The Roles of Output on L2 Vocabulary Acquisition: Noticing, Retrieval and Retention

Sun-Hee Kwon
(Pusan University of Foreign Studies)


A number of studies have recently proposed that input alone does not adequately promote both receptive and productive aspects of L2 vocabulary, and output is as necessary a condition as input for full development of L2 learners’ lexical proficiency. The goal of the present study is to address the question of what specific functions of output may affect L2 vocabulary development. This study considers that output could positively affect L2 learners’ vocabulary acquisition in terms of three aspects: a) noticing, b) retrieval, c) retention. The experiment was conducted with sixty-three learners of English in South Korea over the course of eight weeks. The results of ANOVA analyses of vocabulary and written production tests indicated that output plus input treatments affected more positively participants’ noticing, retrieval and retention of target vocabulary than the input only treatments did. The findings of this study suggest the importance of output in developing L2 vocabulary knowledge, as well as argue for the need to consider effective classroom instruction involving both input and output for receptive and productive vocabulary knowledge.

I. INTRODUCTION

The roles of input and output have been at the fore of second language acquisition (SLA) theory for the last twenty years, and the body of research has yielded important insights for the field of second language (L2) vocabulary acquisition. Previous vocabulary acquisition research has built an argument for productive vocabulary as a natural outcome after receptive vocabulary acquisition (Day, Omura, & Hiramatsu, 1991; Krashen, 1989; Nagy & Anderson, 1984; Nagy, Herman, & Anderson, 1985). Such studies have mainly focused on the importance of relevant input in promoting L2 learners’ vocabulary knowledge, rather than on the importance of output. Recently, however, a number of
studies have proposed that output is as necessary a condition as input for acquisition of higher levels of both receptive and productive vocabulary knowledge (de la Fuente, 2002; Ellis & He, 1999; Hulstijn & Laufer, 2001; Lee, 2003). Central to the theoretical argument of these studies is the Output Hypothesis (Swain, 1995, 1998), which views language production as the trigger of attention to language forms and meanings.

According to Swain, learners acquire a L2 when they are forced to produce output that contains linguistic forms in their interlanguage system. Swain argues that producing language induces learners to test hypotheses about comprehension of input or about linguistic correctness. Thus, language production forces learners to process language forms and meanings more deeply than if they were simply exposed to input. Therefore, in Swain’s view, input is a necessary but insufficient condition for full development of L2 proficiency. Stressing the importance of both input and output for promoting language acquisition, Swain argues that output is not only the result of but also the cause of L2 acquisition.

The recent proliferation of literature on productive lexical proficiency is largely a result of interest in the role of output in L2 acquisition. Studies by Laufer (1998) and Lee (2003) provide empirical evidence that L2 learners’ productive vocabulary knowledge does not develop in the same way as receptive vocabulary knowledge does; when receptive vocabulary knowledge increases, only a very small portion of receptive vocabulary becomes productive without instruction on productive vocabulary use. Studies by de la Fuente (2002), Ellis and He (1999), and Hulstijn and Laufer (2001) provide evidence that output has been found to be the critical condition that promotes both receptive and productive vocabulary knowledge.

The results from the studies above lead to two important conclusions: 1) the development from receptive to productive vocabulary does not occur naturally in L2 acquisition, and thus, different instructional conditions are needed in order to convert receptive into productive vocabulary, and 2) effective instruction for productive vocabulary acquisition necessarily involve an output task condition. These conclusions, then, raise a question: What specific functions of output could affect L2 vocabulary development? While previous studies have shown that output plays an important role in L2 vocabulary acquisition, few studies have addressed in which aspects L2 learners benefit from production of output. In addressing this issue, the current study argues that output may positively affect L2 learners’ vocabulary acquisition in terms of three areas: a) noticing, b) retrieval, c) retention.

First, the question of how noticing interacts with output in L2 vocabulary acquisition has hardly been discussed in the previous studies, regardless of evidence that L2 learners have primarily shown their noticing benefits in their lexical domain—rather than in the grammatical or semantic domain—acquisition as a result of production (e.g., Erlam, 2003;
Izumi & Bigelow, 2000; Van den Branden, 1997). Thus, the current study hypothesizes that output provides an opportunity to promote L2 learners’ noticing on their lexical knowledge, either on their deficiency of vocabulary or, on the gap between their receptive and productive vocabulary knowledge. Second, previous studies have shown that the lemma (meaning) and lexeme (form) connection causes the principal problem for retrieval of L2 lexical knowledge (Wesche, Paribakht & Ready, 1994). This study supports the argument that output may serve as a crucial means for strengthening lemma-lexeme connections, by providing evidence that output is not a supplementary, but rather the necessary condition for gaining easy access to and efficient control over vocabulary knowledge. Third, previous studies have yielded mixed results on whether output production provides learners a better chance to retain lexical information than input alone does (Benati, 2005; Horibe, 2003; Kim, 2001). The rationale underlying the effectiveness of output is that production induces high cognitive involvement, and is more effective in leaving deep imprints in memory than simple exposure to target words in the input. The current study therefore focuses on whether output affords learners a greater chance to retain lexical knowledge that input alone does.

II. LITERATURE REVIEW

1. Previous Studies of Output and L2 Vocabulary Acquisition

A number of recent studies have addressed the productive aspects of L2 vocabulary knowledge. Focusing on the functions of input and output, these studies attempt to explain how input and output affect vocabulary acquisition. One of the studies to investigate L2 productive vocabulary acquisition was a study conducted by Ellis & He (1999). This study was the first empirical investigation to document the benefits of output on acquisition of receptive and productive vocabulary in L2. In order to compare the effects of input and output treatments on oral acquisition of L2 target words, three experimental groups were compared: a) the pre-modified input group, which received the input treatment without interaction, b) the modified input group, which received the input treatment with interaction, c) the modified output group, which received the output and input treatments with interaction. The results indicated that there were no differences between the two input groups and that the output with interaction group performed significantly better in the acquisition of target items than the other groups. Ellis and He interpreted these results as suggesting that the interaction between output and dialogic interaction could be a beneficial factor for learners to acquire productive as well as receptive vocabulary knowledge.
Similarly, de la Fuente (2002) conducted an experiment to examine different roles of negotiation prompted by output on the receptive and productive acquisition of words. The study yielded similar findings to those of Ellis and He (1999): only negotiated interaction that incorporated output appeared to have promoted both receptive and productive acquisition of words, as well as an increase in productive word retention.

While these two studies were concerned with whether the output treatment was more advantageous for target word acquisition as opposed to the input treatment, a study by Lee (2003) examined the effects of different instructional techniques on improving productive vocabulary use. Lee investigated productive vocabulary use in the writing of secondary ESL learners in Canada and found that vocabulary instruction, by means of various learning strategies including reading, writing, and comprehension of target vocabulary and target language learning within grammar exercises, significantly increased the productivity of the target vocabulary. Lee proposed that systematic vocabulary instruction could help to convert receptive vocabulary into productive vocabulary.

An experimental study by Van Gelderen, Snellings and De Glopper (2004) was also concerned with L2 learner’s productive lexical knowledge and its relationship to writing. In their research with Dutch secondary school students learning English as an L2, participants showed a significant enhancement in their speed of lexical retrieval as a result of training. Similar results were also found by Schoonen and Verhallen (1998), who found that lexical retrieval and sentence-building training was correlated with participants’ increased use of target words in their writing. This result supports the claim that production practice plays an influential role in enhancing learners’ productive use of L2 vocabulary.

