The Effect of Output Processing on Subsequent Input Processing: A Free Recall Study

Wataru Suzuki
Miyagi University of Education, *Japan*
Nobuya Itagaki and Tomomi Takagi
Miyagi University of Education, Japan
Tomoyuki Watanabe [1]
Sendai Shirayuri Women's College, Japan

ABSTRACT

There is now growing evidence that output promotes second language acquisition. Recently, interest has been shown in examining (a) the effect of output processing on subsequent input processing, and (b) what factors mediate that effect. An experiment was conducted for two proficiency levels of Japanese learners of English under two conditions (output and non-output). First, participants in the output condition wrote a story in English based on four cartoon pictures (output task processing). Second, participants in both conditions read a model story describing the four cartoon pictures (subsequent input processing). Third, both sets of participants were asked, without any advance notice, to recall what they had read (written free recall test). Performance on the written free recall test suggests that (a) output tasks facilitated subsequent input processing, and (b) complex relationships existed among L2 proficiency levels, experiment conditions, and linguistic domains during subsequent input processing. These findings are discussed with reference to second language acquisition research. Implications for pedagogical practice are also considered.

INTRODUCTION

It has been hypothesized in the field of second language acquisition (SLA) that producing a second language (L2) helps learners to learn it. This does not mean that learners learn language first and produce it later. Since language production (i.e., output) is seen as part of the *process*, and not merely the *product*, of L2 learning (Swain, 2005), L2 learning takes place when learners attempt to produce their developing L2 knowledge. There is growing evidence that output promotes L2 acquisition (see Keck, Iberri-Shea, Tracey-Ventura, & Wa-Mbaleka, 2006; Muranoi, 2007; Swain, 2005 for reviews). Interest has been shown with respect to two particular

[1] Wataru Suzuki is a lecturer at Miyagi University of Education in Japan. Nobuya Itagaki is a professor at Miyagi University of Education in Japan. Tomoyuki Watanabe is an associate professor at Sendai Shirayuri Women's College in Japan. Tomomi Takagi is a graduate from Miyagi University of Education, Japan. Correspondence should be sent to Wataru Suzuki, 149 Aza-Aoba, Aramaki, Aoba-ku, Sendai, Miyagi, 980-0845, Japan. Email: suzukiw@staff.miyakyo-u.ac.jp

issues, which form the foci of the present article: (a) the effect of output processing on subsequent input processing, and (b) factors mediating that effect.

Output in SLA Research

Before examining the role of output in L2 acquisition, we illustrate the overall process of L2 acquisition outlined by Gass (1997). Gass' six stages comprise input, apperceived input, comprehended input, intake, integration, and output. Input is any linguistic information available to learners via listening and reading. Learners perceive this input differently, depending on their prior knowledge and their reaction to affective factors such as motivation and anxiety (i.e., apperceived input). The apperceived input is analyzed and comprehended according to meaning and/or form (i.e., comprehended input), after which it can become intake. During the subsequent integration phase, selected intake is converted into long-term memory, causing restructuring and automatization of current linguistic knowledge (DeKeyser, 2007). Output is the sixth and final component of Gass' (1997) L2 acquisition process.

Since Swain's (1985) seminal paper, the field of SLA has generated encouraging, though not conclusive, findings for four roles of output: (a) noticing (e.g., Hanaoka, 2007; Izumi, 2002), (b) hypothesis-testing (e.g., Ellis & He, 1999; McDonough, 2005), (c) metalinguistic reflection (e.g., Storch, 2008; Swain & Lapkin, 2002), and (d) fluency/automaticity (e.g., Bygate, 2001; DeKeyser, 1997). First, in producing language, L2 learners likely notice the gaps between what they want to say and what they can actually say, a process which is considered to facilitate intake. Second, production of the L2 as hypothetically appropriate statements may help learners ascertain whether their command of the target language is accurate. Third, by engaging in metalinguistic reflection on their use and knowledge of the target language, learners may facilitate the integration of those hypotheses into their interlanguage systems. Fourth, language output facilitates access to integrated knowledge via enhanced fluency and automaticity.

The output-triggered psycholinguistic processes described above may help learners to process new relevant information during subsequent input processing. To borrow Gass' (1997) term, output processing serves as a priming device in which learners are prompted to process relevant information during subsequent input processing. Subsequent input processing may also serve as a reinforcing device in which learners can confirm or disconfirm information about lexical and grammatical problems encountered during output processing (Gass & Alvarez Torres, 2005).

Previous Studies on Output and Subsequent Input

Several studies have examined the effect of output processing on subsequent input processing (Hanaoka, 2007; Izumi, 2002; Izumi & Bigelow, 2000; Izumi, Bigelow, Fujiwara, & Fearnow, 1999; Qi & Lapkin, 2001; Song & Suh, 2008). In Qi and Lapkin (2001) and Hanaoka (2007), participants engaged in three-stage writing tasks in which they (1) wrote first drafts, (2) compared these drafts with a reformulated version or native speaker models, and (3) revised their drafts. In Qi and Lapkin's study, two participants who wrote the draft were asked to think aloud about the difference between their own drafts and the reformulated version they received in Stage 2. In Hanaoka's study, participants were asked to take notes about linguistic problems they noticed during Stage 1, and perceived differences between their drafts and native speakers' models during Stage 2. In both studies, participants noticed their linguistic problems during

output, found solutions in the relevant input (reformulation/model), and incorporated them into subsequent revisions. Although both studies demonstrated the effect of output and noticing on subsequent input processing, the validity of the think-aloud and note-taking procedures they used should be reconsidered (Ericsson & Simon, 1993). The positive effect of noticing on L2 learning might have been mediated by the think-aloud and note-taking procedures themselves. Without these procedures, the participants' attention might not have been directed to target grammatical and lexical features.

