of comprehension exercises followed by a page of vocabulary exercises. The vocabulary section has exercises such as gap filling in a reading, manipulation of target words, and finding the appropriate definition in a context.

3.4   *Measures*

A scale for measuring vocabulary knowledge was required. Paribakht and Wesche (1992) developed the Vocabulary Knowledge Scale, or VKS. The scale was refined and published again in 1997 (Paribakht & Wesche, 1997). Since that time the scale has been widely used in research on second language vocabulary acquisition. One of the values of the scale is that the differences between the categories used in the scale are large enough to be self-perceived, yet small enough that gains can be seen within a relatively brief instructional period. This aspect of the scale was important for the current study. The five levels of the VKS are as follows:

1. The word is not familiar at all.
2. The word is familiar, but its meaning is not known.
3. A correct synonym or translation is given.
4. The word is used with semantic appropriateness in a sentence.
5. The word is used with semantic appropriateness and grammatical accuracy in a sentence.

The five points are meant to represent a continuum of word knowledge that goes from unknown to fully known.

Meara (1996) was critical of the VKS, stating that the levels suggest a progression, and vocabulary acquisition doesn't always follow a standard linear progression. According to Meara, it is possible for a student to write a correct sentence without full understanding of the target word. Students can just reproduce the word in the context in which it was first learned or a set phrase and earn a high rating on the VKS. Therefore, Meara did not consider the VKS fully reliable, and suggested that it is better to focus on the development of the lexicon than on decontextualized words. Meara's points are well taken, but the current study is focused on following a discrete list of words, which is what the VKS was designed to do.

Perhaps the strongest criticisms of the VKS appeared in an article by Bruton (2009). The major shortcoming of the VKS was found to be that it is a single instrument designed to test both receptive and productive knowledge. Bruton argued that different tests are needed to measure receptive and productive vocabulary. The categories were also found to be inadequate in measuring partial word knowledge. Furthermore, a large number of studies used modifications of the VKS, suggesting serious flaws in the original form.

These arguments are certainly well thought out and logical. However, Paribakht and Wesche (1997) established an acceptable level of reliability (r = .89) using a test-retest method (Paribakht & Wesche, 1997, p. 180). Therefore, for the current study it was deemed an appropriate basis for the development of an assessment tool. Even Bruton (2009) conceded, "… the quantification of the scores on the VKS is restricted to totals per nominal category, which could be used to plot percentage changes per category over time (p. 295)." Because the goals of the current study had to do with plotting changes in nominal categories over time, it was decided to adopt the VKS, with modifications. Although a wide variety of vocabulary assessment tools exist (see N. Schmitt, 2010), the VKS, modified to compensate for the criticisms put forth in the literature, was deemed most appropriate for this study.

3.5   *Pre-Course Vocabulary Knowledge Self-Assessment*

In order to establish that vocabulary learning had taken place in this study, it was necessary to determine how well the students knew the target words prior to taking the reading class. To control for the practice effect (participants learning from the pre-test), a self-reported measure was used to determine the initial level of word knowledge. Although many researchers are critical of self-reported measures, self-ratings have been found to be both reliable and valid measures of communicative language abilities (Bachman & Palmer, 1989; Ross, 2006). Self-reporting has been found to be accurate when participants are assessing what they do not know and less accurate when reporting what they

do know (Heilman & Eskenazi, 2008). As the current study is about learning vocabulary, the most important distinction to make was which words were unknown to the participants at the start of the study.

Self-reporting's main threat to validity occurs when something is to be gained (e.g. grades, placement, or exemption from a task) from a high rating (Saito, 2003). In the current study however, the vocabulary exercises were simply homework, with no reason to give a false high rating. Therefore, it was determined that a self-reported measure would be the most appropriate tool to determine initial vocabulary knowledge. To this end, the researchers developed a scale, which they call the Pre-Exercise Self-Reported Vocabulary Evaluation Scale (or PESVES). The scale is based on the VKS, but differs from the VKS in that the VKS uses both self-reported and observed responses, while the PESVES uses only self-reported responses. The PESVES is a self-evaluation rubric (see below) designed to be easy for students to determine their level of lexical knowledge. The 160 target words were each given a score by the participants using the PESVES.