While the findings from the above studies provide evidence in favor of a role for output in productive vocabulary acquisition, they have several limitations: First, the studies by Ellis and He (1999) and de la Fuente (2002) were concerned with the function of output within the “negotiation” process. In other words, these studies’ output tasks did not involve the learners in production alone, since the negotiation variable (e.g., learners’ talking in pairs, or talking with a native speaker) was combined with the output variable. Although positive effects were found for the output groups in both studies, a main factor for these effects was not the output variable alone, but the combination of the negotiation and output variables. Thus, it is necessary to isolate the output variable in order to avoid confounding effects of the treatment, in addition to examining the functions of output with greater precision.

A second limitation is that there is little discussion of which specific functions of output may play a role in productive vocabulary acquisition. While the functions of output proposed by Swain have been investigated in connection to other language aspects, (for instance, the noticing function in relation to grammar), little discussion exists regarding which specific functions of output may affect productive vocabulary acquisition. As
McLaughlin and Heredia (1996) argued, production in output tasks does not mean solely mechanical training or practice. Instead, it includes complex processes that may prompt the internal cognitive processing necessary for acquisition of the language. Thus, exploring specific functions of production in output tasks relative to L2 vocabulary acquisition could have important implications for effective productive vocabulary instruction in L2 classrooms.

Regarding the potential functions of output, this study argues that output could positively affect L2 productive vocabulary acquisition in terms of a) noticing, b) retrieval, and c) retention. The theoretical and research backgrounds underlying these assumptions are discussed in the next.

2. Roles of Output on L2 Vocabulary Acquisition

1) Noticing

The question of how output enables L2 learners to acquire vocabulary more effectively than input or other conditions do might be answered with the concept typically called noticing or attention. This term derives from cognitive psychology, and the claim that no learning can occur without attention has been a focal point of recent SLA research. Numerous studies have investigated how attention can aid the L2 learning process, and have proposed that attention is a key for enabling L2 learners to analyze and incorporate linguistic forms and meanings into their L2 linguistic system. One of the most recognized proposals explaining the role of attention in L2 learning is the noticing hypothesis proposed by Schmidt. In a series of articles (1990, 1992, 1995), Schmidt formulated the hypothesis to illustrate that conscious attention is crucial for converting input to intake, which is then comprehended and passed into the internal processor for further analysis. The concept of comprehended input, or intake has also been incorporated into the Input Processing model devised by VanPatten and colleagues (VanPatten & Cadierno, 1993; VanPatten & Sanz, 1995; VanPatten & Oikkeneon, 1996). VanPatten argued that only internalized and comprehended input can be incorporated into the learners’ interlanguage system.

However, the theoretical tenets of the noticing hypothesis and Input Processing model were originally limited to the interaction of attention with input as a way to incorporate new features into the L2 linguistic system. The issue of the interaction of attention with output had been neglected until Swain’s output hypothesis. Swain (1995) pointed out that there are different types of noticing based on what is noticed, such as 1) noticing a form in the input 2) noticing one’s interlanguage deficiencies or holes and 3) noticing the gap between the interlanguage and the target language. In terms of output, the second and third
types of noticing are related. Especially for productive vocabulary, the second type of noticing—notice the gap—is closely related to L2 learners’ reporting deficiencies in producing sufficient vocabulary when engaging in speaking or writing, in spite of their relatively successful achievements in comprehension (Leki & Carson, 1997; Raimes, 1985; Saville-Troike, 1984; Silva, 1992). These problems may be caused either by L2 learners’ simply not knowing what to say in order to express their intended meaning, or by having difficulty in activating the words that they already know (Henriksen, 1999). In either way, learners are less likely to notice their lack of target vocabulary until they are forced to produce it, and it is in this sense that output plays a crucial role in bringing learners’ attention to target vocabulary and in motivating them to reduce the gap between their receptive and productive vocabulary knowledge.

The question of how noticing interacts with output in L2 vocabulary acquisition has hardly been mentioned in the literature, despite evidence that learners’ noticing during production occurs most often in the lexical domain. For example, Erlam’s study (2003) found that vocabulary was the largest factor in explaining learners’ noticing during production, and a study by Gass and Torres (2005) exploring the effects of interaction on L2 acquisition, indicated that learners benefited most from a lexical production activity during interaction. Other studies by Van den Branden (1997), and Izumi and Bigelow (2000) also have made similar findings, but these studies did not stress the fact that learners’ noticing occurred most frequently in the lexical domain. In short, these studies support the argument that noticing may assist L2 learners to perceive their lack of vocabulary, as well as raise the need for more detailed exploration of the potential effects of output instruction on L2 lexical knowledge development.

2) Retrieval

One of the most promising sources of evidence to support the functions of output in relation to lexical knowledge is the body of research investigating lexical retrieval. For example, a study by de Bot (1996) has evaluated the role of output at the lexical level. Based on Levelt’s production model, she argues that the main function of output is enhancement of lexical fluency, that is, increasing the speed of lexical retrieval. de Bot’s basic logic underlying this claim is that the retrieval of a word is a two-step process, and the connection between these two processes is strengthened by output, for several reasons. First of all, de Bot argues that the lemma (meaning) and lexeme (form) are not activated simultaneously; in fact the lemma must be activated first, and then the lexeme. That is, L2 learners access lemmas first, which contain semantic and syntactic information about lexical items, and then try to retrieve phonological and morphological information for the selected lemma. Second, de Bot claims that the strength of the connection between the
lemma and the lexeme determines how quickly the transition from declarative to procedural vocabulary knowledge takes place. Additionally, the connection between lemma and lexeme can be strengthened more effectively and efficiently via production on one’s own rather than by input from external sources. Production, she argues, gives the learner a greater chance to notice any mismatch between lemma and lexeme than does input, since production requires a greater degree of focused attention in order to coordinate semantic and articulatory-level processing. Thus, although de Bot does not refute the roles of various other factors (e.g., the degree of linguistic difference between L1 and L2, the quality and quantity of input) on developing L2 vocabulary procedural knowledge, and also states that output does not play a major role in the acquisition of completely new declarative knowledge, she argues that output is the crucial condition for the conversion of declarative to procedural L2 vocabulary knowledge.

Another study by Wesche, Paribakht and Ready (1994) has also focused on the lemma and lexeme connection in L2 vocabulary learning. They collected introspective data from adult ESL students from various backgrounds. In their study, students had to read a passage in English and then summarize it. They also had to indicate what lexical problems they had encountered in the passage. Some students had difficulty finding the right word meanings they had practiced before and later produced the right meaning by repeating the word form. This example suggests that somehow, the connection between the L2 lemma and lexeme is less stable and that lexical information in the L2 lexicon is both less available and less accessible to learners.

An experimental study by Snellings, Van Gelderen, and De Glooper (2002) investigated whether L2 learner’s lexical retrieval could be improved by classroom instruction. The study showed that Dutch secondary school students learning English as an L2 showed significantly enhanced speed (i.e., faster reaction time to a lexical decision task, and used a greater amount of the target words in the translation task) of lexical retrieval in the posttest after four weeks’ of classroom instruction. The authors argued that such enhanced lexical retrieval was transferred to narrative writing ability since students in the experimental groups used the trained words more often in their narrative texts and showed significant improvement in their content expression.

In summary, the evidence accumulated from the previous studies supports the theoretical claim that output may serve as a crucial means to strengthen connections between the lemma and the lexeme. Such strengthened connections between the lemma and the lexeme enable learners to have easy access to and efficient control of vocabulary knowledge stored in their L2 language system, and can also contribute to an increase in lexical fluency (retrieval). In other words, when L2 learners are engaged in output production, they are required to actively solve problems of word form or of word meaning on their own. Through such active processing of lexical information, a learner can achieve faster, more
precise and automatic use of vocabulary knowledge than when just hearing or reading L2 vocabulary.

3) Retention

The question of which factors affect the retention of L2 vocabulary is of great interest to researchers. The general consensus from the field of cognitive psychology is that retention of new information depends on the amount and the quality of attention that individuals pay to various aspects of words. That is, having more qualitative and quantitative associations with existing knowledge increases the chances that the new information will be retained. This so-called “elaboration” process is considered a key concept to explain how output production can enable learners to retain lexical information.