Izumi and his colleagues (Izumi, 2002; Izumi & Bigelow, 2000; Izumi et al., 1999) compared an output group with a non-output group in terms of noticing and acquisition (cf. Song & Suh, 2008). In these studies, the grammatical targets differed (e.g., English relativization in Izumi, 2002; the English past hypothetical conditional in Izumi & Bigelow, 2000, and Izumi et al., 1999; and the English past counterfactual conditional in Song & Suh, 2008). Because most of these studies were similar in terms of research design, we will describe Izumi's (2002) carefully designed study. Based upon his previous experiments (Izumi & Bigelow, 2000; Izumi et al., 1999), Izumi (2002) attempted to ascertain whether producing output would promote the noticing and learning of English relativization (i.e., object-of-preposition type). He asked two groups of adult L2 learners, an output and a non-output group, to read a short story containing many examples of English relativization (i.e., input 1). While reading the story, the learners were asked to take note of any words that they thought might be useful for subsequent tasks (note-taking 1). Afterwards, the output group performed a text reconstruction task (output task 1), and the non-output group performed a text comprehension task (non-output task 1). Then, both groups were asked to re-read the same story (input 2) and to take notes for their respective purposes (note-taking 2). Finally, the output group performed a second text reconstruction task (output task 2), and the non-output group performed another text comprehension task (non-output task 2). For one of the measures of noticing, Izumi counted tokens of language produced during note-taking 1 and 2 (i.e., head nouns, relative pronouns, and prepositions) and compared the means between the output and non-output groups. It was predicted that participants in the text reconstruction condition (output task 1) would take note of more words (especially involving relativization) important for their subsequent text reconstruction task than non-output group participants. That is, cognitive processes (e.g., noticing) triggered by output would facilitate subsequent input processing. As predicted, the output group outperformed the non-output group on their learning of English relativization as measured by production and comprehension posttests. However, no difference was found in the amount of note-taking 2 between the output and non-output groups. Similar findings were seen in Izumi and Bigelow (2000), Izumi et al. (1999), and Song and Suh (2008).

Purpose and Rationale

In our opinion, the studies described above demonstrate the effect of output on L2 acquisition, but do not elucidate its effect on cognitive processes (e.g., noticing). Following their findings, one might argue that L2 learners acquire the target language without noticing (cf. De Jong, 2005; Williams, 2005). This counter-intuitive finding, as well as methodological issues and the importance of output in L2 acquisition, motivated the present study. Based on Gass' (1997) L2 acquisition model, we argue that the note-taking task employed by most previous researchers (except that in Qi & Lapkin, 2001) may not be a sensitive enough measure to capture cognitive

processes such as noticing or intake (see Izumi, 2003). In our study, we used a free recall task to measure the effect of output processing on subsequent input processing.

Three things should be noted here. First, we do not argue that a free recall task is more appropriate than a note taking task for investigating the effect of output on subsequent input processing. Rather, this study explores the use of a free recall task. We recognize that cognitive processes related to L2 acquisition can be also conceptualized from an information processing perspective in cognitive psychology. As Gass (1997) argues, the model of L2 acquisition is clearly based upon the information processing model of human cognition. In this paradigm, free recall, among other elicitation methods, has been a valid and representative measure to explore various cognitive processes (Anderson, 2005).

Second, we do not equate recall performance with acquisition. During model essay comprehension, recall performance may simply reflect that selected information is being processed in working memory (i.e., intake). Acquisition, in SLA research, is generally defined as leading to changes in long-term memory as measured by improvement from pretest to posttest. However, there may be an initial stage of storing information (i.e., intake) which triggers long-term memory changes.

Third, unlike many SLA studies, we did not use verbal protocols in the current study. Some types of verbal protocols may influence participants' cognitive processes during the task, thus changing task performance (Ericsson & Simon, 1993).

Many SLA researchers have attempted to explain possible variables affecting the effect of output on L2 acquisition: (a) task type (text reconstruction vs. picture-cued writing; Song & Suh, 2008), (b) level of L2 proficiency (advanced vs. intermediate; Hanaoka, 2007; Swain & Lapkin, 1995; Williams, 1999), (c) linguistic domain (grammar vs. vocabulary; Gass & Alvarez Torres, 2005; Qi & Lapkin, 2001), and (d) modality (oral vs. written output; Izumi & Izumi, 2004). We chose L2 proficiency and linguistic domain out of these variables. The reason for choosing L2 proficiency as a variable is that SLA findings have consistently shown that cognitive processes (e.g., noticing) generated by output are likely affected by participants' L2 proficiency (Leeser, 2004; Swain & Lapkin, 1995; Williams, 1999). Thus, we explore the influence of L2 proficiency on the link between output and subsequent input processing.