The advantage of the Pre-Exercise Self-reported Vocabulary Evaluation Scale (PESVES) over the VKS is that the PESVES can be used with a large number of vocabulary words. In the current study, 160 vocabulary words were targeted for study by 32 students, giving a total number of 5,120 items. Using the unmodified VKS with the same two classes would require the evaluation of over 3,000 sentences by the teacher. The amount of time required to hand score the sentences would be burdensome. Additionally, it was assumed that many of the words were unknown. This could lead to many of the students refusing to write sentences, which would have an impact on the results.

3.6   *Matching Tests*

The post-test consisted of two parts, a matching exercise and written sentence production. To this end, a simple matching test was devised. The vocabulary

TABLE 1   *The Pre-Exercise Self-reported Vocabulary Evaluation Scale (PESVES)*

Pre-Exercise Self-reported Vocabulary Evaluation Scale

0   No answer
1   I have never seen this word before.
2   I have seen this word, but I am not sure I know what it means.
3   I know the meaning of this word, but I am not sure I always use it correctly.
4   I know what it means and I can use it easily.

items were grouped in units of ten and definitions were obtained or adapted from the online version of Longman's Dictionary of Contemporary English (http://www.ldoceonline.com/). The vocabulary words had blanks in front of them and the definitions had letters in front of them. The participants were asked to put the correct letter in the blank in front of the vocabulary word. This was done to determine if the participants could recognize the word-meaning connection at a fairly low level of receptiveness. Both the words and definitions were provided; the participants just had to recognize which definition went with which word. Such an activity could be termed passive recognition of vocabulary (Laufer, Elder, Hill, & Congdon, 2004).

3.7 *Sentence Writing Exercises and Evaluation Rubric*

The second part of the post-test required students to write complete sentences using the target vocabulary. A sheet was prepared with all 160 of the vocabulary items. The sheet had a blank line next to each word on which to write a sentence. The students were instructed to write sentences of ten words or more. The reason for this was to give the raters ample evidence by which to assess the accuracy of the word knowledge. For example, from a sentence such as "It is a crisis." the depth of the student's vocabulary knowledge of the word "crisis" is hard to infer.

A rubric was created to analyze the sentences, the Productive Vocabulary Evaluation Rubric (PVER). The PVER's levels correspond with the levels in the PESVES.

TABLE 2   *Productive Vocabulary Evaluation Rubric (PVER)*

0	Student didn't write any sentence or answer, or had the word confused with another word.
1	Completely incorrect usage, doesn't show understanding of the meaning or semantics.
2	Shows low understanding of the meaning of the word, may understand grammatical function or part of speech.
3	Shows understanding of the meaning, but uses the word incorrectly. May use incorrect collocations or awkward usage (Example for the target word **'tight'**: "the road became tight." Instead of "The road became narrow.").
4	Shows an understanding of the meaning and usage of the word. Ideally, students are using the word in a complicated way such as changing the part of speech, using idiomatic expressions, collocations, or adding inflectional or derivational morphemes.

3.8   *Procedures*

Before commencing each reading chapter, the students in both classes were given the self-reported measure with the ten target vocabulary words for the chapter to determine existing word knowledge. During the reading lessons, the students were taught the target words directly in an identical fashion by the same teacher (one of the authors of this study). The instructor followed a student-centered approach to increase comprehension by giving pair work or group work tasks during which students summarized each paragraph and discussed the content of the paragraph with other groups. This allowed the students to take the ownership of their own learning and gave them the opportunity to discover and create the meanings of the words through negotiation. It also lowers their dependency on teachers as the main source of knowledge. In student-centered classroom students do not necessarily expect teachers to approve, correct, advise or praise them (L. Jones, 2007). Rather, they construct knowledge through cooperative and collaborative activities with their peers in the classroom.