Arguably, the output condition induces more of the elaboration process from learners than the input condition does, due to the linguistic complexity (i.e., semantic, syntactic, and phonological processing) involved in producing output. Additionally, the operation of the elaboration process leads to better retention of vocabulary, which is supported by empirical research conducted by Hulstijn & Laufer (2001). The authors collected data from EFL learners in the Netherlands and Israel to examine the effects of three different tasks on retention of vocabulary. The different tasks consisted of a reading comprehension task with marginal glosses, a reading comprehension plus fill-in-the-blank task, and a composition task using the target words. The results of the posttest showed that retention of ten unfamiliar target words was highest for the composition task, lower in the reading plus fill-in-the-blank task and lowest in the reading task. Due to these findings, Hulstijn and Laufer argue that the superior performance on the composition task may provide support for the output hypothesis, given that the production (composition) task required the learners to stretch their linguistic resources. Additionally, a number of studies have also added evidence that output positively affected the retention of L2 vocabulary. Studies by Ellis & He (1999), Hulstijn & Trompetter (1998), Joe (1998), and Nobuyasi & Ellis (1993) have indicated that vocabulary used in production tasks was recalled better than words practiced in non-production tasks. In explaining the superiority of output tasks over others for vocabulary recall, these authors suggest that output tasks may require a deeper level of processing of new words than other kinds of tasks.

In spite of the above studies supporting the role of output in vocabulary retention, a number of studies have questioned the effect of output on vocabulary retention. For instance, Dekeyser and Sokalski (1996) investigated L2 learners’ acquisition of Spanish object-clitic pronouns, and argued that while the effects of output practice tended to fade over time, those of input practice were more durable. Horibe (2003) and Sakai (2004) have also argued that they did not find evidence to support the output condition having
promoted long-term retention of specific linguistic features better than the input condition. Although these studies did not focus on vocabulary acquisition, the results from these studies require further validation of output’s effect on L2 retention.

III. METHODOLOGY

1. Research Question

Based on the theoretical and empirical research discussed in the previous sections, research hypotheses were formulated to guide this study.

**Hypothesis 1**: L2 learners exposed to input plus output vocabulary instruction on target words will gain higher levels of noticing of target words than learners exposed to input only vocabulary instruction.

**Hypothesis 2**: L2 learners exposed to input plus output vocabulary instruction on target words will gain higher levels of retrieval of target words than learners exposed to input only vocabulary instruction.

**Hypothesis 3**: L2 learners exposed to input plus output vocabulary instruction on target words will gain higher levels of retention of target words than learners exposed to input only vocabulary instruction.

2. Group Design

The group design, having a total of sixty-three (N=63) EFL students enrolled in the second grade at a high school in Korea, was conducted as follows: First, fourteen students were assigned to 1) the control group (N=14), and sixteen students to 2) the input group (N=16). The control group received no vocabulary treatment, while the Input group received vocabulary treatments involving input activities. Next, thirty-three students were assigned to the output group, receiving vocabulary treatments involving output as well as input (N=33). Last, thirty-three students in the output group were randomly assigned to one of two subgroups: 1) a group completing both the input and output tasks in the sequence of input-output (IO, N=16), 2) a group completing both the input and output tasks in the sequence of output-input (OI, N=17). The reason to split the output group in terms of the task sequence is to see if there is any difference when the output tasks are presented before the input tasks, in comparison with the input first and the output presentation. All students were randomly assigned in each of four groups, and they had been taught with the same curriculum in English by the same instructor for the school year in which the study was
conducted. To ensure that there were no significant differences in general English proficiency between groups prior to the experiment, their final term exam scores, which was conducted one month before the start of this study, were compared. It was confirmed that the average scores in each group were similar. \( F (2, 63) = 0.083, p = .92 \).

### 3. Experiment Procedure and Treatment Tasks

The study employed a pretest-posttest design carried out over a period of two months. All participants in the experimental groups completed the following: pretest-treatment-two posttests. Also, ten participants selected from the input and output groups participated in an informal interview. The control group completed the same tests, except for receiving the treatment. All of the treatment was conducted in the written mode, and consisted of six sessions.

The various tasks presented to the experimental groups all involved the same topic and the same target vocabulary, but varied depending on the different treatment conditions. The subjects were given a series of tasks related to pictures showing an accident occurring on a road. The tasks for each group consisted of a variety of activities for describing the pictures, such as a word-recognition activity, comprehension questions, a short-answer activity, and a narrative writing activity. The subjects were requested to complete each given task using their out-of-class study periods, which were normally given in a 50-minute block after regular class hours. The instructor assisted them in carrying out the tasks, but did not provide direct coaching. All of these tasks required the learners to primarily attend to the meanings of the target words, but the tasks were also aimed to draw learners’ attention to target vocabulary by asking them to solve the given problems or questions, which necessitated use of the target words. Thus, the tasks were essentially meaning-based activities, which provide both the information and the incentive for learners to solve the linguistic problems in the tasks. The procedure and the task types in each experimental group are described below in Figure 1 and Table 1 and 2.

<table>
<thead>
<tr>
<th>Tasks Used in the Input Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type of the task</td>
</tr>
<tr>
<td>-----------------</td>
</tr>
<tr>
<td>Input 1</td>
</tr>
<tr>
<td>Input 2</td>
</tr>
<tr>
<td>Input 3</td>
</tr>
<tr>
<td>Input 4</td>
</tr>
<tr>
<td>Input 5</td>
</tr>
<tr>
<td>Input 6</td>
</tr>
</tbody>
</table>
### TABLE 2
**Tasks Used in the Output Group**

<table>
<thead>
<tr>
<th>Type of the task</th>
<th>Activity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Input 1</td>
<td>passage reading 1 + comprehension question 4</td>
</tr>
<tr>
<td>Input 2</td>
<td>short answer (multiple choice)</td>
</tr>
<tr>
<td>Input 3</td>
<td>passage reading 2 + comprehension question 4</td>
</tr>
<tr>
<td>Output 1</td>
<td>word matching (fill-in-the-blank) (Appendix D)</td>
</tr>
<tr>
<td>Output 2</td>
<td>short answer writing (Appendix E)</td>
</tr>
<tr>
<td>Output 3</td>
<td>narrative writing</td>
</tr>
</tbody>
</table>

### FIGURE 1
**Treatment Procedure**

![Treatment Procedure Diagram](image)

4. **Target Vocabulary**

The target vocabulary was selected from items on the pretest. To avoid participants’ potential exposure to the target vocabulary prior to the experimental period, the least known 30 vocabulary items in the pretest were selected. The target items were all related to descriptions of a six-frame picture, and the selected target items consisted of 4 nouns, 22 verbs, and 4 adjectives, a total of 30 vocabulary items.

**Target vocabulary (30 items)**

*frightened, brave, reward, break down, punish, ride, hurt, scratch, pick up, pull over, worry, fix, narrow, continue, balance, steer, consider, complain,*
wonder, knock down, lose, regret*, lane, approach, hedge, downhill, honk, consequence, intimidate, horn

The target words inside [ * ] in the list above belong to the General Service List (West, 1953) which consists of a list of the 2,000 words considered most useful to learners of English as foreign language, and the other items are listed either in the University Word List (Nation & Xue, 1984) or in the Academic Word List (Coxhead, 2000), which contain words beyond the 2000 most frequent ones.

