Promoting Pushed Output Through Task Complexity and Task Condition

Yuka Muraoka

Abstract

The present study examines whether pushed output would be promoted through task complexity and task condition separately or interactively. Task complexity consists of tasks with a few elements (+F) and those with many elements (−F). Task condition factors comprise of open tasks (+O) and closed tasks (−O). The participants were 99 Japanese college students who were randomly assigned into one control group and four experimental groups: +F+O, +F−O, −F+O, and −F−O. All the experimental groups engaged in a picture description task containing the target forms, namely English dative verbs. The same feedback and repeated output opportunities were provided to all the groups four times in total with an interval of two weeks. Pretests and two posttests were given to measure the extent of improvement. Major findings are: (a) there were no main effects of task complexity and task condition; (b) there were no interactional effects; (c) larger amounts of overgeneralized errors were produced by the −O than by the +O, implying that the former had promoted pushed output to a large extent than the latter.

Key Words: Pushed Output, Task Complexity, Task Condition, Dative Verbs

Introduction

In the field of second language acquisition (SLA) research, the significance of output activities as opposed to input-based activities has recently gained increasing attention. One theoretically important suggestion has been made by Swain (1985, 1991, 1993, 1995, 1998; Swain & Lapkin, 1995), who argues that the provision of output opportunities in meaningful contexts is useful to help L2 learners move to more accurate and target-like production of a target language. Swain further suggests that L2 learners have to be pushed to stretch their interlanguage to simulate them to move from semantic processing to grammatical proc-

essing, which may lead to accurate production. There are many plausible ways to promote such pushed output in pedagogical terms. In line with task-based teaching, Robinson (2001 a, 2001b, 2003, 2005) proposes that task features such as task complexity and task condition separately or interactively promote pushed output. Much has not been known yet how task features promote pushed output and influence L2 learners' interlanguage development. The present study attempts to investigate the extent to which task complexity and task condition separately or interactively promote pushed output.

In the following, the illustrations of pushed output and task variables that may facilitate pushed output are presented, followed by discussions on indications of pushed output. Then, the framework proposed by Robinson with respect to task complexity and task condition is elucidated. Dative verbs as a target structure are introduced. Next, research questions, method, results, and discussions will be presented.

Task Variables that Promote Pushed Output

Pushed output is a theoretically important construct and has to be distinguished from a normal sense of output (Swain, 1985, 1993; Swain & Lapkin, 1995). Swain (1993) posits that if the goal of L2 learning is to develop native-like proficiency of a target language, "just speaking and writing are not enough" (p. 160). Learners can get their meaning across in spite of incorrect grammatical forms and sociolinguistically inappropriate expressions. Pushed output denotes that learners need to be "pushed toward the delivery of a message that is not only conveyed, but that is conveyed precisely, coherently, and appropriately" (Swain, 1985, p. 248). Learners need to be put in a situation where their interlanguage is stretched to their fullest, thereby, with more mental effort, reflecting on the source of their output and thinking of some ways to enhance comprehensibility, appropriateness, and accuracy (Swain, 1993, 2000).

Swain and Lapkin (1995) argue that one way to accomplish pushed output is by "pushing learners beyond their current performance level" (p. 374). It has been argued that for L2 learning to take place learners need to be exposed to a linguistic situation where comprehension or production is beyond their current ability to process. When learners encounter a language context which they cannot handle with their existing interlanguage, they may be more pushed to resolve the language problem at hand. Swain (1998) suggests that the study by Tarone and Liu (1995) demonstrate a piece of useful evidence in this regard. It was found that Bob, a Chinese boy who is learning English through natural interactions in Aus-

tralia, displayed a faster acquisition rate of English interrogative forms in interaction between Bob and a researcher, compared with other interactional contexts such as between Bob and his peers or his teachers. Tarone and Liu (1995) argue that the Bob-researcher interaction especially provides important opportunities where Bob is pushed to produce output beyond his current interlanguage level by receiving relevant input from a researcher in a subsequent turn.

One speculation is that tasks in which learners are apt to be pushed beyond their current linguistic levels might be often difficult or challenging ones. In this regard, it can be postulated that L2 learners may be more likely to be pushed beyond their limitations in difficult tasks than in easy tasks. Then, the distinction between difficult versus easy tasks comes into a play. One complicated issue, however, is how difficult versus easy tasks can be empirically operationalized and constructed. In the literature, two suggestions have been put forward by Skehan (1996, 1998) and Robinson (2001a, 2001b, 2003, 2005). The present study employs the framework created by Robinson since it neatly differentiates a range of possible factors that would affect learners' performance during task implementation and systematically integrates all of them into one fixed framework. In addition, Robinson incorporates the concept of 'pushed output' into the framework and proposes that tasks with cognitive complexity are more likely to trigger pushed output compared to tasks with less cognitive complexity. This claim appears to be consistent with the argument presented above that pushed output might be facilitated in tasks that push the limits of learners' current interlanguage systems. Robinson's proposal will be illustrated in detail in a later section.