We chose linguistic domain as a second variable because a number of SLA researchers argue that since linguistic domains (e.g., lexis and grammar) are considered to be processed and stored differently, L2 learners may require more or less focused attention depending on the linguistic domain in question (Hulstijn, 2003; Schmidt, 2001; VanPatten, 2007). In studies by Gass and her colleagues (Gass & Alvarez Torres, 2005; Gass, Svetics, & Lemelin, 2003), learners required less focused attention for vocabulary learning than grammar learning. Mackey, Gass, and McDonough (2000) demonstrated that learners are more accurate in interpreting the intention of corrective feedback on lexical errors when compared with morphosyntax errors. Also, several studies have shown that (a) L2 learners predominantly notice lexical problems more than grammatical ones (Hanaoka, 2007; Williams, 1999) and (b) high proficiency learners notice grammatical problems more than less proficient learners (Qi & Lapkin, 2001; Swain & Lapkin, 1995). Because little is known about the relationships between L2 proficiency and linguistic domain, we explore this connection in relation to the effect of output on subsequent input processing.

Research Questions

One major question and two related questions guide the present study. These are formulated as follows:

1. Does output processing facilitate subsequent input processing?
2. What is the influence of L2 proficiency on subsequent input processing facilitated by output processing?
3. Is there a relationship among treatment types (output vs. non-output), L2 proficiency levels (low-intermediate vs. high-intermediate), and linguistic domains (vocabulary vs. grammar) during subsequent input processing?

METHOD

Participants

The participants were 116 Japanese learners of English as a foreign language (EFL), 66 of whom were high school students, and 50 were university students. The first participant group was composed of first year high school students (ages 15-16) who had studied English for almost four years. We judged the high school participants to be at a low-intermediate proficiency level of EFL based on their scores in the high school entrance examination administered by the local prefectural government. The examination mainly focuses on grammar, vocabulary, reading, and listening, but not speaking. The second participant group of third year university students (ages 21-23) was judged to be at a high-intermediate proficiency level because they had been studying English for at least nine years. They had also been admitted as English Education majors based upon their higher than average scores on the two-stage national university entrance examination, which focuses on grammar, vocabulary, reading, writing, and listening, but not speaking. The two participant groups will hereafter be referred to as the low-intermediate and high-intermediate groups[2]. We randomly divided each of these groups into two subgroups, with 23 from the high-intermediate group and 31 from the low-intermediate group assigned to the output condition, and 27 from the high-intermediate group and 35 from the low-intermediate group to the non-output condition, described below. The four treatment groups are outlined in Table 1.

[2] The descriptions we provided were *impressionistic* (Thomas, 2006). The practical rationale to categorize the high school and university students as low-intermediate and high-intermediate proficiency was supported by results that showed a significant effect of L2 proficiency on recall performance (see Table 3). However, our results may not accurately distinguish between the effects of L2 proficiency and maturity because of the age differences between the low- and high-intermediate learners.

TABLE 1
The Number of Participants in the Four Treatment Groups

	Treatment type		
Proficiency level	Output	Non-output	Total
Low-intermediate	23	27	50
High-intermediate	31	35	66
Total	54	62	116

Procedure and Materials

One of the researchers implemented the experimental procedure with the high-intermediate group. A high school teacher we had contacted conducted the experiment with the low-intermediate group. For both groups, the experiment was conducted during 30 minutes of regularly scheduled high school and university EFL lessons. The study consisted of four phases (see Table 2). First, participants in the output condition were given ten minutes to write a story in English based upon four cartoon pictures (see Appendix A: Output task). Second, participants in both conditions had three minutes to read a model story written by a native speaker of English describing the four cartoon pictures (see Appendix B: Input task). In the input task, participants in both conditions were provided with the cartoon pictures together with the model story. Third, participants in both conditions solved math exercises (e.g., 26 + 50 = ?, 24 × 115 = ?, 1216 ÷ 32 = ?) for two minutes (math task). Finally, participants in both conditions were asked to take seven minutes to recall, in writing, what they had read (recall task).

TABLE 2
Experimental Procedures

	Output group	Non-output group
1	Write story (10 min)	
2	Read model story (3 min)	
3	Perform math task (2 min)	
4	Recall model story (7 min)	

Four things should be noted here. First, when participants in the output condition wrote their stories based on the cartoon pictures, we provided them with the storyline written in Japanese. The reason we provided the storyline is to assure that the story was not open to varying interpretations. In Hanaoka (2007) and Swain and Lapkin (2002), participants reported that the native speakers' interpretations of the pictures were different from their own. Our pilot study also showed that when the storyline was not provided, some participants wrote their stories differently from the model story written by a native speaker of English. The gaps between their written production and the model might have caused participants to experience difficulty engaging in *cognitive comparisons* (Ellis, 1997) of differences between their production and the model. It has been suggested that cognitive comparisons are important for L2 acquisition (Doughty, 2001; Gass, 1997; Swain, 2005). Reducing the gaps between output (written production) and subsequent input (model) would be more likely to foster cognitive comparisons.

Second, we asked participants in both conditions to perform the math task (two minutes) before the recall task. The reason for this was that we attempted to minimize the effect of rote memory on the recall task. If we had asked participants to recal l the text immediately after reading it, the effects of rote memory and output processing on the recall task could have been confounded. A math task irrelevant to the experiment served to tease these effects apart.

Third, participants were not allowed to use dictionaries during the experiment. We wanted to ensure that participants would work through their linguistic problems with language production on their own. Furthermore, 10 minutes was deemed too short to write a story using a dictionary.