The grouping arrangements during the instruction were not tracked. In each lesson the students formed pairs or groups without any teacher interference. The vocabulary words in Reading Explorer's texts are highlighted and students were encouraged to become aware of the context of the words and their collocations. The students were encouraged to use the newly learnt vocabulary in the classroom. Due to time limits imposed by the syllabus, the students were not asked to use all the target words, but rather each pair of students were given one or two words to use in one or two sentences. After writing the sentences, each pair or group shared their sentences with others by reading aloud, and received immediate feedback on the accuracy of the use of the words. The classroom activity was designed to enhance their confidence to use the target vocabulary in their homework assignments.

The intervention was conducted by one of the authors. He was the instructor of the reading course and has 15 years of experience of teaching adult university students. The nature of the intervention was developed by frequent interaction between the two researchers, who critically discussed the activities to be carried out in these reading courses.

In addition to the tasks above, the writing class had two writing tasks. During the classes, as the first post-reading task, the students were trained to use the target vocabulary in sentences or short original compositions. These compositions varied in length from a few sentences to multiple paragraphs. Variation in style also occurred, some of the students wrote dialogs, others wrote narratives or expository pieces. This was done in pairs or groups of three. Each pair or group was given two of the vocabulary words by the teacher. The

students were then asked to write a meaningful piece of text with one or two sentences using the target words from the chapter. The writings were discussed and compared by the whole class. This activity gave students the opportunity to consider the surrounding contexts of the words and better comprehend the meanings so they could improve their productive skill. This activity took about ten minutes of class time to complete.

As the second post-reading task, the students were asked to write a paragraph at home in which they used all 10 of the target words together in the same text. The students had to submit their writing the day after the assignment was given. As the highlighted vocabulary words in the reading lesson chapters were thematically related, it was fairly easy for them to create contexts. Students reported that the writing homework took about 30 to 60 minutes to complete. The control group performed neither of these extra activities in the classroom or at home. In order to control the instructional time that both groups were given, the control group students were engaged in activities given in the extra materials of the course book that allowed them to do word matching, gap filling, listening and video watching relevant to the chapter.

At the end of the course, after the eighth week, the students were given identical tests designed to measure the depth of their word knowledge. The first test was a matching test. All 160 words were given in 16 groups of ten words each on one exam. Both the target words and the definitions were provided. The students just had to select the correct definition from a list and write the letter associated with the definition in a blank next to the target word. Ample time was allotted and every student finished the matching exam. Scoring was done by simply marking the number of incorrect matches and subtracting from 160 to give the number of correct matches.

After that, the sentence writing assignment was done in two sessions. During each session the students were asked to write 80 sentences for a total of 160 sentences. Again, ample time was given; however, many students did not write all of the sentences. Roughly 5,000 sentences were generated and analyzed in the current study. Two independent analyses were conducted, one by each of the authors, and the results were compared. The authors are experienced teachers who have worked with foreign language students for 11 and 15 years respectively. One of the researchers is a native speaker of English, while the other is a non-native EFL teacher. This yielded a number of different judgments and intuitions regarding the assessed word knowledge. The native speaker's intuition contributed to the evaluation of idiomatic uses and nuances of meaning expressed in the sentences, while the non-native speaker's judgments about the degree of accuracy and meaning enriched the interpretations regarding the rating process. The non-native teacher was aware of what the students were

attempting to convey more often than the native speaker. The raters reflected upon the rubric and insightfully discussed what each category in the rubric referred to. However, discussion over specific words was kept to a minimum. A 77% inter-rater reliability was established. This would be considered "substantial agreement" according to the Landis-Koch kappa benchmarking scale (Landis & Koch, 1977). The Landis-Koch scale has six levels, "substantial" is level five, just under "almost perfect". This indicates general agreement between the raters and that the rubric was fairly well understood by both of them. It should be noted that, to control for researcher bias, when the raters disagreed, the lower of the two choices was used.

