THE JOURNAL OF ASIA TEFL Vol. 3, No. 4, pp. 193-225, Winter 2006

The Effect of Task Complexity and Language Proficiency on Task-Based Language Performance

Tomohito Ishikawa
Aoyama Gakuin University, Tokyo

This study examined the effects of task complexity and language proficiency on L2 written narrative production based on the analyses of 52 written narratives produced by Japanese high school students under different task complexity conditions. Task complexity was manipulated along the proposed task complexity dimension of Here-and-Now vs. There-and-Then in Robinson (2001a, 2006). Language proficiency was defined as scores on the Michigan English Placement Test (MEPT). In order to investigate the effects of task complexity and proficiency, four modes of production metrics were employed as dependent variables: accuracy, structural complexity, lexical complexity and fluency. The results showed that the effects of task complexity and language proficiency were largely independent; however, there were signs of interactions particularly in the cases of accuracy and lexical complexity.

THE EFFECT OF TASK COMPLEXITY AND LANGUAGE PROFICIENCY ON TASK-BASED LANGUAGE PERFORMANCE

Recent years have seen increasing interest in task-based language teaching (TBLT). Early attempts in India described for instance by Prabhu (1987) attracted attention to, Long and Crookes' (1993) influential paper paved the

way for the use of pedagogical tasks within the context of second language (L2) teaching and second language acquisition (SLA) research. In fact, recent publications on TBLT (Ellis, 2003; Garcia Mayo, 2006; Nunan, 2004; Robinson, 2001a; Skehan, 1998, to name a few) all speak volumes for the importance and potential of the use of tasks in L2 teaching, learning, and performance.

Convergence on the significance of pedagogic tasks in teaching and researching has also produced divergence on views and more questions than answers. Generally speaking, there have been issues of: the definition of "task" (e.g., Bygate, Skehan, & Swain, 2001); needs analysis (e.g., Long, 2005); assessment (e.g., Brown, Hudson, Norris, & Bonk, 2002); grading and sequencing of tasks (e.g., Long & Crookes, 1993); the effect of manipulating task complexity on learning, performance, and retention (e.g., Robinson, 2001a; Skehan, 1998); task's interactional conditions that affect the quality and quantity of meaning negotiations (e.g., Pica, Kanagy, & Falodun, 1993); the role of individual differences in proficiency, aptitude, motivation etc. (e.g., Robinson, 2005), and so forth.

The present study is an attempt to examine the effects of task complexity manipulation and language proficiency on L2 written task performance. In a broad picture, which is described below, task complexity can be viewed as pedagogic task's inherent information processing demands that can be graded, sequenced, and manipulated by teachers. In contrast, language proficiency, which have been defined in various ways in pedagogical and research contexts (see Wolfe-Quintero, Inagaki, & Kim, 1998), can be conceived as a type of individual differences that defy external manipulation. One pedagogical challenge is then how teachers can adjust their online pedagogical interventions in the face of the encounter of both. Empirical knowledge base for such a difficult pedagogical decision making is no doubt valuable. Unfortunately, there have been a few studies that have investigated relationships between task complexity manipulation and language proficiency. One of the major motivations of the present study is then to fill in a missing piece to the picture by explicitly addressing the issues of task complexity and

proficiency interactions.

ROBINSON'S TRIADIC FRAMEWORK FOR TBLT

Task Complexity, Task Difficulty, and Task Condition

Robinson (2001a, b, 2005, 2006) has argued that increasing the cognitive demands of L2 tasks leads to increases in the accuracy and complexity of L2 speech production, and also to greater learning of task input. The present study adopts Robinson's triadic framework because it is one of the most comprehensive and systematic TBLT frameworks (also see Skehan, 1998, for another framework). Especially, on the conceptual level, Robinson's framework is quite successful in distinguishing three task components: task complexity, task difficulty, and task condition.

First, task complexity refers to pedagogic tasks' information processing demands on memory, attention, and reasoning (Robinson, 2001a). This category of task dimensions is characterized as cognitive in nature, and each dimension represents inherent and relatively fixed demands of pedagogic tasks. Therefore, they are manipulable by teachers and syllabus designers prior to task performance. According to Robinson, task complexity should be the sole anchorage for the decision-making processes about task sequencing in TBLT.

In contrast, task difficulty refers to the L2 learner's "perceptions" of difficulty that are influenced by L2 learner's ability (e.g., proficiency) as well as affective variables (e.g., anxiety). Some variables within this category are subject to temporal change (e.g., motivation) in contrast to relatively stable ability variables (aptitude, intelligence, proficiency, etc.). Concerning prospective task sequencing decisions, the learner's perceptions are impossible to assess prior to task implementation, and it is for these reasons that prospective sequencing decisions that are based on task difficulty are impossible to make. However, they are extremely important indicators for online methodological

decisions in language classroom.

Thirdly, the component of task condition refers to the interactive demands of tasks, which consists of participation variables (e.g., open- vs. closed-tasks) and participant variables (e.g., gender). According to Robinson (2001b), task condition is largely constrained by the target task that is identified by needs analysis and is to be approximated by series of pedagogic tasks; therefore, whereas task condition is a crucial factor if a series of pedagogic tasks are to bear a close similarity to a target task, it does not play a central role concerning sequencing decisions. In what follows, I review the task-based literature that explicitly addresses the issues on the relationship between task complexity and language proficiency.

LITERATURE REVIEW

Despite the rapidly growing interest in TBLT, the effect of learner proficiency on task performance has been a relatively neglected area (but see Robinson, 2005). Studies so far have tended to treat language proficiency as controlling variable rather than examining its interaction with task complexity manipulation. Naturally, there have been a handful studies so far that addressed the effect of language proficiency on L2 task performance explicitly. To my limited knowledge, except for Wigglesworth (1997), who examined the effect of planning time and proficiency in a testing situation, there are only two experimental studies that investigated the effect of task complexity and proficiency in experimental studies: Kuiken, Mos, and Vedder (2005) and Kawauchi (2005).

Kuiken, Mos, and Vedder (2005) manipulated task complexity by varying the number of elements to be considered in a writing task (also see Robinson, 2001a). Specifically, they asked their participants to write a recommendation letter to a friend about where to visit for a holiday. Five destination choices were given and the participants were required to choose only one based on a varying number of criteria (i.e., three in the simple and six in the complex

task). The participants were 62 Dutch learners of Italian studying at the University of Amsterdam. They were divided into two proficiency groups based on their cloze test scores and the amount of learning experience. Kuiken, et al. (2005) used three categories of L2 production measures: syntactic complexity; lexical variation; accuracy.

Their results showed that there were no task complexity effects on lexical and syntactic complexity although low-proficiency group's linguistic complexity showed a trend. In contrast, analyses on accuracy data yielded significant interactions between task complexity and proficiency; namely, greater written accuracy was observed when task complexity and proficiency were both high. The low proficiency group was generally unaffected by the difference in task complexity. Kuiken, et al. (2005) interpreted these findings as follows: (a) the significant task complexity effect on accuracy on the part of the high proficiency learners was taken as support for Robinson's prediction that greater task demands lead to greater accuracy (Robinson, 2001a); (b) the lack of task complexity effect on the part of the low proficiency learners was ascribed to the existence of a threshold level, beyond which task complexity effects could be observed.