5. Testing Instruments

1) Vocabulary Tests

In this study, learners’ vocabulary knowledge was measured using a format adopted from the Vocabulary Knowledge Scale (VKS). The VKS, developed by Wesche and Paribakht (1996) for vocabulary assessment of ESL learners, uses a scale combining self-report and demonstrated knowledge: students are asked to indicate their degree of knowledge for each target word, as well as demonstrate the specific use of the target word in a written form. The original scale ratings use five scoring categories, ranging from complete unfamiliarity (a score of 1), through recognition of the words and some idea of their meaning (a score of 2 3 4), to the ability to use the word with grammatical and semantic accuracy in a sentence (a score of 5). The receptive-productive dimensions of vocabulary knowledge are presumably involved in these five distinct self-report categories, and since the current study aimed to assess gains of receptive and productive vocabulary knowledge in relation to different kinds of instruction, it should therefore take advantage of the VKS scale. In the current study, however, the format was modified from a five-scale to a four-scale category as follows:

TABLE 3

Modified Vocabulary Test Format

<table>
<thead>
<tr>
<th>Please choose the category that best describes your understanding of the following word</th>
</tr>
</thead>
<tbody>
<tr>
<td>FRIGHTENED</td>
</tr>
<tr>
<td>A. I don’t remember having seen this word before.</td>
</tr>
<tr>
<td>B. I have seen this word before, but I don’t know what it means.</td>
</tr>
<tr>
<td>C. I know this word. It means ___________ (synonym, or translation)</td>
</tr>
<tr>
<td>D. I can use this word in a sentence:</td>
</tr>
<tr>
<td>(Student’s sample response) D: I frightened darkness</td>
</tr>
</tbody>
</table>

All participants took the tests in class for 30 minutes. No dictionaries, translations or teacher help were allowed, and no feedback was offered. The 40 items in the two posttests were the same as those used in the pretest, but the order of the items was randomly changed.
2) Written Production Tests

The main aim of the written production tests was to examine whether there were any differences between groups regarding participants’ productive use of target words after the treatments. All the subjects performed two post-treatment written production tests for 30 minutes in their regular English class. The written production test was not given before the vocabulary treatment since the writing test itself might function as the output condition potentially blurring the investigation of the effects of the different tasks in the current study. Thus, the participants were given only two written tests after the treatment. During the test, they were provided with a worksheet that contained a picture related to the writing topic, and were instructed to support their ideas when writing by including descriptions of the pictures as much as possible. There was no limit on the length of their writing, and upon completion, participants’ writings were collected by the instructor. No dictionaries, translations or teacher help were allowed during the test, and no feedback was offered after the test.

6. Scoring

In the pre- and post-vocabulary tests, answered items were scored from 1–4 points (Table 4). Thus the maximum score for the vocabulary tests was 160 points. In the two writing tests, the total number of target words used was counted. For both the vocabulary and written production tests, when words other than the target words produced in the sentence had spelling errors, they did not affect the scoring. However, a target vocabulary item was scored as correct only when correct in both meaning and spelling. Target items that were used correctly more than once were counted only once. Grammatical errors were not considered unless they seriously hindered understanding the meaning of the target vocabulary.

<table>
<thead>
<tr>
<th>TABLE 4</th>
<th>Scores in the Vocabulary Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Categories</td>
<td>Scores</td>
</tr>
<tr>
<td>A</td>
<td>1</td>
</tr>
<tr>
<td>B</td>
<td>2</td>
</tr>
<tr>
<td>C</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>3</td>
</tr>
<tr>
<td>D</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>4</td>
</tr>
</tbody>
</table>
7. Informal Interview

The main aim of the informal interview was to explore how output affected learners’ target vocabulary learning, and in which learning processes the participants were engaged during the output activities. The interview was conducted after the delayed posttests were completed, and two students from the input and eight students from output group volunteered for the researcher’s interview. In order to provide a chance for the learners to go over and think about the questions before the interview, they were distributed prior to the actual interview. Questions were about the tasks in terms of target vocabulary learning, such as the difficulty of each task, and the tasks they found interesting/useful (Appendix F). The participants were allowed to take notes, and bring them to the interview. Thus, the interview followed a semi-structured format. The interview was conducted individually in Korean and recorded by the researcher in a written form. No attempt was made to analyze quantitatively the data from the interview.

IV. RESULTS

1. Vocabulary Tests

To investigate the effects of the output condition on participants’ target word acquisition, two sets of analyses were conducted; 1) comparison of the control, input, and the output groups (OI+IO), and 2) comparison of the Input-Output (IO) and Output-Input (OI) groups.

The second set of analyses is to see if there is any difference between the different task presentations, Input-Output and Output-input. The first set of analyses was conducted with the raw scores of the control, input, and the output group to investigate the effects of the output condition. Two independent variables and one dependent variable were involved in this analysis. The independent variables were the treatment group having three levels (control, input, output group) and time having three levels (pretest, posttest 1, posttest 2). The dependent variable was the score of the vocabulary test. The descriptive measures indicate that while there was no difference between the three groups at the start of the study, the mean scores of two experimental groups were higher than those of the control group on both the immediate and delayed posttests (Table 5). The results of a repeated measures ANOVA (Table 6) confirm this finding. A repeated measures ANOVA with between-subject (Group) and within-subject (Time) variables (3x3 ANOVA design) was conducted, and revealed a significant effect for group, for time and for the interaction between group and time. Tables 7, and 8 show the results of one-way ANOVA for each test, and the Scheffe post hoc
comparisons for the posttests, respectively (The alpha level for this study was set at .05).

### TABLE 5

**Descriptive Statistics on VOC Tests for the Control, Input, and Output Groups**

<table>
<thead>
<tr>
<th>Group</th>
<th>Pretest M</th>
<th>Pretest SD</th>
<th>Posttest 1 M</th>
<th>Posttest 1 SD</th>
<th>Posttest 2 M</th>
<th>Posttest 2 SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control</td>
<td>98.57</td>
<td>13.37</td>
<td>105.71</td>
<td>13.45</td>
<td>106.41</td>
<td>12.99</td>
</tr>
<tr>
<td>Input</td>
<td>99.44</td>
<td>13.45</td>
<td>129.63</td>
<td>8.17</td>
<td>117.69</td>
<td>9.62</td>
</tr>
<tr>
<td>Output</td>
<td>98.00</td>
<td>12.33</td>
<td>123.67</td>
<td>13.91</td>
<td>120.52</td>
<td>13.78</td>
</tr>
</tbody>
</table>

### TABLE 6

**Repeated Measures ANOVA on VOC Tests for the Control, Input, and Output Groups**

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>SS</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group</td>
<td>2</td>
<td>3978.86</td>
<td>4.59</td>
<td>.014</td>
</tr>
<tr>
<td>Error</td>
<td>60</td>
<td>25989.11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time</td>
<td>2</td>
<td>13689.29</td>
<td>277.71</td>
<td>.000</td>
</tr>
<tr>
<td>Time*Group</td>
<td>4</td>
<td>2774.65</td>
<td>28.14</td>
<td>.000</td>
</tr>
<tr>
<td>Error</td>
<td>120</td>
<td>2957.58</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### TABLE 7

**One-way ANOVA for Each VOC Test for the Control, Input, and Output Groups**

<table>
<thead>
<tr>
<th>Test</th>
<th>SS</th>
<th>DF</th>
<th>F</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-test</td>
<td>4.062</td>
<td>2</td>
<td>.01</td>
<td>.990</td>
</tr>
<tr>
<td>Posttest 1</td>
<td>4693.77</td>
<td>2</td>
<td>14.75</td>
<td>.000</td>
</tr>
<tr>
<td>Posttest 2</td>
<td>2055.68</td>
<td>2</td>
<td>6.38</td>
<td>.003</td>
</tr>
</tbody>
</table>