Indications of Pushed Output

To achieve pushed output, L2 learner's interlanguage has to be stretched in some ways, which may prompt the operation of syntactic processing or hypothesis testing (Kowal & Swain, 1997; Swain & Lapkin, 1995; Swain, 1993, 2000). According to Swain, in producing language, L2 learner's attention is somehow paid to a particular linguistic feature because they notice there is a gap or a hole in their interlanguage. When learners discover the gap or their deficiency, they may attempt "an analysis of incoming data (syntactic analysis of input) or an analysis of existing internal linguistic resources, in order to fill the knowledge gap" (Kowal & Swain, 1997, p. 293). Furthermore, if syntactic analyses of language are conducted in learners' heads, some sorts of modifications at the levels of phonology, morphosyntactic, and semantic may appear in their output, resulting in *modified output*

(Swain & Lapkin, 1995). Modified output, therefore, refers to some qualitative changes appearing in learners' output.

Shehadeh (2002) maintains that qualitative changes in output can be directed toward comprehensibility, correct, accurate, or target-like. Besides, these categories, it may be also significant to scrutinize erroneous modifications. During their attempts to modify their output, L2 learners may sometimes produce inaccurate modifications based on their misconception of grammar. Syntactic analysis does not always lead to the enhancement of performance toward accuracy; learners may sometimes engage in wrong analysis, "leading to incorrect hypotheses and inappropriate generalizations" (Swain & Lapkin, 1995, p. 384). The inspection of inaccurate modifications, which can be categorized as "errors," would also provide valuable insights into understanding what mental processes are triggered by output. When considering pushed output, it seems important that studies take into consideration both accurate and inaccurate modifications learners make in their production. To this end, the present study not only examines positive developmental changes in terms of accuracy but also looks into types of errors in production.

Task Complexity and Task Conditions

Stressing the importance of language task as the unit of analysis for language teaching, Robinson proposes that increasing the cognitive demands of tasks may promote attention to and noticing of some forms in input through interaction and negotiation, leading to interlanguage development in terms of accuracy and complexity (Robinson, 2001a, 2001b, 2003, 2005)[1]. Robinson's framework makes distinctions among a factor that pertains to design features of tasks (task complexity), a factor that relates to learners' perception of task demands (task difficulty), and a factor that is relevant to the nature of task participation (task condition). All of these factors may interactively influence learners' task performance and L2 development. Task complexity, task condition, and their interactional effects are the focus of the study.

According to Robinson, task complexity comprises two dimensions: a resource-directing dimension and a resource-dispersing dimension. This study manipulates the resource-directing dimension and controls the resource-dispersing dimension (e.g., ＋planning time, ＋single task, ＋prior knowledge). The resource-directing dimension denotes that learners are pushed to draw their attention to linguistic forms when tasks impose great cognitive demands on them. This dimension has three subcategories: ＋/-few elements, ＋/-here-and-

now, and +/-no reasoning demands. Among them, +/-few elements is the subject of the study. It is suggested that tasks with more elements (-few elements) would be cognitively more complex than tasks with few elements (+few elements) because the former require learners to use a wider range of linguistic resources to describe distinctive elements than the latter (Anderson, Brown, Shillcock, & Yule, 1984; Brown & Yule, 1983). In the present study, the feature of -few elements (−F) was operationalized as a sequence of a story composed of five pictures where more than three people of the same gender are involved (i.e., three women). On the other hand, the feature of +few elements (+F) was defined as a sequence of a story composed of three pictures where only two people with different gender are involved (i.e., one man and one woman). Since the description of the strip containing many similar elements is generally more demanding than that of the strip containing few dissimilar elements, it can be presumed that L2 learners would be pushed beyond their linguistic limitation to a larger extent in the former task than in the latter task.

Regarding task condition, Robinson identifies two broad variables: participation variables (e.g., open/closed, one-way/two-way, and convergent/divergent) and participant variables (e.g., gender, familiarity, power/solidarity). This study focused on the participation variable, namely open vs. closed task conditions. The distinction between open and closed tasks is specified by Long (1989). Closed tasks entail a single correct solution and require accurate production whereas open tasks indicate there is no predetermined correct solution and there are more variable answers. In the present study, the closed task condition (-open task [−O]) requires the memorization of sentences depicting the cartoon strip while the open task condition (+open task [+O]) demands the description of the strip by using assigned words (e.g., the names of characters). Accordingly, a primary difference in +/−O conditions is whether or not memorization (i.e., accurate reproduction of assigned texts) is required during the task. Since closed tasks oblige learners to memorize sentences, it can be argued that L2 learners are more likely to be pushed beyond their linguistic limitations by being forced to use structures that are difficult for or unfamiliar to them. In open tasks, L2 learners have a chance to avoid difficult structures and can use easy forms to describe pictures. On these accounts, it can be speculated that closed tasks may promote pushed output to a larger extent than open tasks.

Furthermore, it is claimed that task complexity and task condition in conjuncture would play a significant role in the promotion of learners' pushed performance. However, much has not been known to date regarding the interaction among task complexity, task condition,

and pushed output; the close examination of such interwoven relationships is of great importance as an empirical issue.