Kawauchi (2005) investigated the effect of strategic planning and language proficiency on L2 oral narrative production by Japanese college students. The participants of her study constituted three different proficiency groups: Low EFL (TOEFL 420-480); High EFL (TOEFL 510-580); Advanced ESL (TOEFL 550-610). Using a within-subject experimental design, she compared L2 oral narrative production under unplanned and planned conditions. Analyses were conducted using four categories of production measures: accuracy; structural complexity; lexical variation; fluency.

The main findings of her study were that regarding structural complexity and lexical variation, High EFL learners received the greatest benefits, whereas Low EFL learners gained the most in accuracy terms. Her interpretations of the findings were that (a) the lack of strategic planning effects of Advanced ESL learners was due to ceiling effects; (b) Low EFL learners were able to direct attention to language form thanks to strategic

planning activities; (c) High EFL learners received the greatest benefits because they had available L2 knowledge, which became accessible under planning conditions.

To conclude this section, because the studies conducted differ in a number of respects, direct comparisons are difficult to make. However, they do offer some insights into the results obtained on the present study as we will see in the Discussion section.

RESEARCH QUESTIONS

The current study posed the following research question: *What are the effects of manipulating task complexity and language proficiency on L2 written task performance?* To answer the research question, the factors of task complexity and language proficiency were used as independent variables (hereafter "Task Complexity" and "Proficiency Group"). Each independent variable had two levels (Hear-and-Now vs. There-and-Then, which are described below; low- vs. high-proficiency groups). Given the lack of preceding studies and cross-study differences in task complexity dimensions manipulated, and in L2 production measures employed, the present study does not formulate any directional hypotheses.

THE PRESENT STUDY

Task demands

Based on the triadic framework, the present study operationalized one of the proposed "task complexity" dimensions: Here-and-Now (HN) versus There-and-Then (TT). In the HN condition, participants, after viewing the cartoon for five minutes, wrote narratives for 30 minutes based on a prompt written in the present tense. In writing narratives, they were allowed to view

the cartoon strip. On the other hand, in the TT condition, participants, after viewing the cartoon strip for five minutes, were asked to write narratives without viewing the cartoon strip for 30 minutes based on a prompt written in the past tense. In the sections to follow, the cognitive demands of the present study's tasks and their potential effects on L2 written discourse are considered in three realms: absence of shared contexts, memory demands, and attentional demands.

Absence of shared contexts. Unlike previous studies that investigated the effect of manipulating the HN-versus-TT dimension, the current research does not feature a shared context in both conditions. In the preceding oral production studies (e.g., Gilabert, 2006; Rahimpour, 1999; Robinson, 1995), the listener was present in both HN and TT conditions; therefore, the context was shared between the speaker and the listener. In both conditions of the present study, and due to the nature of the writing task, the reader is absent and the context is not shared. It is therefore assumed that the communicative demands of both conditions are more demanding than those of the preceding studies.

The absence of shared contexts may direct learner's attention to accuracy of article use (Robinson, 1995). In narratives, a canonical use of articles is introducing a participant and subsequent mentions to the same participant (e.g., Celce-Murcia & Larsen-Freeman, 1999). Once a participant is introduced, the reference to the same participant needs to be established so that the reader can retrieve the entity mentioned in previous discourse from working memory (Ariel, 1990).

Memory demands. Another way to characterize cognitive task demands of the current study is in memorial terms. The HN and TT conditions seem to impose differential memory demands, which may in turn affect the nature of information processing. For example, because learners in the TT condition need to memorize and retrieve the storyline and details, and subsequently produce a coherent narrative, they may be pushed to ruminate on the storyline, to infer the relationships between events, and to create larger informational chunks to facilitate memory encoding, storage and retrieval (Robinson, 1995).

This is similar to Bartlett's (1932) conception of "effort after meaning," which helps to establish elaborated semantic representations prior to task performance (Dechert, 1987; Levelt, 1989). Thus, task demands in the TT condition may encourage deeper semantic processing than those in the HN condition, which may establish more elaborated output plans, out of which more complex language can be unpacked.

Such effects of elaborative conceptualization may be captured by the use of syntactic complexity measures (e.g., S-nodes per T-unit). They may capture the assumed deeper semantic processing in terms of subordinating and embedding clauses, which are associated with the discourse structure of narratives and with "the construction of higher-order events in which event phases are subordinated and interrelated" (Berman & Slobin, 1994, p. 13).

Another type of production measures that may capture some aspect of "effort after meaning" is lexical variation measures. Deeper semantic elaboration may motivate the learner to explain, assess, predict, or interpret events in a narrative in addition to simply describing the main storyline. If so, deeper semantic processing may increase the chance to use a wider range of lexical items.

Crucially, whether such deeper conceptual planning can be done successfully is claimed to hinge on plannability of discourse (e.g., written vs. spoken discourse, Ochs, 1979), including availability of planning time (Crookes, 1989). Although macro-planning imposes attentional demands (Levelt, 1989), given the nature of writing as opposed to speaking and the permission of planning time (i.e., five minutes), the current study's processing conditions may offer sufficient opportunities to make the discourse more plannable. Furthermore, whether or not such conceptual plans and their associated linguistic codes can be accessed during task performance also depends on another processing constraint; namely, available attentional resources.

Attentional resource availability. As suggested above, the realization of production plans as a result of semantic elaboration may be affected by attentional demands during task performance. For example, in the HN

condition, the learner needs to produce written output as he or she monitors the written output and the visually presented material. Under such conditions, the learner is likely to be required to shift attention from the written output to the strip cartoon then to writing and so forth. Furthermore, this type of attention shifting requires overt eye movements from the writing sheet to the cartoon and vice versa (see Ellis, 2005a; Robinson, 2003; Wickens & Hollans, 2000). Hence, it can be expected that attentional resources available to learners are more difficult to co-ordinate in the HN condition. On the other hand, learners in the TT condition may be better able to co-ordinate attention resources; hence, fluency may be greater. Furthermore, freeing up attentional resources are claimed to boost the effect of directing attentional resources to the predicted code features (Robinson, 2001a); therefore, attention to accuracy (target-like use of articles) as well as syntactic and lexical complexity may be benefited more in the TT.

Based on the characteristics of the current study's task demands, the absence of the cartoon strip in the TT condition may entail: (a) absence of shared context, which directs attention to accuracy; (b) greater memory demands, which encourages "effort after meaning," and therefore facilitates greater use of complex language and various linguistic items; (c) differences in the co-ordination of available attentional resources, which may enhance fluency and facilitate attention to accuracy and complexity.

Production measures

Motivated by the above argument for the hypothesized task demands on L2 written production, the present study used four production measures that were thought to reflect the assumed task demands based on Wolfe-Quintero, Inagaki, and Kim's (1998) proposals concerning accuracy, structural complexity, lexical complexity and fluency.