4   Results

The data were tabulated in Excel and then entered into SPSS for statistical analysis. Because this study was done with a quasi-experimental design, t-tests were chosen as the best way to analyze the results. T-tests are commonly used to compare the difference between the means of two sample populations. The current study compared two groups, so ANOVA or other test was not used. The assumption of normality was met in every case of statistical difference by the Shapiro-Wilk test as well as the KS test. Levene's test indicated an equality of variances. The confidence level was set at 95% (p = .05), which is a normal level to control for type I errors. The descriptive statistics of the two groups and the significance levels of the t-tests can be seen in Table 3 below.

The PESVES scores constitute the pre-test or initial position of the participants. The higher PESVES scores at level one and two and the lower PESVES scores at levels three and four of the control group indicate that the control group students were less familiar with the target words than those of the writing group at the beginning of the intervention. The difference was not found to be statistically significant. Because the two groups are within the normal range of similarity, differences in the PVER (the post-test sentence writing) cannot be attributed to pre-existing vocabulary knowledge. The significant differences in the PVER scores are most likely due to the writing intervention.

Interestingly, the writing group, on average, had worse performance (roughly 3% lower) on the matching exercise (M = 137.69, SD = 14.45) than the control group (M = 143.07, SD = 34.83). Although not statistically significant, this finding was surprising for two reasons. Firstly, the writing group reported a higher initial knowledge of the target words. Secondly, the writing group had to perform a cognitive task involving the target words that the control group was not required to do. Therefore, we must conclude that hypothesis 1 is not supported.

TABLE 3 *Descriptive statistics and significance*

| | Control (n = 16) | | Writing (n = 16) | | | |
	M	SD	M	SD	T (30)	Sig. (2-tailed)
PESVES-0	1.25	5.00	.00	.000	1.00	.333
PESVES-1	50.13	17.05	41.56	20.21	1.296	.205
PESVES-2	36.31	13.88	30.69	8.36	1.388	.175
PESVES-3	21.06	9.71	22.63	9.28	-.465	.645
PESVES-4	50.00	19.65	65.13	24.51	-1.926	.064
Matching	143.0*	14.45*	137.7	34.83	-.555	.583
PVER -0	67.81	24.94	18.75	11.07	7.192	**.000**
PVER -1	1.19	1.91	.63	.89	1.071	.296
PVER -2	11.00	8.22	12.50	6.56	-.570	.573
PVER -3	23.81	12.04	35.69	10.74	-2.944	**.006**
PVER -4	56.19	17.37	93.63	17.44	-6.084	**.000**

*n = 15

The productive word knowledge, or the ability to use the words in complete sentences, was examined. This yielded several significant results. First, a significant difference in the level zero sentences was found between the control group (M = 67.81, SD = 24.94) and the writing group (M = 18.75, SD = 11.07), t (30) = 7.19, p = .000, d = 2.73. The control group refused to write many of the sentences, while the writing group was not so reluctant. Cohen's effect size value (d = 2.73) shows that the writing exercises had an extremely large impact on the ability and/or willingness to produce sentences incorporating the target vocabulary. Additionally, significant differences were found in level 3 sentences for the control group (M = 23.81, SD = 12.04) and the writing group (M = 35.69, SD = 10.74), t (30) = -2.94, p = .006, d = 1.04 and the level 4 sentences for the control group (M = 56.19, SD = 17.37) and writing group (M = 93.63, SD = 17.44) t (30) = -6.08, p = .000, d = -2.15. An analysis of the Cohen's effect size values (d = 1.04, and d = -2.15, respectively) shows that the intervention had a very large effect on the participants ability to produce a sentence which accurately used the target word. Based on these results, we conclude that hypothesis 2 (the writing group will be able to demonstrate a deeper understanding of the

vocabulary items by performing better on sentence writing exercises), was sup-
ported.