Dative Verbs

The target structures are dative verbs. The acquisition of dative verbs has been suggested to be difficult for L2 learners in general regardless of their L1 background due to the complex nature of the structures themselves (Carroll & Swain, 1993; Celce-Marcia & Larsen-Freeman, 1999; Ellis, 1991; Hawkins, 1987; Mazurkewich, 1984a, 1984b, 1985). Difficulty in learning dative verbs may arise from the fact that although many dative verbs allow dative alternation, there are verbs that do not allow this. In order to distinguish which verb allows or disallows the dative alternation, L2 learners have to learn morphophonological and semantic constraints. The morphophonological constraint indicates that monosyllabic verbs of Germanic origin (native) permit the dative alternation whereas polysyllabic verbs of Latinate origin appear only in the prepositional construction[2] (Green, 1974; Pinker, 1989). Semantic restrictions are somewhat complicated. Pinker (1989) suggests that the indirect objects can appear in the double-object construction only when they are animate and prospective possessors of the direct objects (i.e., the broad-range rule); however, this constraint can not account for all the examples in English. There is an additional rule (i.e., the narrow-range rule) that applies to particular subclasses of dative verbs[3]. Furthermore, L2 learners have to discriminate *to*-dative verbs between *for*-dative verbs in the prepositional complement. *To*-dative verbs often denote the meaning of transfer while *for*-dative verbs are benefactive.

In sum, the complete acquisition of dative verbs requires the learning of the following three properties: (a) dative alternation, (b) distinction between monosyllabic datives and polysyllabic datives, and (c) distinction between *to*-dative verbs and *for*-dative verbs. The study examined whether the participants could learn the following eight dative verbs: four *to*-dative with monosyllabic (*lend, send*) and polysyllabic (*suggest, explain*) verbs and four *for*-dative with monosyllabic (*buy, bake*) and polysyllabic (*create, select*) verbs. To explore generalizability, four dative verbs from each category (*show, describe, cook, and purchase*) were included as test items.

Research Questions

Based on the arguments presented above, three research questions were formulated:

1. Does a task with higher cognitive demands generate higher accuracy scores of dative verbs than a task with lower cognitive demands?
2. Does a closed task generate higher accuracy scores of dative verbs than an open task?
3. Is there a significant interaction between task complexity and task condition?

Method

Participants

The participants were 99 (31 males and 68 females) Japanese freshmen enrolled in three different private universities in Tokyo at the ages of 18−20, comprising of five intact class-rooms. Their majors were film making, nursing, and economics. They were at the interme-diate level of English proficiency and had completed six years of English education in Japa-nese junior high and high school. The participants from four intact classrooms were ran-domly assigned to one of the four experimental conditions (see Table 1). One whole intact class served as a control group.

Table 1   *Four Experimental and One Control Groups*

	Task Complexity	
Task Condition	[+Few Elements]	[−Few Elements]
[+Open Task]	+F+O (N = 18)	−F+O (N = 25)
[−Open Task]	+F−O (N = 20)	−F/−O (N = 17)
Control Group	(N = 19)	

Note. The +F indicates a cognitively less demanding task while the −F a cognitively more demanding task. The +O refers to an open task while the −O a closed task.

Procedures

Two types of Pretest were given prior to two weeks before the actual treatments began. A treatment was provided four times in total with an interval of two weeks. Immediately after the last treatment was completed, the first immediate posttests (Posttest 1) were provided. The second posttests (Posttest 2) were given seven weeks later to examine the long−term ef-fects of the treatments. The control group was asked to take only Pretest and Posttest 1.

A task in each treatment was composed of four sessions: Preparation, Output, Feedback, and Correction. All of the participants were engaged in the activity of story telling prompted

with a cartoon strip having either three pictures (+F) or five pictures (−F). In the Preparation Session, the participants in the −O were directed to memorize written sentences depicting the strip with three or five pictures while those in the +O were asked to look at the cartoon strip and prepare the description of the drawing. In the Output Session, the −O were asked to reconstruct the sentences they memorized as accurately as possible by just looking at the same cartoon strip without sentences. The +O was directed to depict each cartoon drawing by using assigned words (e.g., names of the characters). In the Feedback Session, the −O was exposed to the same cartoon strip with the sentences and was instructed to check if their reconstruction was accurate or not. Conversely, the +O was provided with the same cartoon strip with the model utterances depicting each drawing at this occasion. Then, they were asked to look for differences between their own writing and the model sentences. At the Correction Session, both groups were directed to reproduce the description of the same cartoon strip by making necessary corrections (see Appendix A for materials used in this study).

This sequence was repeated once again so that the participants engaged in the second version of the materials. For Treatments 1 and 3, which included monosyllabic to−dative (*lend, send*) and *for*-dative (*buy, bake*) respectively, the second version consisted of the same picture and the same story line; however, the participants were exposed to the alternated form of the same dative verbs (e.g., "Kenji lent Tomoko some money" in the first version and "Kenji lent some money to Tomoko" in the second version). Through exposure of the two versions, they were presumed to notice that there were two different ways of expressing the same meaning. Treatments 2 and 4 contained polysyllabic *to*-dative (*suggest, explain*) and *for*-dative (*select, create*) respectively. In these treatments, the same verbs appeared in the prepositional construction in both first and second versions with different pictures but similar story lines. The study speculated that the participants would notice a distinction between monosyllabic and polysyllabic dative verbs through the engagement of the tasks. It was speculated that the differences in the organization of the tasks would prompt the participants to realize that there might be some differences existing between the verbs appearing in Treatments 1 and 3 and those appearing in Treatments 2 and 4.