### TABLE 8

**Scheffe post hoc Comparisons on VOC Posttests for the Control, Input, and Output Groups**

<table>
<thead>
<tr>
<th>Test</th>
<th>Groups</th>
<th>Diff.</th>
<th>SD</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Posttest 1</td>
<td>Control vs. Input</td>
<td>23.91</td>
<td>4.61</td>
<td>.000</td>
</tr>
<tr>
<td></td>
<td>Control vs. Output</td>
<td>17.95</td>
<td>4.02</td>
<td>.000</td>
</tr>
<tr>
<td></td>
<td>Input vs. Output</td>
<td>5.96</td>
<td>3.84</td>
<td>.308</td>
</tr>
<tr>
<td>Posttest 2</td>
<td>Control vs. Input</td>
<td>11.54</td>
<td>4.64</td>
<td>.053</td>
</tr>
<tr>
<td></td>
<td>Control vs. Output</td>
<td>14.37</td>
<td>4.04</td>
<td>.003</td>
</tr>
<tr>
<td></td>
<td>Input vs. Output</td>
<td>2.83</td>
<td>3.86</td>
<td>.766</td>
</tr>
</tbody>
</table>

Table 5 – Table 8 indicate the results comparing the control, input and output 1 and 2 groups; 1) a significant difference between the control and the two experimental groups on the two posttests, 2) the input group’s superior performance over the output group in posttest 1 at a non-significant level, 3) the output group’s superior performance over the input group in posttest 2 at a non-significant level, and 4) the significant difference between times for both the input and output groups.

The second set of analyses was conducted to see if there is a difference between
output-input and input-output groups. A repeated ANOVA design was conducted with the raw scores of the IO, and OI groups on the vocabulary tests. Table 9 presents the descriptive statistics comparing the groups in terms of the mean scores for the pretest and the two posttests. It shows no significant difference between the [IO] and [OI] groups. Also, the ANOVA analysis shown in Table 10 revealed that there was no main effect for group, nor for the interaction of group and time. Therefore, no post hoc tests were run for each test.

### TABLE 9

Descriptive Statistics on VOC Tests for the [IO] and [OI] Groups

<table>
<thead>
<tr>
<th>Group</th>
<th>Pretest</th>
<th></th>
<th>Posttest 1</th>
<th></th>
<th>Posttest 2</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
<td>M</td>
<td>SD</td>
<td>M</td>
<td>SD</td>
</tr>
<tr>
<td>Input-Output</td>
<td>98.45</td>
<td>11.31</td>
<td>128.26</td>
<td>11.84</td>
<td>124.48</td>
<td>12.52</td>
</tr>
<tr>
<td>Output-Input</td>
<td>98.79</td>
<td>12.45</td>
<td>125.88</td>
<td>14.54</td>
<td>123.85</td>
<td>13.75</td>
</tr>
</tbody>
</table>

### TABLE 10

Repeated Measures ANOVA on VOC Tests for the [IO] and [OI] Groups

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>SS</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group</td>
<td>1</td>
<td>38.22</td>
<td>.08</td>
<td>.767</td>
</tr>
<tr>
<td>Error</td>
<td>62</td>
<td>26726.36</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time</td>
<td>2</td>
<td>31337.27</td>
<td>508.84</td>
<td>.000</td>
</tr>
<tr>
<td>Time*Group</td>
<td>2</td>
<td>60.52</td>
<td>.93</td>
<td>.377</td>
</tr>
<tr>
<td>Error</td>
<td>124</td>
<td>3818.26</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

2. Written Production Tests

Two sets of analyses were conducted to investigate the effect of the output condition on participants’ target word acquisition; 1) comparison of the control, input, and the output groups (OI+IO), and 2) comparison of the Output-Input (OI) and Input-Output (IO) groups. In comparing 1) control, input, and the output groups (OI+IO), two independent variables and one dependent variable were involved in this analysis. The independent variables were the treatment group having three levels (control, input, output group) and time having two levels (posttest 1, posttest 2). The dependent variable was the number of target words counted in the writing tests. The raw scores of the written tests were submitted to a 3x2 ANOVA design.

The descriptive measures shown in Table 11 indicate that the mean scores of the output groups were higher than those of the other groups on both tests. The results of the repeated measures ANOVA also confirm this result (Table 12). The repeated measures ANOVA conducted on the raw scores of the two WP tests revealed a significant effect for group, but not a significant effect for time, nor for the interaction between group and time. As a
follow-up test, one-way ANOVAs were conducted on the raw scores of both WP tests (Table 13). The ANOVAs conducted on tests 1 and 2 showed that there were statistically significant differences between the three groups, and the post hoc Scheffe tests (Table 14) indicated that the output group performed significantly better on the posttest 2 than the control and input group.

### TABLE 11
**Descriptive Statistics on WP Tests for the Control, Input, and Output Groups**

<table>
<thead>
<tr>
<th></th>
<th>Posttest 1</th>
<th></th>
<th>Posttest 2</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
<td>M</td>
<td>SD</td>
</tr>
<tr>
<td>Control</td>
<td>4.14</td>
<td>1.09</td>
<td>3.86</td>
<td>1.51</td>
</tr>
<tr>
<td>Input</td>
<td>6.19</td>
<td>1.47</td>
<td>5.81</td>
<td>1.42</td>
</tr>
<tr>
<td>Output</td>
<td>7.48</td>
<td>2.55</td>
<td>7.27</td>
<td>2.18</td>
</tr>
</tbody>
</table>

### TABLE 12
**Repeated Measures ANOVA on WP Tests for the Control, Input, and Output Groups**

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>SS</th>
<th>F</th>
<th>Sig</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group</td>
<td>2</td>
<td>227.74</td>
<td>18.12</td>
<td>.000</td>
</tr>
<tr>
<td>Error</td>
<td>60</td>
<td>377.03</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time</td>
<td>1</td>
<td>2.31</td>
<td>1.51</td>
<td>.224</td>
</tr>
<tr>
<td>Time*Group</td>
<td>2</td>
<td>.14</td>
<td>.04</td>
<td>.954</td>
</tr>
<tr>
<td>Error</td>
<td>91</td>
<td>136.85</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### TABLE 13
**One-way ANOVA for Each WP Test on the Control, Input, and Output Groups**

<table>
<thead>
<tr>
<th>Test</th>
<th>SS</th>
<th>df.</th>
<th>F</th>
<th>Sig</th>
</tr>
</thead>
<tbody>
<tr>
<td>Posttest 1</td>
<td>110.87</td>
<td>2</td>
<td>12.97</td>
<td>.000</td>
</tr>
<tr>
<td>Posttest 2</td>
<td>117.01</td>
<td>2</td>
<td>16.50</td>
<td>.000</td>
</tr>
</tbody>
</table>

### TABLE 14
**Scheffe post hoc Comparisons on WP Tests for the Control, Input, and Output Groups**

<table>
<thead>
<tr>
<th>Test</th>
<th>Groups</th>
<th>Diff.</th>
<th>SD</th>
<th>Sig</th>
</tr>
</thead>
<tbody>
<tr>
<td>Posttest 1</td>
<td>Control vs. Input</td>
<td>2.04</td>
<td>.75</td>
<td>.024</td>
</tr>
<tr>
<td></td>
<td>Control vs. Output</td>
<td>3.34</td>
<td>.65</td>
<td>.000</td>
</tr>
<tr>
<td></td>
<td>Input vs. Output</td>
<td>1.30</td>
<td>.63</td>
<td>.107</td>
</tr>
<tr>
<td>Posttest 2</td>
<td>Control vs. Input</td>
<td>1.96</td>
<td>.68</td>
<td>.017</td>
</tr>
<tr>
<td></td>
<td>Control vs. Output</td>
<td>3.42</td>
<td>.60</td>
<td>.000</td>
</tr>
<tr>
<td></td>
<td>Input vs. Output</td>
<td>1.46</td>
<td>.57</td>
<td>.035</td>
</tr>
</tbody>
</table>

As the second set of analyses, a repeated 2x2 ANOVA design was conducted to see if there is a difference between output-input and input-output groups on the writing tests. The independent variables were the task sequence having two levels, [IO] [OI], and the time factor having two levels; the dependent variable was the number of target words counted in
the writing tests. Table 15 presents the descriptive statistics comparing the groups in terms of the mean scores on two writing tests, and indicates that the mean scores of the [OI] group were slightly higher than those of the [IO] group on both tests. However, the ANOVA results shown in Table 16 reveal that there was no main effect for time, nor for group, nor for the interaction of group and time.