To test hypothesis 3, that the writing group would have greater overall vocab-
ulary gains, the results were examined by comparing the differences in the
PESVES and PVER scores. The results were taken from Table 3 and are com-
pared in Table 4 below.

The percentage change in the level zero scores is not reported because
only one student omitted a few of the answers in the PESVES, while many
students didn't write all of the sentences called for in the PVER. Showing a
percentage change from zero to any number is not meaningful. Comparing the
raw difference scores at level 0 is the best way to consider the differences of the
two groups at that level. The higher number of level zero words for the control
group show a higher inability or unwillingness to write sentences.

As for the level one words (completely unknown words), they have all but
disappeared in both groups (98% reduction). This is not surprising, consider-
ing the words had been studied for eight weeks. Both groups also show marked
decline in the number of level 2 words (-70% and -59% respectively). However,
in the level 3 and 4 words, the writing group is showing more gains than the con-
trol group. This indicates that the writing group has gained a deeper productive
knowledge of the words than the control group. Specifically, the writing group
is demonstrating a better knowledge of the semantic and syntactic features of
the target words. This provides further support for hypothesis 2 and contributes
to the support of hypothesis 3.

TABLE 4   *Comparison of group means*

	PESVES	PVER	Difference	Percent
Control level 0	1.25	67.81	66.56	N/A
Writing level 0	0	18.75	18.75	N/A
Control level 1	50.13	1.19	-48.94	-98
Writing level 1	41.56	0.63	-40	-98
Control level 2	36.31	11	-25.31	-70
Writing level 2	30.69	12.5	-18.19	-59
Control level 3	21.06	23.81	2.75	13
Writing level 3	22.63	35.69	13.06	58
Control level 4	50	56.19	6.19	12
Writing level 4	65.13	93.63	28.5	44

The breadth of learning, or number of words brought into the productive vocabulary, was also higher for the writing group. Looking at Table 2 one can see a 13.06 word increase at level 3 for the writing group compared to a 2.75 word increase for the control group. Additionally, the writing group is showing a 28.5 word increase at level 4 compared to a 6.19 word increase by the control group. The writing group made greater gains in the number of vocabulary words they could use. Because the writing group showed greater gains in both the depth and breadth of vocabulary learned, it must be considered that Hypothesis 3 is supported.

5 Discussion

5.1 *Major Findings*

It was found that there was a significant difference in the writing group's ability to write sentences that accurately use the target vocabulary (level 3 and 4) after the intervention. This indicates that training students to use vocabulary productively in sentences and short compositions will help them learn target vocabulary more effectively. The writing group showed gains in the number of words brought into productive vocabulary. The writing group was able to demonstrate semantic and syntactic knowledge as well as lexical knowledge during the PVER sentence writing task.

A related significant finding was that the students in the control group showed a greater reluctance to writing the sentences asked for in the PVER. If the high number of blank answers were due to fatigue alone, it would seem logical that the writing group would have a higher number of blank answers because they had additional homework requirements. However, this was not the case. It could be possible that the refusal to write was due to some kind of psychological resistance. Another reason the control group may have refused to write the sentences is that they simply could not perform the task. Whether it was because of fatigue, resistance, or lack of knowledge, the control group did not demonstrate any ability to use roughly 42% of the target words. The writing group refused to write sentences for only about 12% of the target words.