Finally, unlike other studies (Izumi, 2002; Izumi & Bigelow, 2000; Izumi et al., 1999; Song & Suh, 2008), we did not inform both participant groups in advance about the sequence of the tasks. That is, participants were asked to read the model story without any indication that they would later be tested for its retention. This may have prevented participants from consciously memorizing the L2 input.

Scoring and Analysis

In our examination of participants' intake during subsequent input processing, we used two categories (i.e., word and grammar) to analyze how much of the model story participants accurately recalled. (N.B., spelling errors were permitted for this analysis.)

For the word level analysis, we counted the total number of words per sentence participants accurately recalled from the model story. We scored one point for each word and calculated the average of the scores for the two groups. As the total number of words in the model story is 83, the maximum word score is 83.

For the grammar level analysis, we focused only on 20 predetermined grammatical expressions (see Appendix B). We awarded one point for each correctly recalled expression. As we had 20 expressions, the maximum grammar score was 20. The rationale for selecting these expressions was primarily pedagogical and practical rather than theoretical. Prior to the current study, two of the authors conducted a pilot study with a high school teacher (a non-native speaker of English) who implemented the experimental procedures with his 40 students. We asked the teacher to select target grammatical expressions for the model essay which he wanted his students to pick up or to have at least receptive knowledge about. For the current study, we examined the target expressions the high school teacher had selected and chose 20 expressions which had the lowest pilot study recall rate.

Since coding and scoring can result in subjective interpretations, we adopted a very strict criterion for recall accuracy. We defined recall accuracy as the usage of the exact words and grammatical expressions in the model essay. We did not award points for any different words or grammatical expressions that the participants used in the recall task, even when their meanings were similar to those of the original words and expressions in the model essay. Since the coding categories were transparent in the data, we felt that a second coding was not necessary.

For statistical analysis, we conducted an analysis of variance (ANOVA) for each dependent variable (word and grammar) with two independent variables (L2 proficiency level and experimental condition). The procedure of using a Bonferroni-adjustment on alpha to protect against an inflated familywise (or experimentwise) Type I error rate was recommended since multiple univariate ANOVAs were performed (Huberty & Morris, 1989). Therefore, we set the

p-value at .025 (i.e., .05 divided by two dependent variables). We used a statistical package for the social sciences (SPSS) version 12.0 to conduct the statistical analyses.

RESULTS

Table 3 shows the mean and standard deviations of the recall scores of word and grammar analyses for the two experimental conditions and the two L2 proficiency groups. Multiple univariate ANOVAs on the means produced (a) significant main effects of L2 proficiency level and type of condition on word and grammar scores, and (b) only the significant interaction between L2 proficiency and type of condition on the grammar score, which will be reported below.

TABLE 3
Means and Standard Deviations of the Recall Scores of Word and Grammar for Two Types of Experimental Condition and Two Levels of L2 Proficiency

Participants	Type of experimental condition			
	Output condition		Non-output condition	
	Word	Grammar	Word	Grammar
Low-intermediate proficiency group	31.13 (10.79)	3.84 (2.25)	28.83 (12.87)	4.06 (2.6)
High-intermediate proficiency group	56.35 (13.47)	10.38 (4.12)	47.13 (10.29)	7.26 (3.39)

Note*:* Standard deviations are given in parentheses.

For L2 proficiency level, the findings from the ANOVA on the word and grammar scores were significant: F (1, 112) = 87.9, $p < .00$, partial $\eta^2 = .44$, and F (1, 112) = 69.80, $p < .01$, partial $\eta^2 = .39$, respectively. The high-intermediate group likely obtained higher scores for all two variables than did the low-intermediate group (see Table 3).

For type of condition, the findings from the ANOVA on the word grammar scores were also significant: F (1, 112) = 7.62, $p < .01$, partial $\eta^2 = .06$, and F (1, 112) = 6.20, $p < .01$, partial $\eta^2 = .053$, respectively. Participants in the output condition tended to obtain higher scores on all two levels than those in the non-output condition (see Table 3)[3].

For the interaction between L2 proficiency and type of condition, only the findings from the ANOVA on grammar were significant: F (1, 112) = 8.2, $p < .01$, partial $\eta^2 = .07$. The high-intermediate group likely obtained higher grammar scores in the output condition than the non-

[3] The non-output group participants also recalled the story they had read to a significant degree. In this study, the non-output group participants read the model story without any indication that they would later be tested for its retention. Thus, the recall performance for the non-output group indicates the amount of incidental learning in the sense that participants must have picked up words and grammatical forms simply by engaging in the reading activity.

output condition, while the low-intermediate group tended to obtain similar grammar scores in both the output and non-output conditions.

DISCUSSION

Output-Triggered Processes and Subsequent Input Processing

Our first research question was, "Does output processing facilitate subsequent input processing?" Participants in the output condition obtained higher recall scores than those in the non-output condition. In other words, participants who wrote a story based on the cartoon pictures (i.e., output task) and then read a model story (i.e., input task) recalled the model story more accurately than those who engaged in the input task only. When participants produced output and then received relevant input, they were prompted to convert some of the information in the input into memory (i.e., intake). When a subsequent task (e.g., free recall) allowed for access to this intake, participants could successfully retrieve the information in the input.

Cognitive processes which are believed to underlie the effect of output on L2 acquisition (see Swain, 2005) may be employed to explain this finding. Producing language, under some circumstances, facilitates noticing, which may be conducive to L2 acquisition. In producing language, L2 learners likely notice the gaps between what they want to say or write in the target language, and what they can actually say or write. Producing language may encourage L2 learners to notice what they have not yet learned (i.e., noticing the hole). In Schmidt's (2001) terms, noticing processes during production may facilitate not only awareness at the level of noticing (i.e., simple detection of error) but also awareness at the level of understanding (i.e., noticing of gaps and/or holes).