### TABLE 15
Descriptive Statistics on WP Tests for the [IO] and [OI] Groups

<table>
<thead>
<tr>
<th>Group</th>
<th>Posttest 1</th>
<th></th>
<th>Posttest 2</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M  SD</td>
<td>M  SD</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Input-Output</td>
<td>7.65 2.73</td>
<td>7.58 2.44</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Output-Input</td>
<td>8.09 2.73</td>
<td>7.82 2.33</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### TABLE 16
Repeated Measures ANOVA on WP Tests for the [IO] and [OI] Groups

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>SS</th>
<th>F</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group</td>
<td>1</td>
<td>3.73</td>
<td>.33</td>
<td>.567</td>
</tr>
<tr>
<td>Error</td>
<td>62</td>
<td>700.07</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time</td>
<td>1</td>
<td>.90</td>
<td>.47</td>
<td>.492</td>
</tr>
<tr>
<td>Time*Group</td>
<td>1</td>
<td>.34</td>
<td>.18</td>
<td>.671</td>
</tr>
<tr>
<td>Error</td>
<td>62</td>
<td>118.20</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

3. Informal Interview

The informal interview was conducted with ten students from the input and output groups volunteered to be interviewed by the researcher. The participants were asked to discuss the tasks they completed, and after the interview, the researcher translated into English the Korean notes transcribed during the interview. In comments that the participants made regarding the tasks, most students who engaged in the output activities mentioned that they found the picture-related tasks interesting. Also, some indicated the advantage of the picture-related task and of reading the passage regarding the picture description.

Participant 10: I think that the (production) activities that I did at the beginning were exciting. And I was a little happy that I realized that I already knew some of the words! Actually, I got motivated to know more words after the picture (production) activity, so I carefully read the reading passage to find out if there were words that I wanted to know for ‘sup’ (=‘bush’ in English).
Participant 2: When I was reading the passage, I noticed that the words I needed to describe the angry driver in the picture was ‘honking the horn’. I had no idea of what words had to be used, when given the picture..., because we usually do passages about some kind of report or news for reading activities in class. Not with the pictures... For me, it was interesting to make a story and read the passage about the picture.

In terms of the difficulty of the test, however, participants reported that they found the word recognition task (input task 2 for the input group) and the word-matching task (output task 1 for the output group) difficult to complete. The word recognition and matching tasks contained 30 word items, and each item was presented within the sentence. For each item, the participants were asked to read first an underlined target word in the sentence, and then read a similar expression given at the end of the sentence (input task 2), or fill in the blank with one of the words given in the word box (output task 1). The followings are some of the participants’ reports on these tasks:

Participant 1: I understood what was going on in the task (input task 2), but for some sentences it was hard to understand the meaning. And the extra word which was supposed to be the synonym of the word underlined in the sentence [target word] made me feel frustrated because sometimes I didn’t know either of them.

Participant 4: The task, the task where I was asked to find a synonym to replace the word (output task 1) underlined within a sentence was very difficult. If I could have had more contexts, such as paragraphs, then I could find it more easily. But even if I had had more contexts, I could hardly guess some of words, such as ‘honk’. I had to use a dictionary a lot when I was doing the task.

One possible reason that participants felt difficulty in the word recognition/matching task might have been the number of vocabulary items presented. To deal with thirty new items in one task could have been a demanding task for the participants. Also, each target word was presented with a short sentence, but without explicit explanation of its meaning or usage. Such a presentation was expected to provide learners a chance to understand the context in which the target word was used. However, the participants in the current study seemed to find it demanding rather than useful.

In addition, some of participants found the output activity itself difficult to deal with. Especially, some reported that what they needed most to do the output activity was
‘knowing the words.’

Participant 3 : It was difficult for me to find the right word. I tried to guess the meaning within the sentence, but I had to use the dictionary many times. And sometimes, the words I found in the dictionary were those, I believed, I had known. When I saw the definition of the word, I realized that it was familiar to me... I think I forgot the word....

Participant 6 : ... I am not sure if the (production) activities helped me a lot because honestly, I did not have any knowledge to describe the picture in English when I was given the task. All of the picture tasks were hard for me, and I wished I could have some sample passages or a vocabulary index or something.

The excerpts above imply that when their vocabulary proficiency is not yet adequate to deal with the task, the tasks could become too overwhelming for the learner to appreciate the effects of the activities.

In summary, the output tasks were found to be enjoyed by the participants, and they seem to have provided a strong motivation for the participants to do the tasks. However, some of the output tasks were found to be challenging, and it was shown that the effects of the output tasks might have been influenced by learners’ vocabulary proficiency levels. These findings imply that examination of other variables, such as a learner’s proficiency level or other affective factors, might help us to have a more comprehensive picture regarding the effects of output and task interactions on L2 learners’ vocabulary acquisition.

V. DISCUSSION

Hypothesis 1: L2 learners exposed to input plus output vocabulary instruction on target words will gain higher levels of noticing of target words than learners exposed to input only vocabulary instruction.

The question of how noticing interacts with output for L2 vocabulary acquisition is answered by qualitative and quantitative findings from this study. First, learners were expected to notice target words during their production, and the result of the informal interview revealed that providing an opportunity to produce target vocabulary played a crucial role in helping participants’ noticing the gap between their receptive and productive vocabulary knowledge. In the informal interview, some participants reported that they were
surprised at the fact that they could hardly produce the words when doing output tasks, and this made them curious about what the answers to the tasks were, and why they had such difficulty doing the task. Here is one of excerpts;

*I was shocked at my poor vocabulary knowledge. Actually, I am not bad at English, and my scores on tests have been good. After completing the first task (output task 1), I became curious about what the answers were. The next day, the teacher gave me the answers to the questions, but no explanations. When I was reading the passage two or three days later, I realized, ‘Ah, ‘pull over’ is used in this way...’*

This statement indicates that the output activity affected the participant’s increased attention to target words; she evaluated her vocabulary knowledge, and noticed the gap between what she knew and what she needed to know how to do while doing the output tasks. When a sample usage of target words was given, the participant became conscious of how the target words were used in the context. The verbal report of this participant consequently provides support for production of output as an effective tool for increasing learner’s attention to target words.

The quantitative results, however, do not fully support the assumption that production might lead participants to notice the gap between their receptive and productive vocabulary, and that noticing the gap induced by production may lead to better retention of what they noticed. First, although the increases in the mean scores on vocabulary posttest 2 and the two writing tests suggest that the output group attended to the target form better than the input group did, there were not statistically significant group differences. Second, although there was not a significant group difference, the input group showed higher gains of target words than the output group on the immediate vocabulary test (posttest1). Furthermore, there was no difference between the output first and input first presentation group. The output prior to input presentation was expected to further enhance learners’ noticing of target words, but it appears that the output first presentation did not induce greater improvements on vocabulary and writing tests.