An interesting finding was that the treatment did not have a significant effect on the matching tests. The matching tests, according to Laufer, et al. (2004), could be termed as passive recognition, the lowest of four levels of vocabulary knowledge described in their article. As the matching task included both the target words and the definitions, it was considered that this test would show the receptive skill of the participants. Surprisingly, the writing group scored lower than the control group, although the writing group had worked more with the

words. However, this finding supports Barcroft's (2004) assertion that writing activities can have an inhibiting effect on the ability to recognize new words. In a subsequent study Barcroft (2006) wrote, "… requiring learners to write target words in sentences can decrease their ability to learn those words by depleting processing resources that otherwise could be used to encode target word forms and establish form-meaning connections." (p. 494).

An additional finding was the difference between the PESVES and PVER scores within and between the groups as shown in Table 2. This was most pronounced in level 4. By the end of the study, the control group students were able to actually produce an increase of just over 6 sentences at level 4, an increase of 12.38%. The writing group, on the other hand, realized an increase of almost 30 sentences, or 43.76%. The Pearson product-moment correlation coefficient (r = -.73) indicates that 73% of this difference between the two groups can be explained by the productive tasks during the intervention. This means that 73% of the 43.76% increase was due to the intervention, while 27% of the increase was due to individual factors. This finding must be viewed cautiously due to the tendency for individuals to exaggerate about known items in self-reported measures such as the PESVES.

Assigning learners productive tasks in the form of sentence writing and original text writing using the words taught in reading lessons seems to be an effective tool to enhance vocabulary knowledge. It also helps them to learn how to use this knowledge in their language. The negative effect of writing sentences on vocabulary learning reported by Barcroft (2004) is related to new words rather than known words. Apparently, making the words known to students through teaching vocabulary in a reading context prior to productive vocabulary activities is beneficial.

The two complementary measurement tools, the PESVES and the PVER, utilized in this study were found to be useful. The advantage of PESVES/PVER over the VKS is that it can be used with a relatively large amount of vocabulary. In the current study, 160 vocabulary words were targeted for learning by 32 students. The VKS accepts self-reporting only at levels 1 and 2, but requires students to demonstrate their vocabulary knowledge at the higher levels. The thousands of sentences that would need to be evaluated using the VKS in this case would take a great deal of time and energy. Additionally, if one is interested in measuring the impact of a course of study on students' vocabulary knowledge, one may very well want to control for the practice effect. The PESVES was found to be a convenient method to obtain data that could be analyzed in a meaningful way.

The PESVES and PVER have been designed to be complementary measurement devices. They are designed to be easy to use both by the student and the

teacher. Admittedly, it took a lot of time to analyze the sentences at the end of the experiment. However, the combination of the two measures provides some very useful information.

One drawback of the PESVES/PVER combination is the lack of precision. A similar criticism has been made about the VKS. The numerical levels do not give an adequate reflection of partial word knowledge. Apparently, the perfect method for assessing vocabulary learning has yet to be developed. Recent improvements of computer assisted methods such as computer adaptive tests or the Coh-Metrix system are promising (Laufer & Goldstein, 2004; Graesser, McNamara, Louwerse, & Cai, 2004). However, for this study, it was found that the PESVES/PVER combination was relatively easy to administer, controls for the practice effect, and provides results that can be examined categorically. Therefore, we recommend other teachers and researchers to try utilizing this combination of measurement devices in their work.

5.2 *Relation of Findings to Other Studies*

The results of the current study are in line with an enduring and popular theory of memory. With their Levels of Processing theory, Craik and Lockhart (1972) were responding to research on memory that was conducted mostly in the late sixties. These studies from the sixties focused on methods of memorization such as presentation rates, scaled properties of stimuli, serial position, or the form of the memory test (Lockhart & Craik, 1990). Craik and Lockhart postulated that memorization is not the result of a special encoding process, but the byproduct of normal cognition. In other words, the ability to remember a certain item is related to the way the individual cognitively processes that item when encountered. Although many cognitive functions have been identified, Craik and Lockhart argued that no specific "committing to memory" function exists. To get a subject to remember a word, Lockhart and Craik make the assertion that it is important to control the "orienting task". This will control the "depth of processing" (a term which Craik and Lockhart admit is ambiguous) and the quality of recall or recognition.