First, in order to capture the hypothesized increased attention to the use of English articles, a production measure of target-like use articles was employed. Spelling mistakes concerning the distinction between *a* and *an*

were ignored in coding the data.

Regarding structural complexity, following the previous HN-versus-TT studies of Gilabert (2006), Rahimpour (1999), and Robinson (1995), a measure of S-nodes per T-unit was adopted. An S-node is equivalent to a verb phrase (VP) (both finite and infinite) (also see Wolfe-Quintero, et al., 1998). If the learners in the TT condition are pushed to connect and motivate events temporally, causally, and evaluatively and to pack more information into a verbal unit, S-nodes per T-unit should be sensitive to such task-induced changes in learners' linguistic products. In the present study, a T-unit was defined as a main clause plus any subordinating clauses (Hunt, 1965) and it could occur across periods. In addition, sentence fragments were not counted as T-units (Hirano, 1991).

The present study also used a lexical variation measure in order to capture the lexical influence of "effort after meaning." The measure was "word types squared divided by the total number of words" (WT^2/W), a type-token measure described by Wolfe-Quintero et al. (1998). Type-token measures have been criticized because they are affected by text length. The measure used in the present study "may not be perfect" (Wolfe-Quintero et al., 1998, p. 108), but does take into account the effect of text length. For the sake of simplicity, hereafter I refer to WT^2/W as type-token ratio.

Finally, the present study employed words per T-unit (or T-unit length) as writing fluency measure, which is one of the best measures of writing fluency (Wolfe-Quintero, et al., 1998). Traditionally, length measures are considered as measures of linguistic complexity (e.g., Ortega, 2000). However, as Wolfe-Quintero, et al. (1998) claim, length measures do not reflect complexity itself because there are other ways to increase the length of a production measure. Furthermore, words per T-unit may reflect the degree of automaticity in accessing the lexicon unselectively because the measure is free from specific syntactic categories. For these reasons, the present study defined words per T-unit as fluency measure. In coding the data, hyphenated words were counted as one word. In addition, if a learner wrote *shoppingcart* for *shopping cart*, such instances were counted as two words.

Data coding was conducted by two native speakers of English with MAs in TESOL and me. The inter-rater reliabilities reached above .90. The CHILDES System (MacWhinney, 2000) was used for computing figures. In the section to follow, I will review empirical studies that examined the effects of task complexity and proficiency on L2 task performance.

Participants

The present study is based on written narrative data collected under the HN and TT conditions. Originally, 54 participants (Japanese high school students aged 17-18) were randomly assigned to the HN and TT conditions (T. Ishikawa, 2005, 2006). The present study, however, is based on 52 participants' data because high- and low-proficiency groups were formed in the HN and TT conditions respectively.

Operationalization of Proficiency Groups

Based on the scores of the Michigan English Placement Test (MEPT) Form B (KR21= .81), proficiency-based grouping was conducted. First, a median score on the MEPT was identified in the HN and TT conditions respectively. When more than one case fell on an identified median score, based on randomization procedures, both low- and high-proficiency groups were formed in each task complexity condition (four groups of 13 participants). The resultative ranges of the MEPT scores of each group were as follows: low-proficiency groups in the HN (50-65) and TT (50-60); high-proficiency groups in the HN (67-88) and TT (65-84).

TABLE 1
Descriptive Statistics for the Michigan Placement Test of the Experimental Groups

Task complexity	Low proficiency *M/SD*	High proficiency *M/SD*
Here-and-Now	56.00/4.83	74.77/6.57
There-and-Then	54.87/3.16	74.15/6.66

Table 1 presents the descriptive statistics for the means and standard deviations of the four groups on the MEPT (maximum = 100). A one-way ANOVA was applied to the data set and group mean differences on the MEPT were examined. In the analysis, the factor of Experimental Group (with four group levels) was used as independent variable and the MEPT scores were used as dependent variable.

The results showed that there was a significant main effect for Experimental Group ($F(3, 48) = 52.13$, $p = .0001$). Subsequently, Schaffe's post hoc test was applied. The results showed that all comparisons of the MEPT group means were significant ($p = .0001$) except for the two non-significant group mean differences: (a) one between low-proficiency groups in the HN and TT conditions respectively and (b) the other between high-proficiency groups in the HN and TT conditions respectively ($p >. 90$). In other words, the two low-proficiency groups did not differ from each other on the MEPT group mean scores; the same was true of the group mean difference between the high-proficiency groups. Hereafter, I will refer to the low-proficiency group in the HN condition as "Low-HN"; the high-proficiency group as "High-HN"; the low-proficiency group in the TT condition as "Low-TT"; and the high-proficiency group in the TT condition as "High-TT."

Procedures

Before the experiment, the participants were informed that the writing task would not be considered as part of their school grades and that they would be

given extra points for participating in the experiment. They were also informed that the writing samples would be used for research purposes.

During the experiment, the participants were first given a paper on which instructions for the task were written in their L1 (Japanese). The content of the directions was as follows:

(a) you are allowed to view a strip cartoon for five minutes;
(b) you may take notes (only words and/or phrases) during the cartoon-viewing session but the notes will be removed before writing (see Crookes, 1989);
(c) you are not allowed to name the characters in the strip cartoon (see S. Ishikawa (1995) for a similar treatment to elicit articles);
(d) you will be required to write a narrative

Furthermore, the participants in the HN condition were told that the strip cartoon would be available for them to view while writing, whereas those in the TT condition were told that the strip cartoon would be removed.

After reading the directions, the participants received prompts (the one for the HN condition was written in the present tense and the one for the TT condition was written in the past tense, see Appendix A) and a strip cartoon "the supermarket" from Yule (1997) (Appendix B). They viewed the strip cartoon for five minutes before writing. In addition, the participants in both conditions were given four words paired with their Japanese translation equivalents for them to avoid basic vocabulary problems; however, they were not required to use the words. Those words were simply presented as options. They were *shopping cart*, *bottle*, *checkout counter*, and *shelf*. After the cartoon-viewing session, the directions sheet and notes were removed in both groups and the participants were given new writing sheets. All the participants completed narratives within 30 minutes except for one participant. A majority of the participants finished writing narratives between 25-30 minutes. Sample narratives from both conditions are given in Appendix C.

RESULTS

Table 3 shows the descriptive statistics for the four L2 production measures of the four groups. Because the assumptions of multivariate normality and homogeneity of the covariance matrix were not violated, a multivariate analysis of variance (MANOVA) was applied to the data set, with Proficiency Group (with two levels of high- and low-proficiency groups) and Task Complexity (with two levels of the HN and TT conditions) as independent variables and the four production measures as dependent variables. The MANOVA showed that Task Complexity was significant (Λ = .75, F(4, 45) = 3.67, p = .011, partial η^2 = .25); Proficiency Group significant (Λ = .66, F(4, 45) = 5.92, p = .001, partial η^2 = .35); the interaction between the two independent variables non-significant (Λ = .86, F(4, 45) = 1.82, p = .142, partial η^2 = .14). Partly because the selection of the dependent variables was exploratory in nature, subsequent analyses were made using univariate analyses of variance (ANOVAs).