The amount of time spent in each session was controlled by using a stopwatch. For the Preparation and Feedback sessions, two minutes were assigned; for the Output and Correction sessions, five minutes were allocated. A conductor of the experiments (the author) instructed the participants not to go to the next stage unless they were told to do so.

Measurement

The present study used recognition tests to assess the recognition ability and fill-in-the-blank written production tests to measure the production ability (see Appendix B for example). In the recognition tests, the participants were required to choose one correct use of dative verbs out of four alternatives. In the production tests, the participants were asked to look at pictures. Then, as Celce-Murcia and Larsen-Freeman (1999) suggest, a question such as "What is Jill doing?" was provided as a prompt to elicit the natural use of dative verbs together with indirect and direct objects. Another instruction such as "Answer the question in two different ways if it is possible" was also provided. This instruction seems useful to naturally elicit the production of both the double-object and prepositional constructions. As a response to the question, the participants were required to complete sentences by using the assigned words provided in the parenthesis which consisted of one dative verb, one direct object, and one indirect object. Test items consisted of 14 verbs, including 12 target dative verbs and two non-target verbs as distracters. One point was assigned for a correct choice and the accurate production of a dative verb phrase. A maximum of 18 points was possible in both tests. Since test scores are based on accurate recognition and production abilities, the study will call them "accuracy scores" thereafter.

Analyses

To address research questions, between-groups Analysis of Variance (ANOVA) was separately performed for each test. The independent variables were Task Complexity with two levels (+/−F) and Task Condition with two levels (+/−O); the dependent variable was the accuracy score obtained from the recognition and the production tests. Investigating within-group differences among the tests, one-way repeated measures ANOVAs were performed for each experimental condition by using the accuracy scores from three tests (Pretest, Posttest 1, and Potesttest 2), obtained from both the recognition and the production tests. In the case of the control group, the scores from Pretest and Posttest 1 were submitted to paired *t*-test to explore a difference.

Previous knowledge of dative verbs was examined by using a one-way ANOVA. The result showed there were statistically significant differences among the five groups in the recognition test, $F(4, 103) = 3.143$, $p < .05$, but not in the production test, $F(4, 103) = 2.092$, $p = .87$. A Fisher's LSD post-hoc test revealed that the mean score of the control group was

significantly higher than all the other groups. This difference has to be taken into consideration in interpreting the results of the recognition tests.

As discussed above, pushed output indicates any qualitative changes in L2 learners' interlanguage, both in terms of accuracy and inaccuracy. Therefore, the study also examined types of errors the participants made in the production tests. Error analyses were conducted with respect to three properties of dative verbs. For the sake of brevity, this study only presents the results of the analyses regarding the distinction between monosyllabic and polysyllabic datives. The study calculated the number of overgeneralized errors. The overgeneralized error was identified when the participants wrote such sentences "Bill is describing Jim the story" or "Jim is explaining Mike the answer." The effects of task condition and task complexity on the production of overgeneralized errors were also examined by using nonparametric measures[4].

Results

Descriptive Statistics for the Accuracy Scores of Dative Verbs

Table 2 and Table 3 present descriptive statistics for the recognition and the production tests respectively. Figure 1 and Figure 2 visually display the mean scores of the experimental and control groups. On the whole, in both tests, the accuracy scores went up from Pretest to Posttest 1 but moderately went down from Posttest 1 to Posttest 2 in all the experimental groups except the control group.

The Effects of Task Complexity and Task Conditions

The results of the two-way ANOVAs did not show significant main effects for Task

Table 2　*Descriptive Statistics for Accuracy Scores of Dative Verbs from Recognition Tests*

| Task Condition | Task Complexit | | | | | |
| | [+Few Elements] | | | [–Few Elements] | | |
	Pretest	Posttest 1	Posttest 2	Pretest	Posttest 1	Posttest 2
[+Open Task]	M = 9.06	M = 11.56	M = 10.39	M = 8.40	M = 10.88	M = 9.52
	SD = 3.638	SD = 2.874	SD = 3.415	SD = 2.972	SD = 3.432	SD = 2.801
[–Open Task]	M = 9.35	M = 11.65	M = 10.70	M = 9.24	M = 11.88	M = 9.76
	SD = 2.925	SD = 3.265	SD = 2.473	SD = 2.796	SD = 3.276	SD = 3.308
Control Group	M = 11.53	M = 10.79				
	SD = 2.988	SD = 2.917				

Note. Possible maximum score is 18.