Thus, the cognitive processes that output generates may facilitate drawing the learner's attention to relevant linguistic information in the subsequent input (Izumi, 2002; Schmidt, 2001). As de Bot (1996) argues, "actively making [a] particular trace in memory is more effective than merely perceiving it. The explanation probably lies in the amount of attention invested" (p. 549). In other words, the cognitive processes that output triggers allow learners to engage in a cognitive comparison to notice the gap between their interlanguage and the target language contained in the input (Doughty, 2001; Ellis, 1997; Swain, 2005). When participants experienced difficulty in producing the target language (i.e., output task) and relevant input was immediately available (i.e., input task), they may have noticed the difference between what they had written and what a target language speaker had written to convey the same intention (Izumi, 2002).

Influence of L2 Proficiency on Subsequent Input Processing

The second research question was, "What is the influence of L2 proficiency on subsequent input processing facilitated by output processing?" Low-intermediate proficiency participants showed significantly lower recall scores than high-intermediate proficiency participants, regardless of the level of recall (i.e., word and grammar) and the type of condition (i.e., output and non-output). This finding was not surprising and was anticipated. Poorer recall performance for the low-intermediate learners can be explained by the cognitive processes necessary to perform the input task. While reading the model story in a short period of time (i.e., 3 minutes), learners had to attend to materials in the input (i.e., the model story), temporally

process the materials attended to (i.e., using short-term memory), and store some of these materials at the long-term level (i.e., in long-term memory). When they were asked to incidentally recall what they had read, participants had to access this input from long term memory. This process depended on the extent to which their L2 linguistic systems were automatized/procedualized. However, while reading the model story and recalling it, lower proficiency learners likely accessed their explicit linguistic knowledge through more controlled processing than did high-intermediate learners[4]. To borrow Craik's (1986) term, *self-initiated processes* (e.g., retrieval searches and reconstructive activities) of low-intermediate participants were effortful and deliberate. Moreover, they required cognitive resources that may have overloaded participants' cognitive capacities. Overloaded cognitive capacities may not have allowed low-intermediate learners to effectively process information in the input (i.e., model story) and retrieve the attended-to information from long-term memory. This explanation is in line with general SLA findings: due to their limited memory capacity, low proficiency learners experience difficulty in attending to meaning and form at the same time (VanPatten, 1990; Williams, 1999).

Relationships among Output, Proficiency Level, and Linguistic Domain

Research question 3 was "Is there a relationship among treatment types (output vs. non-output), L2 proficiency levels (low-intermediate vs. high-intermediate), and linguistic domains (vocabulary vs. grammar) during input processing?" The differences in the means of the grammar scores between the output and non-output groups varied as a function of participants' proficiency levels. In the case of high-intermediate learners, producing language (i.e., writing a story) seemed to help them process grammatical information in the model story effectively or deeply enough to remember it for the subsequent recall task. In the case of the low-intermediate learners, however, the output task did not seem to facilitate or hinder their processing of the grammatical information embedded in the model story. It is important to note that the output task helped both high-intermediate and low-intermediate learners recall the model story on the word-level. These results suggest that there may be a threshold of L2 proficiency levels for learners to benefit from output, especially with respect to grammatical processing. The findings are also consistent with Gass and her colleagues' studies which demonstrate that learners require more focused attention for grammar learning than vocabulary learning (Gass & Alvarez Torres, 2005; Gass, Svetics, & Lemelin, 2003; Mackey, Gass, & McDonough, 2000). Moreover, the present study extends previous SLA research by demonstrating the necessity of examining possible L2 proficiency thresholds that may determine the extent to which output processing can facilitate subsequent input processing of different linguistic domains. Clearly, complex relationships among output processing, proficiency levels, and linguistic domains are well worth investigating in future SLA research.

We would like to summarize several of the explanations for our findings. First, the low-intermediate learners may not have been as developmentally ready to process grammatical forms as the high-intermediate learners (Mackey & Philip, 1998; Spada & Lightbown, 1999). Second, the low-intermediate learners might have overlooked certain perceptually non-salient or communicatively redundant grammatical features in the input that did not contribute to meaning,

[4] Controlled processing is defined as "temporary activation of a sequence of elements that can be set up quickly and easily, but requires attention, is capacity-limited (usually serial in nature), and is controlled by subject" (Schneider & Shiffrin, 1977, p. 1).

because their default processing strategy during input processing was to give priority to meaning (VanPatten, 2007). To overcome such a default strategy, they would have had to deliberately pay attention to grammatical forms (Robinson, 2003). However, the attention they could pay to grammatical forms must have been limited when they consumed available processing resources searching for lexical information to understand the meaning of the input (VanPatten, 1990; Williams, 1999). Therefore, the low-intermediate learners failed to perform the recall task at the grammatical level. Third, as the story produced by the low-intermediate learners tended to be different from the model story written by a target language speaker, they might have experienced difficulty in making cognitive comparisons between their output and subsequent input (Doughty, 2001; Ellis, 1997; Swain, 2005). In other words, the degree of overall match between what is produced (i.e., output processing) and what is subsequently comprehended (i.e., input processing) may be important for L2 learning (Lightbown, 2008). Clearly, this needs further empirical investigation.