TABLE 2
Descriptive Statistics for the L2 Production Measures of the Experimental Groups

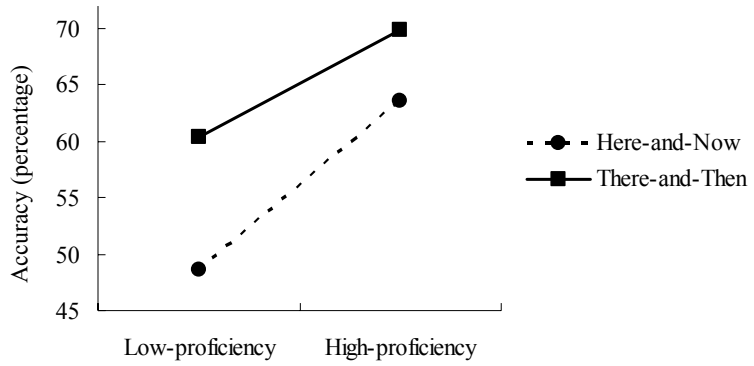
	Here-and-Now		There-and-Then	
	Low-proficiency	High-proficiency	Low-proficiency	High-proficiency
Production metrics	*M/SD*	*M/SD*	*M/SD*	*M/SD*
TLU articles	48.63/13.98	63.63/14.94	60.35/13.49	69.80/19.42
S-nodes per T-unit	1.49/.30	1.69/.30	1.70/.27	1.95/.37
Type-token ratio	36.06/8.21	49.51/10.41	46.62/10.88	47.15/8.2
Words per T-unit	8.16/1.61	9.67/1.57	9.21/1.70	10.76/2.0

Accuracy: Target-like Use of English Articles

Regarding the effect of proficiency on target-like use of English article, ANOVA results showed that Task Complexity was significant (F(1, 48) = 4.26, p = .044, partial η^2 = .081); Proficiency Group significant (F(1, 48) =

7.96, p = .007, partial η^2 = .142); the interaction non-significant (F(1, 48) = 4.10, p = .525, partial η^2 = .008). Thus, the effects of task complexity and proficiency were both additive. Schaffe's post hoc tests detected significant mean differences between the Low-HN and the High-TT groups (p < .01) and between the Low-HN and High-TT groups (p = .018). Figure 1 shows the means of the four groups on the target-like use of English articles. In view of the preceding results, it seems that both task complexity and proficiency influenced the learner's attention allocation independently. However, of crucial importance was that whereas the interaction was not statistically significant, even when task complexity was increased, the High-TT group showed a modest increase of 6.17 in percent accuracy. This contrasted with the group mean difference of 11.72 between the Low-HN and the Low-TT groups.

FIGURE 1
Group Means of Target-like Use of English Articles as a Function of Task Complexity and Proficiency



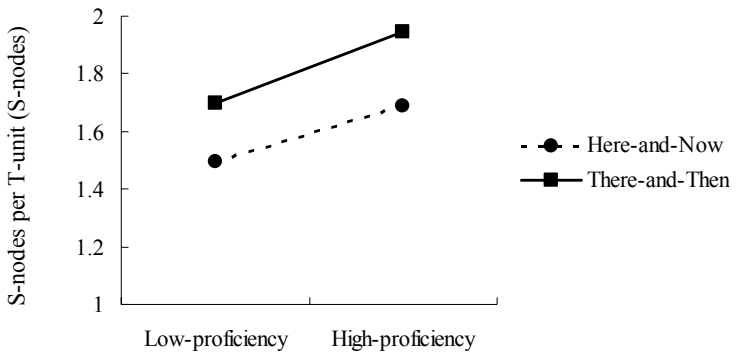
There are two possible explanations for the differential effect. One explanation is that this may be suggestive of the potential interaction between task complexity, proficiency, and the inherent rule complexity (e.g., DeKeyser, 2003; Robinson, 1996; Samuda, 2001). Thus, task complexity

pushes the learner to produce more accurate interlanguage form but such task-induced accuracy-facilitating effects may be dependent on learners' proficiency and/or inherent rule complexity. Alternatively, the result may reflect a simple fact that as the learner's proficiency becomes higher, the leeway for increasing accuracy is also minimized as Kawauchi (2005) observed in her study. Clearly, both explanations are possible and are not mutually exclusive. In view of these interactional possibilities and the greater accuracy improvement on the part of the Low-TT group (approximately 12%), it may be wiser to place a practical emphasis on the finding and conclude that there were some indications that the degree of the effect of increasing task complexity may be dependent on the learner's proficiency level and possibly on inherent rule complexity under investigation.

Structural Complexity: S-nodes per T-unit

Regarding S-nodes per T-unit, the results of a univariate ANOVA showed that Task Complexity was significant ($F(1, 48) = 7.19$, $p = .010$, partial $\eta^2 = .130$); Proficiency Group also significant ($F(1, 48) = 6.58$, $p = .013$, partial $\eta^2 = .121$); the interaction non-significant ($F(1, 48) = .11$, $p = .741$, partial $\eta^2 = .002$). Subsequent Schaffe's post hoc tests produced a single significant mean difference between the Low-HN and High-TT groups ($p < .01$). Figure 2 shows the four group means of S-nodes per T-unit. The overall pattern was that more proficient learners' generally produced complex language and manipulating task complexity also affected the use of complex language independently and positively. The latter effect may be ascribable to the increased conceptual activation during the output planning stage, or what Berman and Slobin (1994) call "relating events in narrative." In other words, functional complexity accompanied structural complexity (Givon, 1985; Robinson, 1995). Of particular interest is the result that even the High-HN group (S-nodes per T-unit = 1.69) was overhauled by the Low-TT group (1.70).

Figure 2
Group Means of S-nodes per T-unit as a Function of Task Complexity and Proficiency