Table 3　*Descriptive Statistics for Accuracy Scores of Dative Verbs from Production Tests*

Task Condition	Task Complexity					
	[+Few Elements]			[−Few Elements]		
	Pretest	Posttest 1	Posttest 2	Pretest	Posttest 1	Posttest 2
[+Open Task]	M = 8. 00	M = 9. 72	M = 9. 33	M = 6. 60	M = 8. 80	M = 8. 28
	SD = 3. 447	SD = 3. 691	SD = 3. 162	SD = 3. 136	SD = 3. 316	SD = 2. 821
[−Open Task]	M = 6. 90	M = 9. 60	M = 8. 35	M = 8. 06	M = 10. 47	M = 8. 76
	SD = 2. 426	SD = 2. 542	SD = 2. 852	SD = 2. 015	SD = 3. 375	SD = 2. 948
Control Group	M = 8. 68	M = 9. 84				
	SD = 3. 433	SD = 3. 862				

Note. Possible maximum score is 18.
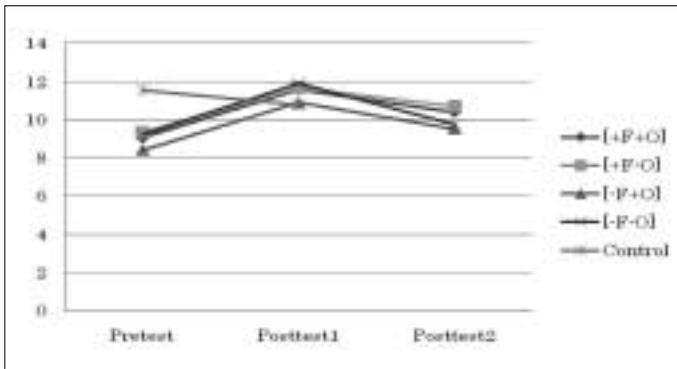
Figure 1　Mean accuracy scores of dative verbs from recognition tests.
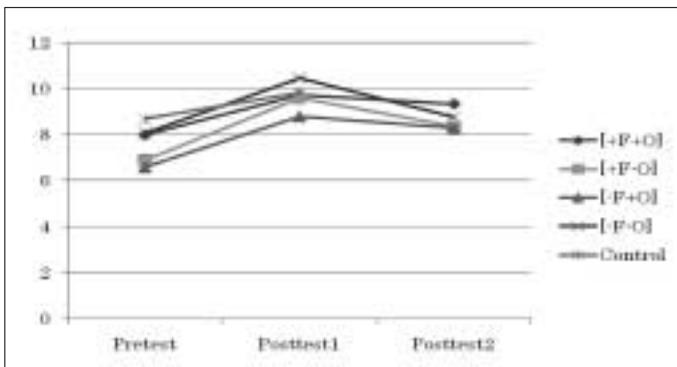
Figure 2　Mean accuracy scores of dative verbs from production tests.

Complexity in the recognition test, $F(1, 76) = .092$, $p = .76$, nor in the production test, $F(1, 76) = .001$, $p = .972$. Similarly, significant main effects for Task Condition were not found in the recognition test, $F(1, 76) = .561$, $p = 456$, nor in the production test, $F(1, 76) = 1.154$, $p = .286$. Significant interactional effects between Task Complexity and Task Condition were not found in the recognition test, $F(1, 76) = .384$, $p = .537$, nor in the production test, $F(1, 76) = 1.547$, $p = .217$. These results seem to show that the participants in all the experimental conditions performed similarly in the recognition and production Posttest 1.

The results of the two-way ANOVAs for Posttest 2 did not reveal that significant main effects for Task Complexity in the recognition test, $F(1, 76) = 1.786$, $p = .185$, nor in the production test, $F(1, 76) = .232$, $p = .632$. Significant main effects for Task Condition were not found in the recognition test, $F(1, 76) = .169$, $p = .682$, nor in the production test, $F(1, 76) = .141$, $p = .708$. Significant interactions between Task Complexity and Task Condition were not found in the recognition test, $F(1, 76) = .002$, $p = .961$, nor in the production test, $F(1, 76) = 1.224$, $p = .272$. Similar to Posttest 1, these results appear to indicate that linguistic performances of the participants were comparable, regardless of different experimental conditions to which they were exposed.

Inspection of Between-Test Differences

The results of the repeated measure one-way ANOVAs using the scores from the recognition tests indicated that significant differences among the tests were found in all of the experimental groups: in the +F+O, $F(2, 34) = 6.547$, $p < .05$; in the +F–O, $F(2, 38) = 5.952$, $p < .05$; in the –F+O, $F(2, 48) = 56.750$, $p < .05$; in the –F+O, $F(2, 32) = 5.182$, $p < .05$. Post-hoc Bonferroni analyses showed that the mean scores of Posttest 1 were significantly higher than those of Pretest in all the experimental groups. However, significant differences in the means were not found between Pretest and Posttest 2, nor between Posttest 1 and Posttest 2 in all of the experimental groups. These findings seem to suggest that the recognition ability of dative verbs similarly improved from Pretest to Posttest 1 in all the groups, regardless of the different treatments in which they were involved. Since there were no significant differences between Pretest and Posttest 2, it can be said that such improvement did not retain until the time of Posttest 2. The repeated measure t-test exploring the test-difference in the control group indicated no significant difference between Pretest and Posttest 1, $t(18) = .995$, $p = .333$. Therefore, contrary to the experimental groups, the recognition ability of the control group did not change over time. Table 4 presents a summary of the

Table 4   *Summary of Between-Test Comparisons in Recognition Tests*

+F+O	+F−O	−F+O	−F−O	Control Group
Pretest < Posttest 1*	Pretest < Posttest 1*	Pretest < Posttest 1*	Pretest < Posttest 1*	Pretest = Posttest 1
Posttest 1 = Posttest 2	Posttest 1 = Posttest 2	Posttest 1 = Posttest 2	Posttest 1 = Posttest 2	
Posttest 2 = Pretest	Posttest 2 = Pretest	Posttest 2 = Pretest	Posttest 2 = Pretest	

*p < .05

between-test comparisons conducted in each group.