To quote directly from Craik (2002) "Using orienting tasks that induced the participants to process words in a deep semantic fashion, it was easy to demonstrate that incidental encoding can yield levels of memory performance that are at least as good as those obtained after intentional learning." (p. 310). Later, Craik makes a distinction between "remembering" and "knowing" saying that knowing may be due to higher levels of processing. "One possibility is that representations at higher levels are more interconnected and networked, thereby providing more access routes for retrieval processes; another is that the general knowledge represented by higher levels is used to interpret new events or plan

new actions, and is therefore accessed more frequently than is specific event information represented by lower levels (pg. 313)." This distinction between knowing and remembering is important for language teachers who want to provide vocabulary instruction. In the majority of cases, language teachers would rather have the students know a word than just remember it. To achieve that aim, target vocabulary must be processed in a deep semantic fashion. The productive action of writing requires such deep semantic processing.

This may also help to explain why, in the current study, the groups performed similarly (3% difference) on the matching test and dissimilarly on the PVER sentence writing activity (roughly 400% difference). The control group simply had no tasks which oriented them towards productive use of the vocabulary. Additionally, they did not have the deep semantic processing activity of writing. Both groups, however, were trained to remember the target words. The contention here is that knowing a word is not a simple matter of remembering the form-meaning connection. Furthermore, teaching which concentrates on the form-meaning connection without any tasks which orient the student toward productive use are of limited value to students. After graduating, when former students find themselves in the work environment, they will seldom have to respond to a vocabulary matching test. However, it is quite likely that they will have to construct sentences that incorporate new vocabulary items. Therefore, for the benefit of the students, it behooves teachers to employ productive tasks such as writing when teaching vocabulary.

5.3   *Comprehensible Input vs. the Output Hypothesis*

One of the issues in this study concerns the nature of vocabulary teaching. There has long been a debate over the relative merits of explicit instruction and implicit acquisition. Krashen is probably the most well-known proponent of implicit acquisition. In the mid-1980s, he developed a theory that "comprehensible input" is sufficient for second language acquisition (Krashen, 1985). Swain (1985, 1995) disagreed. According to Swain, output has the key role in second language acquisition. Swain noticed that immersion students failed to produce language like a native-speaker in spite of considerable comprehensible input. This led her to think that mere exposure to input was not enough to trigger production. Swain and Lapkin (1995) argue that it is when learners are producing language that they become aware of problems and seek ways to produce better output.

Krashen (1998) responded by saying that "comprehensible output" is a rare phenomenon and therefore can't be a major contributor to second language acquisition. Swain contends that "Comprehensible Output" can be a misleading term (Swain, 2007). When she was formulating her hypothesis, "output" was

considered a verb which described the process of production, not the product itself. She did not mean that the output must be understandable, just that the cognitive effort involved in the encoding process can be helpful at times for acquiring a new language. However, she did indicate that the output should be "pushed", or encouraged by the teacher. Swain states that the claim of the output hypothesis is that, under certain circumstances, productive activities facilitate second language learning. Although the debate continues, most practitioners use some combination of instruction and practice.

Rosenshine (2008, 2012) contends that effective teaching is strategic in nature. There is no reason to suppose that vocabulary teaching in a second language is an exception. A strategic approach to vocabulary learning may be beneficial to students' education. A strategic series of vocabulary learning activities done in a particular order may prove beneficial to learners. That order is receptive activities starting with direct vocabulary instruction including explicit lexical explanations, followed by contextual analysis using reading or concordances, and then further elaboration through productive activities such as discussions and writing.

Having considered the debate on the relative values of input and output, it was decided to use both of these approaches in the current study in a strategic manner, starting with input and moving to output. This study supports the view that newly learned words in context through reading lessons are cognitively processed, thus enhancing the awareness of word meaning. The possibility of putting this knowledge into use is enhanced by further cognitive processing in productive activities such as writing. In terms of vocabulary, this meant moving from lexical awareness to semantic elaboration, by moving from listening and reading to discussion and writing.