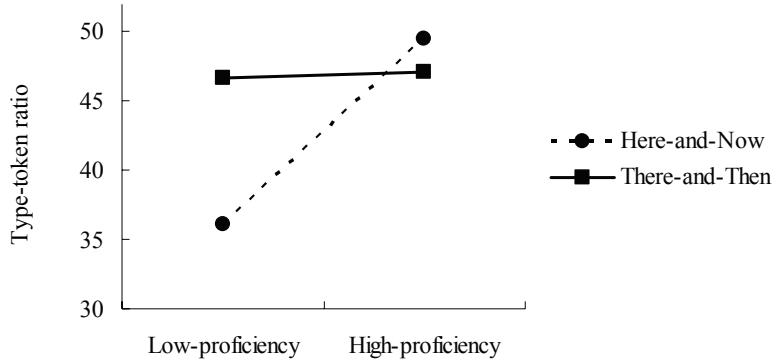


Lexical Complexity: Type-token Ratio

The ANOVA results of the type-token ration produced a somewhat different picture. Thus, task complexity was not significant ($F(1, 48) = 2.41$, $p = .127$, partial $\eta^2 = .048$), but proficiency group was significant ($F(1, 48) = 7.00$, $p = .011$, partial $\eta^2 = .127$). More importantly, there was a significant interaction between task complexity and proficiency group ($F(1, 48) = 5.99$, $p = .018$, partial $\eta^2 = .111$). This means that the effect of proficiency on lexical complexity or variation was dependent on task complexity. Figure 3 makes the nature of the interaction clearer (higher values on the scale indicate more lexical variation). Except for the lowest mean type-token ratio of 36.1 of the Low-HN group, the three other groups attained almost equivalent mean type-token ratios: 46.6 of the Low-TT group; 49.5 of the High-HN group; 47.1 of the High-TT group). In fact, Schaffe's post hoc tests showed that the lowest mean type-token ratio of the Low-HN group was significantly different from the High-HN group ($p = .0089$) and from the High-TT group ($p = .043$), and near-significantly different from the Low-TT group ($p = .058$). Thus, descriptively, the results indicated that as proficiency increases, the

effect of manipulating task complexity on lexical variation disappeared or was overridden by the learner's proficiency level. This in turn implies that the low-proficiency learners received greater benefits when task complexity was manipulated.

FIGURE 3
Group Means of Type-token Ratio as a Function of Task Complexity and Proficiency



In order to confirm that the finding of the absence of task-induced lexical variation effects on the High-TT group were not attributable to the conceptual limit of the writing material itself (i.e., use of the picture cartoon), I also asked a 17-year old female returnee, who spent 12 years in the United States and is fluent both in Japanese and English, to write a narrative based on the same strip cartoon in the TT condition. Within 30 minutes, she produced 336 words (or tokens) with 176 word types. Her type-token ration as defined in the present study was approximately 92 (for the other measures, target-like use of articles = 100%; S-nodes per T-unit = 2.95; words per T-unit = 16.8). The results thus indicated that a higher type-token ratio was attainable even when the same strip cartoon was used.

The fact that the lack of task complexity effects on the High-TT group's lexical variation was not likely to be ascribable to the material's conceptual

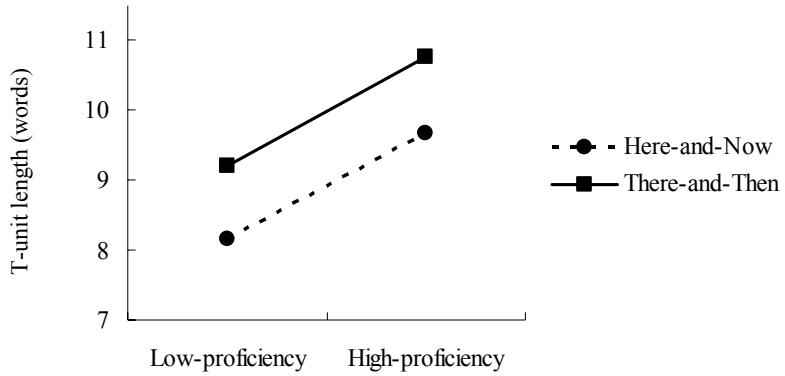
limit calls for another explanation. A more plausible account is that because L2 vocabulary repertoire of the learners' in the High-TT group was relatively restricted compared to the female near-native speaker's, certain vocabulary items occurred with excessive frequency even when conceptual representations were activated in the TT condition. In fact, a great majority of the participants neither used the words *chat*/*chatting* nor *converse*/*conversation*; they almost unanimously relied on *talk*/*talking*. The same held in the frequent reliance on basic words such as *take*, *put*, or *know*. The female returnee's narrative was characterized by additional layers of vocabulary (e.g., *grab*, *replace*, *insist*, or *discover* to name a few). Without doubt the participants of the High-TT group "knew" such words; however, when it came to production those words did not seem to constitute their productive vocabulary, which may have created limiting conditions for the High-TT group's lexical variation.

Fluency: Words per T-unit

Finally, with respect to the fluency measure, words per T-unit, a two-way ANOVA showed that the effect of task complexity was significant ($F(1, 48) = 4.45$, $p = .040$, partial $\eta^2 = .085$); the effect of proficiency group was also significant ($F(1, 48) = 9.35$, $p = .004$, partial $\eta^2 = .163$). The interaction between task complexity and proficiency group was non-significant ($F(1, 48) = .003$, $p = .96$, partial $\eta^2 < .001$). Thus, the effects of task complexity and proficiency group were both additive. Schffe's post hoc tests showed that there was a single significant mean difference; namely, the High-TT group differed from the Low-HN group (a difference of 2.5 words per T-unit). Figure 4 presents the group means of words per T-unit. Figure 4 shows that increasing task complexity led to an increase in approximately one word per T-unit. This effect of manipulating task complexity along the HN-and-TT dimension on T-unit length may reflect the differential degrees of attentional co-ordination in both task conditions. The effect of proficiency, on the other hand, was approximately an increase of 1.5 words per T-unit. This can be attributable to proficiency-related differences in the degree of automaticity in

general plus accessibility to ready-made chunks (Ellis & Barkhuizen, 2005; Kormos, 2006). In principle, the production metric of words per T-unit is global in nature in that the measure does not necessarily reflect selective access to particular word classes and/or ready-made chunks. Accordingly, the effects of task complexity and proficiency observed in the present study should also be viewed as global in nature.

FIGURE 4
Group Means of Words per T-unit as a Function of Task Complexity and Proficiency



SUMMARY

In examining the effects of task complexity and proficiency, the present study employed four types of production metrics of accuracy, structural complexity, lexical complexity and fluency. The overall patterns were rather clear-cut. Thus, the effects of task complexity and language proficiency were generally independent and therefore additive. When we took a closer look at the data, however, whereas the effects of task complexity and proficiency were independent in both structural complexity and fluency, there were some indications of interactions between task complexity and proficiency in

accuracy and lexical complexity. The interactions were clearer when we considered practical significance of the findings. I will discuss potential reasons for the observed interactions in the subsequent section.

DISCUSSION

The results of the present study then showed that the low-proficiency learners seemed to receive greater benefits when task complexity was manipulated from HN to TT. Of particular significance was the consistent results that the four aspects of the low-proficiency learners' performance in the TT condition were by no mean inferior to the high-proficiency learners' in the HN condition. Furthermore, the results of the analysis of target-like use of English articles showed that the low-proficiency learners received approximately doubled benefits compared to the high-proficiency learners when task complexity was manipulated from HN to TT (approximate growths of 12 % vs. 6 %). Such favorable effects of task complexity manipulation on low-proficiency learners were particularly evident in the case of lexical variation, where the mean of the low-proficiency group was almost equal to the means of the high-proficiency groups in both task complexity conditions.