This result is similar to findings from other studies: for instance, Izumi and Bigelow (2000) and Sakai (2004) yielded similar findings that although learners noticed linguistic problems by producing the target language, they had difficulty in attending to and incorporating the subsequent input for later production. Some proposed explanations by these authors are that the difficulty of the tasks might have carried too much of a cognitive demand, or that the task design made the noticing of target forms difficult. In this study, it was also indicated that task difficulty could be one of the possible causes in making participants feel difficulty in giving their full attention to noticing target words. In
participants’ informal interview, some of output tasks were reported as difficult for them to handle without any prior information, and the difficulty of that output task could have strained participants’ capacity for paying attention to target words by requiring more attention to deciphering the task itself. Along with the task difficulty, it seems that a sequence of two consecutive output tasks followed by two consecutive input tasks (i.e., output 1-output 2 - input1 - input2) in this study was not effective in helping participants process target words that were noticed. Some output tasks (task 1 and 2) were probably too demanding for participants, and the participants in the output group might have needed some immediate help when they noticed what they needed to do for output tasks 1 and 2. However, participants in the output group in this study did output task 2 after output task 1, instead of doing the input task. Thus, they did not have a chance to see examples of answers or correct target words until they had been provided with the two input tasks, which included a reading passage and comprehension questions. If a participant’s need had been adequately met with immediate input or instruction, the effect of noticing from output could have been increased.

In summary, regardless of theoretical assumptions and qualitative data favoring the output condition on noticing of target words, the quantitative data fail to fully support the results at a statistically significant level.

Hypothesis 2: L2 learners exposed to input plus output vocabulary instruction on target words will gain higher levels of retrieval of target words than learners exposed to input only vocabulary instruction.

The findings of the current study support the claim that output contributes to enhanced lexical retrieval significantly better than input does (de Bot, 1996; Hulstijn & Trompetter, 1998). It was found in this study that those engaged in the output treatment outperformed those exposed to the input only treatment in production of target words. The output group outperformed the input group on the delayed writing test at a significant level, and on the immediate writing test at a non-significant level. This result yields evidence favoring the output condition for fluent retrieval of target words.

This result implies that even though more input could have an effect on participants’ acquisition of target words in the short-term, it may not apply well to other tasks requiring a heavier cognitive load, such as writing a paragraph using target words, or for long-term recall. In other words, the input condition could provide learners a good chance to understand the meanings of the target words and how they are used in context, but such comprehension is not easily transferred to the production of words when have had no prior opportunity to produce those words. This claim is supported by informal interviews conducted after administration of this study’s posttests, in which participant 2 from the
input group pointed out the difficulty she found in the production of words, which she already knew:

*The answer to the question occurred to me. I mean, I knew the answer, because I saw a similar question and pictures when I did the tasks. But, I couldn’t think of the words! I mean, I feel that I knew the word I wanted, but it just did not come out from my lips...*

The statement above is similar to a phenomenon that has often been observed when L2 learners are engaged in productive tasks: the words they think they know or can comprehend are accessible only to a limited degree when it comes to production. That is, this statement supports Henriksen (1999)’s, Laufer and Nation (1995, 1999)’s, and Laufer and Paribakht (1998)’s claim regarding the different dimensions—receptive and productive—of L2 vocabulary knowledge. The authors argue that word comprehension does not automatically predict correct production of words, since the productive use of vocabulary requires stable connections between the semantic and morphological information associated with the words; however, L2 learners are often unable to make such connections correctly. In other words, the ability to produce words develops at a different rate as the ability to comprehend increases in L2, and this study provides evidence that the input conditions alone do not effectively aid development of the ability to produce target words.

**Hypothesis 3**: L2 learners exposed to input plus output vocabulary instruction on target words will gain higher levels of retention of target words than learners exposed to input only vocabulary instruction.

The previous studies have yielded mixed findings regarding whether production of output can enable learners to retain lexical information better than receiving comprehensible input. While a study by Horibe (2003) found that output production did not promote long-term retention of target features better than input exposure, Hulstijn and Laufer (2001) argued for a positive effect of output on retention of target words. The assumption underlying the superior effect of output on the retention of target words is that producing output induces more of the elaboration process from learners, due to its involvement of multiple linguistic processes than does simple exposure to input, and therefore output leads to better retention of target vocabulary.

The current study has found evidence in favor of the effect of output on long-term retention of target vocabulary. Those engaged in the output tasks outperformed those engaged in the input tasks for the post vocabulary and writing tests although a significant group difference was found only on the delayed writing test. In particular, participants in
the output group showed little declination in their mean scores on the first and second vocabulary posttest, whereas the input group indicated a significant loss of target words on the delayed test. These results provide positive evidence for the superiority of output over input tasks on the long-term retention of target words.

Another pedagogical question considered with the findings of the current study is the effect of task frequency with respect to output and input treatments. The degree to which task frequency helps increase retention of L2 words has been questioned (Ellis, 1995). The results of this study imply that the frequency of tasks may not be a determining factor for retention of words in the long term. In this study, the participants in the input group were exposed to target words twice as often as those in the output group were. Those in the input group, however, did not have a significant advantage in long-term retention. This result also supports the argument by Hustljin and Laufer (2001) that tasks inducing high cognitive effort might be more effective in leaving a deep imprint in memory than simple but repeated exposure to the target words. That is, multiple exposure to new words is desirable for further reinforcement, but when we limit the chances for reinforcement, it appears that tasks with high cognitive involvement are more effective for increasing the chance that the information will be retained.

VI. PEDAGOGICAL IMPLICATIONS AND CONCLUSION

This study has found some evidence favoring the output treatments on L2 learners’ target vocabulary acquisition. The participants who received the output treatments outperformed those who received the input treatments on the two writing tests and the delayed vocabulary test although the significant difference was found on the delayed writing test. It appears that the input treatment was effective in promoting short-term learning effects. However, the input treatment was not as effective as the output treatment in terms of the long-term retention of words and transference of comprehended words to productive use. The fact that the input group demonstrated significant losses of target words on the delayed vocabulary test and it did not outperform the output group neither on writing test 1, nor on test 2 implies that while the input condition can provide learners a good chance to understand the meanings of the target words as well as how well they are used in context, such comprehension is not easily transferred to the production of words when learners are given no prior chance to produce them.

The implications from the current study’s findings are first, that output production should be taken into consideration as an essential condition for full development of L2 vocabulary. The results of the current study have shown that the output treatments were a necessary condition for increasing learners’ noticing of words they needed to know in order
to express their ideas, in facilitating retrieval of learned words, and in extending the retention duration of words they have acquired. In particular, the fact that the output treatments affected participants’ acquisition of target vocabulary in the long-term suggests that production of target words is a crucial factor in developing learners’ implicit and automatic knowledge of L2 vocabulary. This result is consistent with L2 research on the roles of output: output practice increases the chances for the reorganization and restructuring of information (McLaughlin, 1990; Swain, 1995), and plays a crucial role in developing productive language skills (Anderson, 1996; de Bot, 1996). Thus, output practice should necessarily be included in L2 vocabulary instruction in order for full development of learners’ vocabulary knowledge.

In addition, it needs to be pointed out that the input condition was an important factor in developing L2 learners’ vocabulary knowledge, in the sense that it provided necessary information to determine in which contexts L2 vocabulary should properly be used. The results of the current study indicated that when L2 learners were not provided sufficient information regarding how the target words were used in context, they had difficulty with further processing of the meanings of words they had learned. Therefore, providing explicit explanations of in which contexts words can be used is as important as forcing learners to produce the words for developing a higher level of both receptive and productive vocabulary knowledge.

As the limitation of the current study, it concerns individual differences in task performance. In this study it appears that the learners’ attitudes toward the tasks and their English proficiency, could be one of the important factors that affected learners’ performance in participating in and completing output tasks. Thus, if the learners’ attitudes were examined in detail along with the quantitative data, greater depth and breadth could be achieved in examining the effect of the output condition on the learners’ vocabulary learning.