As to the production tests, the results of the repeated one-way ANOVAs revealed significant main effects for Tests in all the experimental groups: in the +F+O, $F(2, 34) = 3.943$, $p < .05$; in the +F−O, $F(2, 38) = 6.132$, p < .05; in the −F+O, $F(2, 48) = 6.726$, $p < .05$; in the −F−O, $F(2, 32) = 6.760$, $p < .05$. The multiple comparisons using Bonfferroni analysis demonstrated that the mean scores significantly increased from Pretest to Posttest 1 in all the groups; however, significant differences between Posttest 1 and Posttest 2 were not found in any of the groups. A significant difference between Pretest and Posttest 2 was found only in the −F+O[5]. These results suggested that production ability had improved from Pretest to Posttest 1 in all the experimental groups, but such improvement did not remain until the time of Posttest 2, except for the −F+O condition. With respect to the control group, the repeated measures t-test indicated that there was no significant difference between Pretest and Posttest 1, $t(18) = -1.649$, $p = .116$. Thus, it shows that the production ability of the control group stayed consistent. Table 5 is a summary of the within-test comparisons in groups.

Table 5   *Summary of Between-Test Comparisons in Production Tests*

[+F+O]	[+F−O]	[−F+O]	[−F−O]	Control Group
Pretest < Posttest 1*	Pretest < Posttest 1*	Pretest < Pos−test 1*	Pretest < Posttest 1*	Pretest = Posttest 1
Posttest 1 = Posttest 2	Posttest 1 = Posttest 2	Posttest 1 = Posttest 2	Posttest 1 = Posttest 2	
Posttest 2 = Pretest	Posttest 2 = Pretest	Posttest 2 > Pretest*	Posttest 2 = Pretest	

*p < .05

Error Analyses

Table 6 displays the descriptive statistics for the mean overgeneralized errors made by the participants in writing sentences using polysyllabic dative verbs (i.e., *select, create, purchase, suggest, explain*, and *describe*). On the whole, the overgeneralized errors appeared to increase from Pretest to Posttest 1, but decreased from Posttest 1 to Posttest 2. Figure 3 graphically demonstrates such changes. The −F−O presents the highest mean of the overgeneralized errors among the other conditions in Posttest 1. Furthermore, the overgeneralized errors produced by the −F−O and the +F−O sharply increased from Pretest to Posttest

1 while those produced by the +F+O and the −F+O moderately increased or did not change significantly.

Table 6　*Descriptive Statistics for Overgeneralized Errors of Polysyllabic Datives*

Tests	+F+O	+F−O	−F+O	−F−O
Pretest	M = .72 SD = 1.487	M = .20 SD = .410	M = .48 SD = .872	M = .41 SD = .712
Posttest 1	M = 1.06 SD = 1.552	M = 1.70 SD = 1.985	M = 1.12 SD = 1.856	M = 2.18 SD = 2.270
Posttest 2	M = .72 SD = 1.320	M = .80 SD = 1.152	M = .68 SD = 1.600	M = 1.06 SD = 1.478

Note. The possible maximum point in each test is 6.
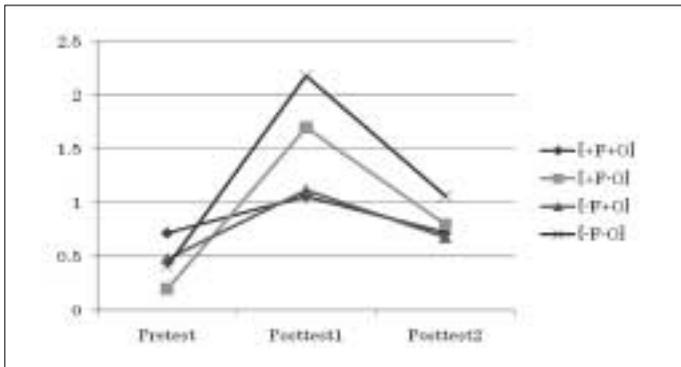
Figure 3　Mean overgeneralized errors of polysyllabic datives from production tests.

To examine the effects of Task Complexity and Task Condition, the data from Posttest 1 and Posttest 2, classified into +/−F and +/−O, were separately submitted to the Mann-Whitney U tests. The results of the Mann-Whitney U tests showed no significant differences existing between the +F and the −F in Posttest 1, $Z = −.149$, $p = .881$, nor in Posttest 2, $Z = −.270$, $p = .787$. However, with respect to the effect of Task Condition, the Mann-Whiney U tests presented that there was a significant difference between the +O and the −O in Post-test 1, $Z = −2.183$, $p < .05$, but not in Posttest 2, $Z = −1.365$, $p = .172$. These results suggest that the distribution patterns of the +O and the −O are significantly different in Posttest 1. The inspection of the histograms indicates that the −O produced more amounts of over-generalized errors compared to the +O in the Posttest 1 production test.