These results, however, may be due to complex interplays among various factors. There remain possibilities that L2 production is influenced by inherent rule complexity (e.g., DeKeyser, 2003; Robinson, 1996; Samuda, 2001). Some rules are more or less complex and subtle that they may elude efficient hypothesis formation and testing. It is not surprising therefore to observe some developmental suspensions for learners of particular proficiency levels. In the case of the present study, target-like use of English articles may be classified as one of such complex rules. Whether or not interactions between task complexity, proficiency, and rule complexity exist is an interesting issue to be addressed in future studies.

There seemed to be ceiling effects as well in the case of target-like use of

English articles. Kawauchi (2005), for instance, observed that in the case of past tense accuracy, her learners did not show improvements at around the accuracy ranges of 70-80% in the case of her Advanced ESL learners. It should be noted that whether there really exist such a ceiling requires empirical investigations through the introduction of other methodological options or other task implementation aspects are needed to be introduced: task repetition (e.g., Bygate, 2001, or [+ prior knowledge] in Robinson's term) or practice (e.g., DeKeyser, 2001); use of various modes of focus on form (FonF) such as proactive FonF techniques (e.g., visual-input enhancement of task prompts), reactive FonF techniques (e.g., provision of feedback) (e.g., Doughty & Williams, 1998; Muranoi, 2000), or increasing awareness through collaborative learning (e.g., Kowal & Swain, 1994). Such methodological options may help the learner to pay attention to form, which is motivated by a semantic need during task performance (Samuda, 2001) although learner individual differences still stay as important variables that affect the results (e.g., Robinson, 2002, 2005). As Bygate (2001) emphasizes, L2 learners are required to integrate form-meaning connections under real-time pressure, and these additional methodological options and task complexity manipulation go hand in hand to help learners to take advantage of noticing opportunities (Long & Robinson, 1998).

The final issue concerns the findings of lexical variation. The results of the current study presented remarkable task complexity effects only on low-proficiency learners under the TT condition. I ascribed the lack of task complexity effects on lexical variation of the High-TT group learners to their relatively restricted vocabulary repertoires and its consequence of relatively excessive frequency of certain lexical items compared with the returnee student. An interesting question is whether or to what extent the learners of the High-TT group were aware of the need to develop their vocabulary knowledge compared with the learners of the other groups. In fact, the learners in the High-TT group showed the most accurate, syntactically complex, and fluent performance; nonetheless, the lexical variation showed underperformance.

A more positive interpretation is possible; that is, the learners in the High-TT group may have been more ready to take up linguistic forms if they were introduced as options prior to task performance. Uptake of linguistic expressions (Robinson, 2004) is clearly associated with the time-honored notion of "islands of reliability" (Dechert, 1984) and is conceptually mediated if Potter and Lombardi's (1990) Regeneration Hypothesis (i.e., even a simple repetition of an utterance is conceptually mediated rather than mechanically being reproduced devoid of conceptual activation) is also applicable to L2 production. Such taken-up linguistic items are not task-essential but useful or "helpful" as options (Loschky & Bley-Vroman, 1993; Robinson, 2004; Smuda, 2001) and are predicted to be retained longer under greater task demands (e.g., Robinson, 2001a; Schneider, Healy, & Bourne, 2002) or strongly perceived to be needed by the learner (e.g., Laufer & Hulstijn, 2001).

In the context of TBLT, Robinson (2004) reports increased learner uptake under more cognitively demanding tasks that require complex reasoning. In the study, participants were asked to sequence a series of pictures and describe the story that imposed varying degrees of reasoning demands (low, intermediate, and high reasoning demands based on the picture arrangement subtest of the Wechsler Adult Intelligence Scale, Revised Version). The tasks were one-way, closed task; so, the information receiver was also required to sequence the pictures based on the information giver's oral narration but was allowed to ask question to the information giver. Within this task performance context, six English phrases ("pre-modified input" in Robinson's term) were presented to the information giver only for each task, which were controlled in structural terms across the three tasks and given in written form on the information giver's picture prompts (e.g., *is carrying a plank* for the simple task; *is hailing a taxi* for the complex task). Those task-relevant English phrases were presented paired with the participants' L1 translations (Japanese), so that the meanings were clear to the information givers. Uptake was coded in two ways: partial uptake (use of content words from the given expressions) or exact uptake (learners' unaltered use of the given

expressions). The results showed that learners significantly increased partial uptake (e.g., uses of *plan* or *hailing*) as the task complexity increased both in total number of partial uptake per task and partial uptake per turn.

Kawauchi (2005, pp. 162-164) also reports her participants' uptake of L2 input (i.e., low frequency lexical items and problematic structural items) from reading texts given prior to task performance (or during strategic planning activities) (also see Raupach, 1984, p. 120, for a similar finding). According to her, Low EFL learners (TOEFL 420-480) took advantage of uptake opportunities, which may have been driven by noticing the hole/gap on the part of learners (e.g., Doughty & Williams, 1998; Izumi, 2003; T. Ishikawa, in press). It may also be possible, especially during writing, for L2 learners to notice the lack of variation, appropriateness, or effectiveness. It is important that in the cases of Kawauchi (2005), Raupach (1984), and Robinson (2004), linguistic expressions given in reading prompts or task sheets were task-relevant. As the learner's proficiency level goes up, different levels of linguistic expressions may be perceived and valued as islands of reliability and they would become likely targets to be incorporated into the learner's linguistic system.

PEDAGOGICAL IMPLICATIONS

The results of the present study have some pedagogical implications. The present study presented additional evidence for the view that task complexity manipulation is a useful form of pedagogical practice in motivating the learner to produce more advanced forms of their L2 (e.g., Long & Crookes, 1993; Robinson, 2006). According to Robinson (2001a), the effect of task complexity on learner language is in part attributable to the nature of conceptualization in performing pedagogic tasks (i.e., resource-directing effects). Pedagogical tasks can be made more complex or simpler; however, the crucial point here is that the connections between task complexity or conceptualization and the learner language should not be arbitrary. This

means that simply manipulating task complexity and assessing the learner language in arbitrary ways could lead to unfairness of assessment (for instance, the assessment of the target-like use of phrasal verbs in relation to the manipulating of the HN-and-TT dimension is perhaps inappropriate if not specified).

Pushing the learner to produce more advanced forms of the learner language is at the same time a challenge for the learner. Some of the limiting conditions are already discussed elsewhere in this article (e.g., L2 knowledge deficits, rule complexity etc.) and promising methodological options and possible ways to optimize task complexity without necessarily reducing tasks' conceptual demands are proposed and examined in the literature (e.g., Doughty & Williams, 1998; Egusa & Yokoyama, 2004; Ellis, 2005b; Robinson, 2001a, 2005, 2006).

CONCLUSION

Despite a number of limitations (e.g., small sample size, single task use etc.), the present study was able to take a small step toward the investigation of the effects of task complexity and language proficiency on L2 performance. Clearly, there are more questions than answers, which is just another way of emphasizing the importance and potential of task-based language learning and performance.