LIMITATIONS AND IMPLICATIONS

The current study was limited in several ways. The first limitation is the fact that students in the output condition performed one task more than those in the non-output condition. The additional task (i.e., output processing) took 10 minutes, which is more than three times as long as the task that participants under both conditions performed (i.e., input processing task). However, in the cognitive psychology literature (see Lightbown, 2008, for a review), there is a general consensus that more durable memory is generated not by time-on-task, but rather by the type and level of processing (Craik & Lockhart, 1972) and the degree of overall match between conditions at learning and conditions at remembering (Morris, Bransford, & Franks, 1977). We admit that having more time on task affects the type and level of processing possible. Not only did the output group have more time-on-task, but also they may have processed the cartoon at a deeper level, which might not have affected the input task, but only the cued recall task (i.e., recall may have come from the output task directly, not the input task). In order to argue that cognitive mechanisms triggered by output, not time spent performing output tasks, generate the effect of output on subsequent input processing, future SLA studies should control for the amount of time spent on the tasks.

The second limitation concerns additional input in the form of the Japanese story the output group received. The storyline may have provided a further source of input. This extra input could have aided the output group's subsequent input processing[5]. Put differently, participants in the output condition gained prior familiarity with the content while those in the non-output condition were deprived of such an opportunity. This puts the non-output participants at a disadvantage. In such an unfair condition, greater recall of input by the output participants may simply mean that they were familiar with the input content. Since they were provided with the storyline written in Japanese, our task may have functioned as a translation task rather than as

[5] The output group was shown the cartoon pictures three times (i.e., during the output task, the input tasks, and the recall task), while the non-output group looked at the pictures twice (i.e., during the input task and the recall task). This additional input for the output group is not a critical difference that affected the performance of both output and non-output groups in the subsequent recall task because the cartoon picture provided in the recall task equally helped participants in both groups remember the model essay.

an output task[6]. We felt that providing the storyline in Japanese was necessary for our limited L2 proficiency Japanese EFL learners to ease their cognitive capacities to generate the story content and language.

Third, we defined recall accuracy as usage of the same words and grammatical forms as the input task. It is reasonable to assume that features that were noticed and taken into memory would be recalled more often in participants' recal led texts. However, cognitive psychology literature suggests that people tend to store information in meaning form, not in exact wording (Anderson, 2005). Therefore, when participants used different words and grammatical forms, it may have indicated that the information was actually better understood and internalized. It would have been better to use a more lenient scoring method in which equivalent words or expressions would have been counted as correct.

Due to several inherent weaknesses in the particular design of this study, we cannot make a strong claim about the effect of output processing *per se* on subsequent input processing. However, cognitive processes which are believed to underlie the effect of output on L2 acquisition may well be a valid explanation for the current findings of this study. Although the methods of analysis in this study (i.e., free recall) did not demonstrate how cognitive processes (e.g., noticing) were actually implicated in the given study, the use of free recall data suggests a new way of measuring cognitive processes triggered by output (e.g., noticing). The study shows that, compared to the non-output group, the output participants processed information during the subsequent input task (i.e., intake) more effectively, making it available for the subsequent recall task. The output-input instructional sequence (i.e., output tasks immediately followed by relevant input tasks) may have effectively activated cognitive processes such as noticing and cognitive comparisons. With a sequence of an output task (i.e., a priming device) followed by its relevant input task (i.e., a reinforcing device), ESL/EFL teachers can create optimal conditions for triggering acquisitional processes such as noticing.

Another interesting finding is that the differences in the means of the grammar scores between the output and non-output groups varied as a function of participants' proficiency levels. This finding suggests a possible threshold of L2 proficiency levels for learners to benefit from output tasks on subsequent input processing. This deserves further investigation. When L2 teachers employ output tasks, they may be responsible for controlling the kinds of processing (e.g., word level vs. grammar level) they want their students to engage in. This decision may depend on learners' L2 proficiency levels. Limited L2 proficiency students may not always be able to pay attention to grammatical forms on output-based tasks.

We hope that this paper will foster future studies on a timely and important topic in current SLA research.

ACKNOWLEDGEMENTS

An earlier version of this paper was presented at the 15th World Congress of Applied Linguistics (AILA 2008), Essen, Germany in August 2008.We are grateful to Linda Borer, Paul Quinn, Manami Suzuki, Bo-Ram Suh, Merrill Swain, and Yasuyo Tomita for their helpful comments on a draft version of this paper. We are also grateful for the comments by the editor and anonymous

[6] If we had provided the non-output group with the story line in Japanese or if we had not provided another output group with the story line, we could have distinguished the effect of output task from that of a translation task.

reviewers. Of course, all remaining errors and omissions are our own. Our sincere thanks also go to the students for their participation in this study.

REFERENCES

Anderson, R. (2005). *Cognitive psychology and its implications* (6th ed.). New York: Worth Publishers.

Bygate, M. (2001). Effects of task repetition on the structure and control of oral language. In M. Bygate, P. Skehan, & M. Swain (Eds.), *Researching pedagogic tasks: Second language learning, teaching and testing* (pp. 23-48). Harlow, England: Longman.

Craik, F. I. M. (1986). A functional account of age differences in memory. In F. Klix & H. Hagendorf (Eds.), *Human memory and cognitive capabilities* (pp. 409-422). Amsterdam: Elsevier.