We feel that similar strategic development of activities can have a positive impact on second language learning in general. A new term for this combination of input and output might be the "Throughput Hypothesis". Briefly stated, the Throughput Hypothesis claims that language teaching is facilitated by strategically designed educational practices that start with receptive activities and move towards productive activities. We are suggesting that this orientation can assist in all language teaching and learning, even the development of receptive skills. For example, reading can be further developed with the aid of productive activities such as comprehension questions, summaries, response essays, and critical analysis discussions. Also, listening skills can be increased with a variety of productive exercises, including note-taking and cloze activities while listening to songs.

Input and output cannot be considered in terms of relative importance without a context. In other words, both input and output are crucial to language

learning, only the focus changes. If the goal is to develop receptive skills, more emphasis may be given to reading and listening. However, if the goal is the development of productive skills, then speaking and writing may be stressed. In either case, learning will be facilitated by the strategic inclusion of a non-focal activity. To state it clearly, every receptive activity should have a corresponding productive activity following it, and every productive activity should be preceded by a receptive activity. By using this strategic approach, second language acquisition will be both more efficient and more effective.

One common goal of learning another language is to speak it with near-native proficiency, or at least to be able to communicate in a real setting. Therefore, knowledge of rules and the ability to recognize words are not enough for many students. Language education should not solely be focused on developing the skill to pass a test. The Throughput Hypothesis suggests developing a production-oriented teaching mode, including writing activities, which will much better serve our students and prepare them for the challenges they will face in modern life. It is important that students start trying to learn vocabulary and stop trying to remember words.

5.4   *Limitations*

As in all studies, this research is not without limitations. First, in an eight-week period 160 words were extensively taught and tested, which could lead to boredom and fatigue in the participating students. Many of the students did not write sentences in the final activity. Many potential reasons for this exist. The main potential reasons for refusal to write the sentences were identified as fatigue, psychological resistance, and inability.

Total isolation of one linguistic feature in assessment measures is unlikely. In the current study, in addition to assessing vocabulary, the examinees were assessed on their ability to write sentences. The sentence writing ability was not tested prior to the start of the study. Perhaps some of the lexical items were known, but the participant may have lacked the writing ability to express the lexical knowledge.

The self-reported initial measure could easily be said to be a shortcoming. However, research has shown that self-reports are in many cases both valid and reliable. The PESVES and the PVER measured the same construct using different methods. These rubrics were careful designed to collect reliable data. However, it must be admitted that the self-reported measure at the outset of the study will be considered by many to be a serious threat to the validity of this study.

The intervention itself was not strictly controlled. The writing assignments were given as homework. The amount of time spent and exact procedures used

by the students were not monitored. Furthermore, the writing assignments were not assessed nor feedback given. Although this is a limitation in terms of the validity of the study, it has implications for teaching. The gains reported by the writing group required only about ten minutes of class time per day.

Also, the Bonferroni correction was not utilized. Although many variables are included in the study, only two groups are being compared. At the time of the study, it was thought that a Bonferroni correction could only be conducted with three or more groups. Therefore, a Bonferroni correction was not applied. In this case, the comparisons which are significant are highly significant ($p <$ .005 in two cases and $p = .006$ in the other). Therefore, likelihood of a type I error in this case is extremely low and it is unlikely that the Bonferroni correction would have substantially altered the interpretation of the data.

Finally, the sample size of 32 students is on the small side. A larger sample size will reveal more generalizable results. This study was also conducted only at one university preparatory program and other programs may provide for different results.

More research is needed in the area of strategic vocabulary learning. Particularly, the impact of the cognitive processes involved in writing on vocabulary acquisition should be more thoroughly explored.