ACKNOWLEDGMENT

I thank Professor Peter Robinson, my mentor at Aoyama Gakuin University, for the considerable academic support he has given me. Thanks are also due to Professor Tomoko Takahashi, my former MA advisor at Soka University of America, Bobert Agajeenian, Jonathan Epstein, Terrie Mathis, my former M.A. teachers at Soka University of America, for their highest

encouragement. I also thank Aiwa Akagi, a retired professor at Soka University, Japan, for his encouragement. All errors remain my responsibility.

THE AUTHOR

Tomohito Ishikawa is a Ph.D. candidate at Aoyama Gakuin University, Tokyo. He is also a part-time lecturer at Soka Women's College and Soka University, Tokyo. His research interests cover task-based language learning and teaching. His publications include *Investigating the relationship between structural complexity indices of EFL writing and language proficiency: A task-based approach* with T. Ishikawa (2005) and *The effect of increasing task complexity along the [±Here-and-Now] dimension on L2 written narrative discourse* with T. Ishikawa (2006).

REFERENCES

Ariel, M. (1990). *Accessing noun-phrase antecedents*. London: Routledge.

Bartlett, F. C. (1932). *Remembering: An experimental and social study.* Cambridge: Cambridge University Press.

Berman, R. A., & Slobin, D. I. (1994*). Relating events in narrative: A crosslinguistic developmental study.* Hillsdale, NJ: Erlbaum.

Brown, J. D., Hudson, T., Norris, J., & Bonk, W. (2002). *An investigation of second language task-based performance assessments.* Second Language Teaching & Curriculum Center, University of Hawaii.

Bygate, M. (1996). Effects of task repetition: Appraising the developing language learners. In J. Willis & D. Willis (Eds.), *Challenge and change in language teaching* (pp. 136-148). Oxford: Macmillan.

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, Essex: Pearson.

Bygate, M., Skehan, P., & Swain, M. (2001) Introduction. In M. Bygate, P. Skehan, & M. Swain (Eds.), *Researching pedagogic tasks: Second language learning,*

teaching and testing (pp. 1-20). Harlow, Essex: Pearson.

Celce-Murcia M., & Larsen-Freeman, D. (1999). *The grammar book: An ESL/EFL teacher' course* (2nd ed.). Boston: Heinle & Heinle.

Crookes, G. (1989) Planning and interlanguage variation. *Studies in Second Language Acquisition, 11,* 367-383.

Dechert, H. (1984). Second language production: Six hypotheses. In H. Dechert, D. Möhle & M. Raupach (Eds.), *Second language productions* (pp. 211-230). Tübingen: Gunter Narr.

Dechert, H. (1987). Understanding producing. In H. Dechert & M. Raupach (Eds.), *Psycholinguistic models of production* (pp. 229-237). Norwood, NJ: Ablex.

DeKeyser, R. (2001). Automaticity and automatization. In P. Robinson (Eds.), *Cognition and second language instruction* (pp. 125-151). Cambridge: Cambridge University Press.

DeKeyser, R. (2003). Implicit and explicit learning. In C. Doughty & M. Long (Eds.), *The handbook of second language acquisition* (pp. 313-348). Malden, MA: Blackwell.

Doughty, C., & Williams, J. (Eds.) (1998). *Focus on form in classroom second language acquisition.* Cambridge: Cambridge University Press.

Egusa, C., & Yokoyama, Y. (2004). The effects of task types on second language speech production among Japanese university students: Fluency, accuracy, complexity and trade-off effects. *Annual Review of English Language Education in Japan, 15,* 129-138.

Ellis, R. (2003). *Task-based language learning and teaching.* Oxford: Oxford University Press.

Ellis, R. (Ed.) (2005a). *Planning and task performance in a second language.* Amsterdam: Benjamins.

Ellis, R., & Barkhuizen, G. (2005). *Analyzing learner language.* Oxford: Oxford University Press.

Garcia Mayo, P. (Ed.) (2006). *Investigating tasks in formal language learning.* Philadelphia, PA: Multilingual Matters.

Gilabert, R. (2006). The simultaneous manipulation along the planning time and +/- Here-and-Now dimensions: Effects on oral L2 production. In P. Garcia Mayo (Ed.), *Investigating tasks in formal language learning* (pp. 44-68). Philadelphia, PA: Multilingual Matters.

Givon, T. (1985). Function, structure, and language acquisition. In D. Slobin (Ed.), *The crosslinguistic study of language acquisition: Vol. 1* (pp. 1008-1025). Hillsdale, NJ. Lawrence Erlbaum.

Hirano, K. (1991). The effect of audience on the efficacy of objective measures of

EFL proficiency in Japanese university students. *Annual Review of English Language Education in Japan, 2,* 21-30.

Hunt, W. (1965). *Grammatical structures written at three grade levels.* Urbana, IL: The National Council of Teachers of English.

Ishikawa, S. (1995). Objective measurement of low-proficiency EFL narrative writing. *Journal of Second Language Writing, 4,* 51-69.

Ishikawa, T. (2005). Investigating the relationship between structural complexity indices of EFL writing and language proficiency: A task-based approach [EFL writing ni okeru kouzoutekifukuzatsusa no hattatsusihyouto fukuzatsusano kannkeino kenshou: task ni motozuku approach]. *JACET Bulletin, 41,* 51-60.

Ishikawa, T. (2006). The effect of increasing task complexity along the [±Here-and-Now] dimension on L2 written narrative discourse. In P. Garcia Mayo (Ed.), *Investigating tasks in formal language learning* (pp. 117-135). Philadelphia, PA: Multilingual Matters.

Ishikawa, T. (in press). On learner uptake: Manipulable sources of variation, motivating factors, and psycholinguistic processes. *Thought Currents in English Literature*, 79, xxx-xxx. The English Literary Society of Aoyama Gakuin University.

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.

Kawauchi, C. (2005). The effects of strategic planning on the oral narratives of learners with low and high intermediate L2 proficiency. In R. Ellis (Ed.), *Planning and task performance in a second language* (pp. 143-164). Amsterdam: Benjamins.

Kormos, J. (2006). *Speech production and second language acquisition.* Mahwah, NJ: Erlbaum.

Kowal, M., & Swain, M. (1994). Using collaborative language production tasks to promote student language awareness. *Language Awareness, 3,* 73-83.

Kuiken, F., Mos, M., & Vedder, I. (2005). Cognitive task complexity and second language writing performance. In S. Foster-Cohen, M.P. García-Mayo, & J. Cenoz (Eds.), *Eurosla Yearbook, Volume 5* (pp. 195-222). Amsterdam: Benjamins.

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

Levelt, W. (1989). *Speaking: From intention to articulation.* Cambridge, MA: MIT Press.

Long, M. (Ed.) (2005). *Second language needs analysis.* Cambridge: Cambridge University Press.

Long, M., & Crookes, G. (1993). Units of analysis in syllabus design: The case for task. In G. Crookes & S. Gass (Eds.), *Tasks in a pedagogical context: Integrating theory and practice* (pp. 9-54). Philadelphia, PA: Multilingual Matters.