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

de Bot, K. (1996). The psycholinguistics of the output hypothesis. *Language Learning*, *46*, 529-555.

De Jong, N. (2005). Can second language grammar be learned through listening? An experimental study. *Studies in Second Language Acquisition*, *27*, 205-234.

DeKeyser, R. (1997). Beyond explicit rule learning: Automatizing second language morphsyntax. *Studies in Second Language Acquisition*, *19*, 195-211.

DeKeyser, R. (2007). Situating the concept of practice. In R. DeKeyser (Ed.), *Practicing in a second language: Perspectives from applied linguistics and cognitive psychology* (pp. 1-18). New York: Cambridge University Press.

Doughty, C. (2001). Cognitive underpinning of focus on form. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 206-257). Cambridge, UK: Cambridge University Press.

Ellis, R. (1997). *SLA research and language teaching*. Oxford, UK: Oxford University Press.

Ellis, R., & He, X. (1999). The roles of modified input and output in the incidental acquisition of word meanings. *Studies in Second Language Acquisition*, *21*, 285-301.

Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis*. Cambridge, MA: MIT Press.

Fleming, G. (1975). *Guided composition for students of English* (2nd ed.). Sevenoaks, Kent: Hodder and Stoughton Ltd.

Gass, S. M. (1997). *Input, interaction, and the second language learner*. Mahwah, NJ: Lawrence Erlbaum.

Gass, S. M., & Alvarez Torres, M. J. (2005). Attention when? An investigation of the ordering effect of input and interaction. *Studies in Second Language Acquisition*, *27*, 1-31.

Gass, S. M., Svetics, I., & Lemelin, S. (2003). Differential effects of attention. *Language Learning*, *53*, 495-543.

Hanaoka, O. (2007). Output, noticing, and learning: An investigation into the role of spontaneous attention to form in a four-stage writing task. *Language Teaching Research*, *11*, 459-479.

Huberty, C. J., & Morris, J. D. (1989). Multivariate analysis versus multiple univariate analyses. *Psychological Bulletin*, *105*, 302-308.

Hulstijn, J. H. (2003). Incidental and intentional learning. In C. J. Doughty & M. H. Long (Eds.), *Handbook of second language acquisition* (pp. 349-381). Malden, MA: Blackwell Publishing.

Izumi, S. (2002). Output, input enhancement, and the noticing hypothesis: An experimental study on ESL relativization. *Studies in Second Language Acquisition*, *24*, 541-577.

Izumi, S. (2003). Comprehension and production processes in second language learning: In search of the psycholinguistic rationale of the output hypothesis. *Applied Linguistics*, *24*, 168-196.

Izumi, S., & Bigelow, M. (2000). Does output promote noticing and second language acquisition? *TESOL Quarterly*, *34*, 239-278.

Izumi, S., Bigelow, M., Fujiwara, M., & Fearnow, S. (1999). Testing the output hypothesis: Effects of output on noticing and second language acquisition. *Studies in Second Language Acquisition*, *21*, 421-452.

Izumi, Y., & Izumi, S. (2004). Investigating the effects of oral output on the learning of relative clauses in English: Issues in the psycholinguistic requirements for effective output tasks. *Canadian Modern Language Review, 60*, 587-609.

Keck, C. M., Iberri-Shea, G., Tracey-Ventura, N., & Wa-Mbaleka, S. (2006). Investigating the empirical link between task-based interaction and acquisition: A meta-analysis. In J. M. Norris & L. Orgega (Eds.), *Synthesizing research on language learning and teaching* (pp. 91-131). Amsterdam: John Benjamins.

Leeser, M. J. (2004). Learner proficiency and focus on form during collaborative dialogue. *Language Teaching Research*, *8*, 55-81.

Lightbown, P. (2008). Transfer appropriate processing as a model for classroom second language acquisition. In Z. H. Han (Ed.), *Understanding second language process* (pp. 27-44). Clevedon, UK: Multilingual Matters.

Mackey, A., & Philip, J. (1998). Conversational interaction and second language development: recasts, responses, and red herrings? *The Modern Language Journal*, *82*, 338-356.

Mackey, A., Gass, S. M., & McDonough, K. (2000). Do learners recognize implicit negative feedback as feedback? *Studies in Second Language Acquisition*, *21*, 557-587.

McDonough, K. (2005). Identifying the impact of negative feedback and learners' responses on ESL question development. *Studies in Second Language Acquisition*, *27*, 79-103.

Morris, C., Bransford, J. D., & Franks, J. J. (1977). Levels of processing versus transfer appropriate processing. *Journal of Verbal Learning and Verbal Behavior*, *16*, 519-533.

Muranoi, H. (2007). Output practice in the L2 classroom. In R. DeKeyser (Ed.), *Practice in second language learning: Perspectives from applied linguistics and cognitive psychology* (pp. 51-84). Cambridge, UK: Cambridge University Press.

Qi, D. S., & Lapkin, S. (2001). Exploring the role of noticing in a three-stage second language writing task. *Journal of Second Language Writing*, *10*, 277-303.

Robinson, P. (2003). Attention and memory during SLA. In C. J. Doughty & M. H. Long (Eds.), *Handbook of second language acquisition* (pp. 630-678). Oxford, UK: Blackwell.

Schmidt, R. (2001). Attention. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 3-32). Cambridge, UK: Cambridge University Press.

Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological Review*, *84*, 1-53.