Discussion and Conclusion

The primary aim of the present study was to examine whether pushed output is promoted through task complexity and task condition separately or in conjunction, as suggested by Robinson (2001a, 2001b, 2003, 2005). The results of the analyses using the accuracy scores from the recognition and production tests by and large imply that pushed output was facilitated to a certain extent, at least in the short term. The accuracy scores of all the experimental groups, who received output-based treatments, improved significantly from Pretest to Posttest 1 in both tests, although such positive improvement were not sustained in the long term (14 weeks). On the other hand, the control group who received no treatment failed to show improvement; the scores remained almost consistent across the tests. Therefore, it can be concluded that the study found positive effects of tasks containing the features of task complexity and task condition in general. However, since the main effects of task complexity and task condition and their interactive effects were not found, it is not clear how these two variables had contributed to the positive improvement. It may be that distinctions between +F and −F and between +O and −O were not evident enough to produce any significant differences.

Nonetheless, in terms of errors, the study found that −O generated significantly larger amounts of overgeneralized errors than +O. It can be argued that this finding seems to indicate that −O had promoted pushed output to a large degree than +O and that the former group learned the rule of dative alternation better than the latter. Generally speaking, overgeneralized errors are generated when L2 learners incorrectly extend their current L2 knowledge to new linguistic contexts. In the case of this study, overgeneralized errors were generated when the participants applied the rule of the dative alternation, which is relevant to monosyllabic datives, to a context where polysyllabic datives were involved. That is to say, a sentence such as "Bill is describing Jim the story" was generated on the basis from a sentence such as "Bill is sending Mary a letter." Applying already-known explicit knowledge to an unfamiliar language context is the mental process, which Swain and Lapkin (1995) suggest as "thought processes of a sort which may play a role in second language learning" (p. 383). The generation of overgeneralized errors may imply that the participants engaged in syntactic analysis (e.g., hypothesis formulation and testing). If they did not pay conscious attention to the rule of dative alternation during the task and if they did not engage in any sort of linguistic analyses, overgeneralized errors would have never appeared in their output.

In this regard, it can be argued that the participants in −O might have been mentally more pushed to modify their output than those in +O. This indicates that −O facilitated the processing of language beyond the participants' current interlanguage limitations and promoted syntactic analysis to a larger extent than +O.

It must be noted that the beneficial effect of −O over +O in terms of pushed output is valid only if overgeneralized errors are viewed as a part of the restructuring process, which may result in ultimate acquisition. Given that overgeneralization of the monosyllabic double-object datives is frequently seen in the course of natural child L1 development (Mazurkewich & White, 1984; Whong-Barr & Schwartz, 2002), it is logical to assume that the production of overgeneralized errors is a part of the natural process of language development. One caveat is that such an effect lasted only for a short time since there was no significant difference between −O and +O in the amounts of overgeneralized errors in Posttest 2 production test. These results appear to imply that without long-term and consistent pedagogical interruptions, the effect of output tasks may vanish over time. Furthermore, since the effect of task condition was found only through the investigation of errors, it seems important to inspect qualitative changes directed both at accuracy and inaccuracy. With respect to the accuracy scores, significant differences between −O and +O were found neither in the recognition tests nor in the production tests. It was only when errors were carefully examined, significant differences were disclosed.

Notes

1. In contrast, Skehan (1998) argues that tasks of excessive difficulty might lead to the depletion of attentional resources necessary to be paid to linguistic forms. This view is contrary to that expressed by Robinson.
2. The rule that monosyllabic datives allow the dative alternation while polysyllabic datives do not allow such alternation cannot be applied to all verbs. In the present study, such exceptional verbs are not taken into account.
3. Some of the dativizable verbs are as follows (Pinker, 1989, p. 119): verbs of giving (give, pass), sending (send, ship, mail), instantaneous causation of motion (throw, toss), communication, illocutionary (tell, ask), creation (build, cook), obtaining (get, find, buy).
4. The data violated the distribution assumptions of parametric tests. The results of the analyses by using nonparametric techniques were interpreted in terms of significant differences in the patterns of the distribution.
5. The significant difference between Pretest and Posttest 2 revealed by the −F+O might have been attributed to the lowest mean score in the pretest production test.

References

Anderson, A., Brown, G., Shillcock, R., & Yule, G. (1984). *Teaching talk: Strategies for production and assessment*. Cambridge: Cambridge University Press.

Brown, G., & Yule, G. (1983). *Teaching the spoken language: An approach based on the analysis of conversational English*. Cambridge: Cambridge University Press.

Carroll, S., & Swain, M. (1993). Explicit and implicit negative feedback: An empirical study of the learning of linguistic generalizations. *SSLA*, 15, 357–386.

Celce-Murcia, M., & Larsen-Freeman, D. (1999). *The Grammar Book: An ESL/ EFL teacher's course*. New York: Heinle & Heinle Publishers.

Ellis, R. (1991). Grammaticality judgments and second language acquisition. *SSLA, 13*, 161–186.

Green, G.. M. (1974). *Semantics and Syntactic Regularity*. Bloomington, IN: Indiana University Press.

Hawkins, R. (1987). Markedness and the acquisition of the English dative alternation by L2 learners. *Second Language Research, 3*, 20–55.