Long, M., & Robinson, P. (1998). Focus on from: Theory, research, and practice. In C. Doughty & J. Williams (Eds.), *Focus on form in classroom second language acquisition* (pp. 14-41). Cambridge: Cambridge University Press.

Loschky, L., & Bley-Vroman, R. (1993). Grammar and task based methodology. In G. Crookes & S. Gass (Eds.), *Tasks and language learning: Integrating theory and practice* (pp. 123-167). Philadelphia, PA: Multilingual Matters.

MacWhinney, B. (2000). *The CHILDES system: Tools for analyzing talk* (3rd ed.). Hillsdale, NJ: Erlbaum.

Muranoi, H. (2000). Focus on form through interaction enhancement: Integrating formal instruction into a communicative task in EFL classrooms. *Language Learning, 50*(4), 617-673.

Nunan, D. (2004). *Task-based language teaching.* Cambridge: Cambridge University Press.

Ochs, E. (1979). Planned and unplanned discourse. In T. Givon (Ed.), *Syntax and semantics 12: Discourse and syntax* (pp. 51-80). NY: Academic Press.

Ortega, L. (2000). *Understanding syntactic complexity: The measurement of change in the syntax of instructed L2 Spanish learners*. Unpublished doctoral dissertation, Honolulu, HI: University of Hawaii.

Potter, M. C., & Lombardi, L. (1990). Regeneration in the short-term recall of sentences. *Journal of Memory and Language, 29,* 633-654.

Rahimpour, M. (1999). Task complexity and variation in interlanguage. In N. Jungheim & P. Robinson (Eds), *Pragmatics and pedagogy: Proceedings of the third Pacific Second Research Forum, Vol. 2*, (pp. 115-134). Tokyo: PacSLRF.

Raupach, M. (1984). Formulae in second language narrative production. In H. Dechert, D. Möhle & M. Raupach (Eds.), *Second language productions* (pp. 114-137). Tübingen: Gunter Narr.

Robinson P. (1995). Task complexity and second language narrative discourse. *Language Learning, 45,* 99-140.

Robinson, P. (1996). *Consciousness, rules, and instructed second language acquisition.* NY: Lang.

Robinson, P. (2001a). 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. (2001b). Task complexity, task difficulty, and task production: Exploring

interactions in a componential framework. *Applied Linguistics, 22,* 27-57.

Robinson, P. (Ed.) (2002), *Individual differences in instructed language learning.* Amsterdam: Benjamins.

Robinson, P. (2003) Attention and memory during SLA. In C. Doughty & M. Long (Eds.), *The handbook of second language acquisition* (pp. 631-678). Malden, MA: Blackwell.

Robinson, P. (2004). Comprehension, cognitive complexity, and task-based language production and acquisition. In D. Smith, S. Nobe, P. Robinson, G. Strong, M. Tani, & H. Yoshiba (Eds.), *Language comprehension: Perspectives from linguistics and language education* (pp. 187-240). Tokyo: Kuroshio.

Robinson, P. (2005). Cognitive complexity and task sequencing: Studies in a componential framework for second language task design. *International Review of Applied Linguistics, 43,* 1-32.

Robinson, P. (2006). Criteria for grading and sequencing pedagogic tasks. In P. Garcia Mayo (Ed.), *Investigating tasks in formal language learning* (pp. 7-26). Philadelphia, PA: Multilingual Matters.

Prabhu, N.S. (1987). *Second language pedagogy.* Oxford: Oxford University Press.

Samuda, V. (2001). Guiding relationships between from and meaning during task performance: The role of the teacher. In M. Bygate, P. Skehan, & M. Swain (Eds.), *Researching pedagogic tasks: Second language learning, teaching and testing* (pp. 119-140). Harlow, Essex: Pearson.

Schneider, V. I., Healy, A. F., & Bourne, L. E., Jr. (2002). What is learned under difficult conditions is hard to forget: Contextual interference effects in foreign vocabulary acquisition, retention, and transfer. *Journal of Memory and Language, 46,* 419-440.

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

Wigglesworth, G. (1997). An investigation of planning time and proficiency level on oral test discourse. *Language Testing, 14,* 85-106.

Wolfe-Quintero, K., Inagaki, S., & Kim, H. (1998). *Second language development in writing: Measures of fluency, accuracy and complexity.* Second Language Teaching & Curriculum Center, University of Hawaii.

Wickens, C., & Hollans, J. (2000). *Engineering psychology and human performance* (3rd ed.). Upper Saddle River, NJ: Prentice Hall.

Yule, G. (1997). *Referential communication tasks.* Hillsdale, NJ: Erlbaum.

Appendix A

Prompts for the HN and TT conditions

Prompt for the HN condition

Today a woman goes to a supermarket. She enters the supermarket through a door. She is wearing a black shirt. She puts her bag in the shopping cart. She is pushing the cart slowly. Maybe she is planning to buy many things for dinner.

Prompt for the TT condition

Yesterday a woman went to a supermarket. She entered the supermarket through a door. She was wearing a black shirt. She put her bag in the shopping cart. She was pushing the cart slowly. Maybe she was planning to buy many things for dinner.

Appendix B

Strip cartoon for the current study taken from Yule (1997)

Note. From *Referential Communication Tasks* (p. 67), by George Yule, 1997, Hillsdale, NJ: Lawrence Erlbaum Associates Inc. Copyright 1997 Lawrence Erlbaum Associates Inc. Adapted with permission.

Appendix C

Sample narratives from the HN and TT conditions

A sample essay by a female student in the HN condition (MEPT = 59)

One day, a woman goes to the supermarket. She enters the supermarket through a door. She is wearing a black shirt. The food corner is crowd with many people. She puts her white bag in the shopping cart and push the cart slowly. When she walks around the icecream corner, she happens to meet her friend. Her friend takes her big baby. They are talking between the whine bottles shelf. Suddenly, the baby takes one of the whine bottles, and he put this into the woman's bag while the woman and the baby's mother are fascinated with taking. She doesn't notice. After that, she goes to the checkout counter. The shoping manager finds the whine bottle in her bag. He takes her to the guest room, and the polce comes to there. She mistakes for the thief! Truelly she does not.

A sample essay by a female student in the TT condition (MEPT = 60)

Last week, a woman went to a supermarket to buy some foods. Maybe she was planning to make special dinner that night. So she needed to go there. She entered supermarket through the door. Pushing the shopping cart she looked for a lot of things vegetables, fruits, beaf etc. When she was pushing the cart, she happened to meet her friend and her friend's child. The woman who was her friend was also one of her neighborfoods. They were talking to each other. The woman's child was sitting on the shopping cart and hit upon to make a mischief. First he took a bottle of wine which was near by him, and put it in his mother's friend's bag. His mother and she was not notion. After talking, she was pushing the cart again and went to registration to finish the shopping. When she went out the supermarket, she was called out "Wait!" by the clerk. And then she was caught and under arrested by a police man because there was a bottle of wine in her bag which wasn't payed. She was very astonished and angry. Then she didn't know what she should do and say so she was at lost.