Song, M.-J. , & Suh, B. R. (2008). The effects of output task types on noticing and learning of the English past counterfactual conditional. *System*, *36*, 295-312.

Spada, N., & Lightbown, P (1999). Instruction, first language influence, and developmental readiness in second language acquisition. *The Modern Language Journal*, *83*, 1-22.

Storch, N. (2008). Metatalk in a pair work activity: Level of engagement and implications for language development. *Language Awareness*, *17*, 95-114.

Swain, M. (1985). Communicative competence: Some roles of comprehensible input and comprehensible output in its development. In S. Gass & C. Madden (Eds.), *Input in second language acquisition* (pp. 235-253). Rowley, MA: Newbury House.

Swain, M. (2005). The output hypothesis: Theory and research. In E. Hinkel (Ed.), *The handbook of research in second language teaching and learning* (pp. 471-483). Mahwah, NJ: Lawrence Erlbaum.

Swain, M., & Lapkin, S. (1995). Problems in output and the cognitive processes they generate: A step towards second language learning. *Applied Linguistics*, *16*, 371-391.

Swain, M., & Lapkin, S. (2002). Talking it through: Two French immersion learners' response to reformulation. *International Journal of Educational Research*, *37*, 285-304.

Thomas, M. (2006). Research synthesis and historiography: The case of assessment of second language proficiency. In J. M. Norris & L. Orgega (Eds.), *Synthesizing research on language learning and teaching* (pp. 278-298). Amsterdam: John Benjamins.

VanPatten, B. (1990). Attending to form and content in the input. *Studies in Second Language Acquisition*, *12*, 287-301.

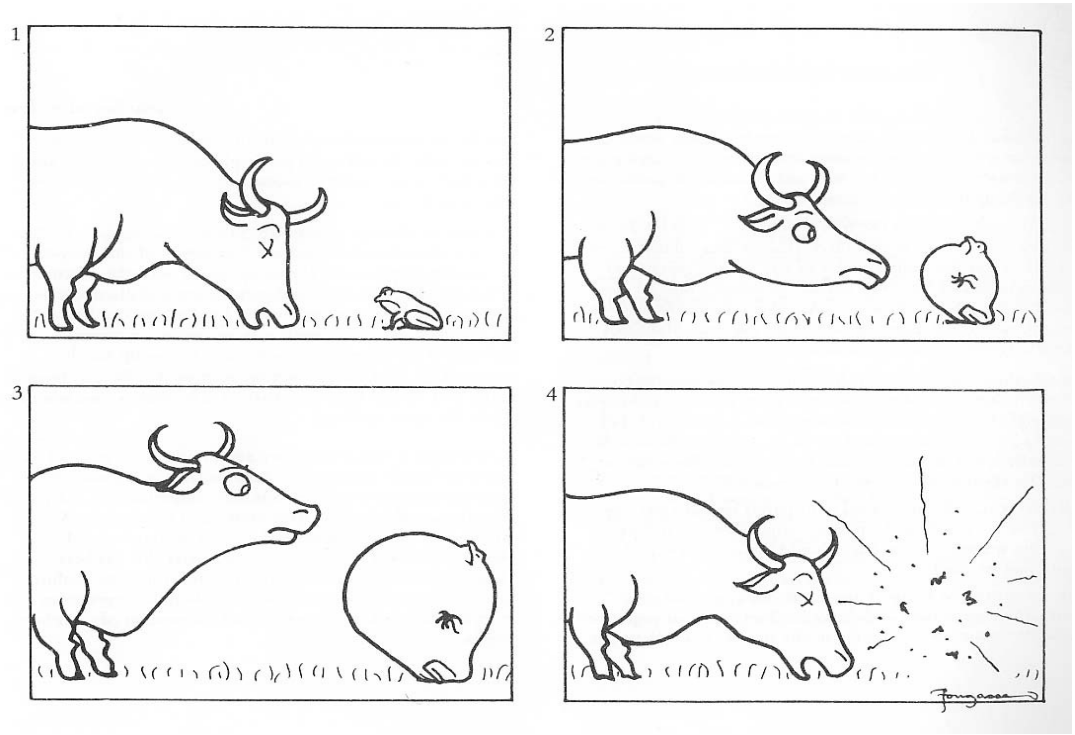
VanPatten, B. (2007). Input processing in adult second language acquisition. In B. VanPatten & J. Williams (Eds.), *Theories in second language acquisition: An introduction* (pp. 115-135). Mahwah, NJ: Lawrence Erlbaum.

Williams, J. (1999). Learner-generated attention to form. *Language Learning*, *49*, 583-625.

Williams, J. N. (2005). Learning without awareness. *Studies in Second Language Acquisition*, *27*, 269-304.

APPENDIX A: Output task

Four Cartoon Pictures



(Fleming, 1975, p. 2)

APPENDIX B: Input task

Model Story

One day, a frog was sitting[1] on the grass, looking at[2] a cow. The cow was eating[3] the grass quietly. The frog thought that the cow was a very big animal[4], and it wanted to be[5] a very big animal, too[6]. So, it began to fill[7] itself[8] up[9] with air. The cow looked at[10] it in surprise[11]. The frog went on[12] filling[13] itself[14] with more air[15] until suddenly — bang![16] It broke into[17] little pieces[18]. The cow went back to[19] eating[20] the grass quietly.

Note: The underlining with the numbers indicates the predetermined 20 target grammatical forms