Kowal, M., & Swain, M. (1997). From semantic to syntactic processing: How can we promote Metalinguistic awareness in the French immersion classroom? In K. Johnson & M. Swain (Eds.), *Immersion education: International perspectives* (pp. 284–309). Cambridge: Cambridge University Press.

Long, H. M. (1989). Task, group, and task-group interactions. *University of Hawaii Working Papers in ESL, 8*, 1–25.

Mazurkewich, I. (1984a). The acquisition of the dative alternation by second language learners and linguistic theory. *Language Learning, 34*, 91–109.

Mazurkewich, I. (1984b). Dative questions and markedness. In F. R. Eckman, L. H. Bell, & D. Nelson (Eds.), *Universals of SLA*. (pp. 119–131). Rowley, MA: Newbury House.

Mazurkewich, I. (1985). Syntactic markedness and language acquisition. *SSLA, 7*, 15–36.

Mazurkewich, I., & White, L. (1984). The acquisition of the dative alternation: Unlearning overgeneralizations. *Cognition, 16*, 261–283.

Pinker, S. (1989). *Learnability and cognition: the acquisition of argument structure*. Massachusetts: The MIT Press.

Robinson, P. (2001a). Task complexity, task difficulty, and task production: Exploring interactions in a componential framework. *Applied Linguistics, 22*(1), 27–57.

Robinson, P. (2001b). Task complexity, cognitive resources, and syllabus design: a triadic framework for examining task influences on SLA. In P. Robinson (Ed.), *Cognition and Second Language Instruction* (pp. 287–318). Cambridge: Cambridge University Press.

Robinson, P. (2003). Attention and Memory during SLA. In C. Doughty and M. Long (Eds.), *The handbook of second language acquisition* (pp. 631–678). Oxford: Blackwell Publishing Ltd.

Robinson, P. (2005). Cognitive complexity and task sequencing: Studies in a componential framework for second language task design. *IRAL, 43*, 1–32.

Shehadeh, A. (2002). Comprehensible Output, from occurrence to acquisition: An agenda for acquisitional research. *Language Learning, 52*(3), 597–647.

Skehan, P. (1996). A framework for the implementation of task-based instruction. *Applied Linguistics, 17*, 38–62.

Skehan, P. (1998). *A cognitive approach to language learning*. Oxford: Oxford University Press.

Swain, M. (1985). Communicative competence: some roles of comprehensible input and comprehensible output in its development. In S. Gadd & C. Madden (Eds.), *Input in SLA* (pp. 235–253). Rowley, MA: Newbury House.

Swain, M. (1991). French immersion and its offshoots: Getting two for one. In B. F. Freed (Ed.), *Foreign language acquisition research and the classroom* (pp. 91–103). Lexington, MA: D.C. Heath.

Swain, M. (1993). The output hypothesis: Just speaking and writing aren't enough. *The Canadian Modern Language Review. 50*(1), 159–164.

Swain, M. (1995). Three functions of output in second language learning. In G. Cook & B. Seidlhofer (Eds.), *Principles & practice in Applied Linguistics* (pp. 125–144). Oxford: Oxford University Press.

Swain, M. (1998). Focus on form through conscious reflection. In C. Doughty & J. Williams (Eds.), *Focus on form in classroom second language acquisition* (pp. 64–81). Cambridge: Cambridge University Press.

Swain, M. (2000). The output hypothesis and beyond: Mediating acquisition through collaborative dialogue. In C. Lantolf (Ed.), *Sociocultural theory and Second Language learning* (pp. 97–114). Oxford: Oxford University Press.

Swain, M., & Lapkin, S. (1995). Problems in output and the cognitive processes they generate: A step towards second language learning. Applied Linguistics, 16(3), 370–391.

Tarone, E., & Liu, G.–Q. (1995). Situational context, variation, and second language acquisition theory. In G. Cook & B. Seidlhoffer (Eds.), *Principles and practice in applied linguistics* (pp. 107–124). Oxford: Oxford University Press.

Whong-Barr, M., & Schwartz, D. B. (2002). Morphological and syntactic transfer in child L2 acquisition of the English dative alternation. *SSLA, 24*, 579–616.

APPENDIX A: Cartoon Strip Used at Treatment 1 (Version 1)

Strip with [+Few Elements]   Strip with [-Few Elements]

Tomoko forgot money today.

Kenji lent Tomoko some money.

So, Tomoko could send her grandmother a package.

Tomoko, Yoko, and Keiko went to the post office.

Tomoko forgot money today.

Yoko didn't have enough money.

So, Keiko lent Tomoko some money.

So, Tomoko could send her grandmother a package.

APPENDIX B: Examples of Recognition and Production Test

Recognition Test

Instruction:

Choose the most appropriate answer among four choices.

Mike (a) bought flowers Monica.
    (b) bought to Monica flowers.
    (c) bought Monica to flowers.
    (d) bought Monica flowers.

Production Test

Instruction: Answer the questions by using the assigned words. Add any word if it is necessary. If you can answer the questions in two different ways, please do so. (Write '×' if it is not possible to write in two ways.)

Q: What is Bill doing?
              (Bill / Fred / lend)
A: ①Bill is _____
                  .
  ②Bill is _____