Suggestions for future research begin with this limitation. The effect of individual differences in developing learners’ productive language skills is one of the most promising topics to investigate. L2 research has indicated that individual factors such as motivation, learning orientation, and other affective aspects implicitly impact language learning. Furthermore, the production of language, in essence, assumes that there are audiences who read or listen. Thus, L2 learners’ motivation, self-confidence or other individual differences in language learning may significantly influence the degree to which L2 learners benefit from a certain type of task in developing productive skills. Consequently, to get a more comprehensive picture regarding how productive skills develop, learners’ affective dimensions with respect to language learning should be further examined.
REFERENCES


In picture 1, a boy is riding his bicycle down a narrow road. A car approaches the boy from behind, and the hotheaded driver impatiently honks his horn at the boy because he wants to pass him. The driver is probably saying, “come on, come on, move!” The boy looks tired and worried as he tries to move to the side of the road. The man in the car angrily drives past the boy on the bike, and the boy quickly steers away from the passing car. Since the road is so narrow, it causes the boy to lose his balance and fall into the hedge lining the road. The car speeds on without stopping to help him. Although his bicycle did not break, the boy got a little hurt. He got a scratch on his cheek from the bushes, and his shirt got dirty. He picks up his bike and continues on his way. A little while later, the boy sees a car that has broken down on the side of the road. There is a lot of smoke coming out of the car, and a man is looking at the engine under the hood. It is the car of the same driver who was so unkind to the boy before. Instead of helping the man fix the car, the boy rings his bicycle bell as he passes the bitter man. Now the driver is dirty and upset and the boy is smiling, thinking “Ha ha, maybe this old man has learned his lesson!”

1. Why did the driver honk his horn at the boy?
   1) The driver was so happy to drive a new car  
   2) The boy was driving too fast 
   3) The driver wanted to pass the boy  
   4) He was frightened by the narrow road.

2. What happened to the boy when the driver was passing him?
   1) The boy got very upset  
   2) The boy was injured badly, so he had to go to the hospital.
   3) The boy fell into the hedge and got a little hurt.  
   4) The boy cried because his bike broke.

3. What happened to the driver’s car?
   1) The car hit another car in the intersection. 2) The car broke down, but the driver fixed it.
   3) The car ran out of gas.  
   4) The car’s engine, so it had to stop.

4. What did the boy do when he saw the car broken down alongside the road?
   1) The boy stopped his bike and helped the driver  
   2) The boy yelled at the driver 
   3) The boy just passed the man who was working on the broken car  
   4) The boy waved his hand to the driver to say hello.
APPENDIX B

INPUT TASK 2 (=INPUT TASK 5)

다음 아래의 각 문제에서 밑줄 친 부분을 대신할 수 있는 단어를 생각해 봅시다. 시제나 주어 동사 일치를 고려해 적절한 형태로 바뀌어진 정답이 문장 끝 부분에 제시되어 있습니다.

3. The gentleman was hit by a car while walking across the street. knocked down

5. The vending machine stopped working so he had to replace it with a new one. broke down

7. He fell into a row of small trees along the side of the road. hedge

10. The lady loses her balance on a narrow road. lane

12. The lady who helped the old beggar got repaid for her kindness: she won 1st prize! rewarded

13. The parking ticket was not found in my pocket, so I moved the car to the side of the road and stopped to remember where I put it. pulled over

15. An angry driver made a short and loud noise at him to pass quickly. honked the horn

(Selected task items are presented)

APPENDIX C

INPUT TASK 3 (=INPUT TASK 6)

다음 그림을 잘 본 후 그림들을 묘사하는 이야기를 생각해 보세요. 그림과 연관된 질문 1~10에 대해 알맞은 답을 보기 중에서 골라 보세요.

(*The same picture used in the input task 1 is presented)

2. What is the driver of the car doing in Picture 1?
   1) He is pulling over  2) He stops the car in the middle of road
   3) He is honking the car’s horn at the boy  4) He is fixing the engine of his car

3. Why cannot the driver pass the boy in Picture 2?
   1) The road is too muddy for him to drive fast  2) The road is too narrow for him to pass the boy
   3) The car breaks down  4) There are a lot of people walking in front of the car

5. What has happened to the boy in Picture 3?
   1) He was knocked down by the car  2) He was riding the bike
   3) He was approaching the car very closely  4) He was running slowly

7. What does the boy see in front of him in Picture 5?
   1) He sees a car that has broken down on the side of the road.
   2) He sees a man holding a sign saying ‘help’  3) He sees a tree fallen onto the road
   4) He sees a car that has gotten a flat tire

(Selected questions are presented)
APPENDIX D
OUTPUT TASK 1

다음 아래의 각 문제에서 밑줄 친 부분과 가장 비슷한 의미를 가진 단어를 페이지 끝에 있는 단어 상자에서 고르세요. 상자 안의 단어는 한번만 사용할 수 있으며 시제나 동사 일치에 맞게 단어의 형태를 바꾸세요.

3. The gentleman was hit by a car while walking across the street.
5. The vending machine stopped working so he had to replace it with new one.
7. He fell into a row of small trees along the edge of the road.
10. The lady loses her balance on a narrow road.
12. The lady who helped the old beggar got repaid for her kindness: she won the 1st prize!
13. The parking ticket was not found in my pocket, so I moved the car to the side of the road and stopped to recall where I placed it.
15. An angry driver made a short and loud noise at him to pass fast.

(Selected task items are presented)

<table>
<thead>
<tr>
<th>frightened</th>
<th>brave</th>
<th>pick up</th>
<th>reward</th>
<th>approach</th>
</tr>
</thead>
<tbody>
<tr>
<td>break down</td>
<td>honk the horn</td>
<td>knock down</td>
<td>punish</td>
<td>pull over</td>
</tr>
<tr>
<td>downhill</td>
<td>worry</td>
<td>ride</td>
<td>impatient</td>
<td>hurt</td>
</tr>
<tr>
<td>hedge</td>
<td>scratch</td>
<td>balance</td>
<td>fix</td>
<td>continue</td>
</tr>
<tr>
<td>consider</td>
<td>wonder</td>
<td>complain</td>
<td>intimidate</td>
<td>cause</td>
</tr>
<tr>
<td>lose</td>
<td>regret</td>
<td>consequence</td>
<td>lane</td>
<td>steer</td>
</tr>
</tbody>
</table>

APPENDIX E
OUTPUT TASK 2

다음 그림을 잘 본 후 그림들을 묘사하는 이야기를 생각해 보세요. 그림과 연관된 질문 1~10에 대해 한 문장 정도로 영작해 보십시오.
(*The same picture used in the input task 1 is presented)

2. What is the driver of the car doing in Picture 1?
3. Why cannot the driver pass the boy in Picture 2?
5. What has happened to the boy in Picture 3?
7. What does the boy see in front of him in Picture 5?

(Selected questions are presented)
APPENDIX F
INFORMAL INTERVIEW QUESTIONS

다음 질문들은 질문자와의 인터뷰 시 토론될 질문들입니다. 질문을 잘 읽어 본 후 본인의 의견을 미리 정리해 봅시다.

1. How did you find a series of tasks you completed for the last two weeks? On which language aspect you think you did (or did not) particularly benefit from the activities you completed? (for example, vocabulary, grammar, writing…) Please feel free to express your opinion.
2. Was there any particular task you found difficult or interesting? And why? Please be specific.
3. How did you find doing production activities? (for example, a short writing activity with a picture). Please be specific.
4. How did you feel about the way the tasks were presented? Did you like the way (writing first or reading passage first or just reading passage) the tasks were presented? And why (not)?

Applicable levels: Secondary, college students
Key words: L2 vocabulary, output, task design

Sun-Hee Kwon
Department of English
Pusan University of Foreign Studies
15 Seokpo-Ro, Nam-Gu
Busan, 608-738, Korea
Email:skwon@pufs.ac.kr

Received in August, 2007
Reviewed in September, 2007
Revised version received in November, 2007