COGNITION HYPOTHESIS AND SECOND LANGUAGE PERFORMANCE: COMPARISON OF WRITTEN AND ORAL TASK PERFORMANCE

SHOKO SASAYAMA
University of Hawai'i at Mānoa

ABSTRACT

This study sought to test Robinson's (1995, 2003b) cognition hypothesis by investigating the effects of cognitive task difficulty on ESL learners' written and oral task performance.[1] Ten exchange students at the University of Hawai'i at Mānoa were given four sets of picture-based narrative tasks: a simple writing task, a difficult writing task, a simple speaking task, and a difficult speaking task. The simple tasks contained fewer characters or main foreground events, but not supporting background events, in cartoon-based stories; on the other hand, the complex tasks included more characters or both foreground and background events. The results indicated that (a) the difficult writing task was more successful than was the difficult speaking task in eliciting complex language production without deteriorating its accuracy; (b) the cognition hypothesis seems to be more relevant to language complexity than accuracy; and (c) accuracy of language production seems to be more susceptible to individual learners' ability to produce accurate sentences than cognitive task difficulty. These findings, therefore, partially support and partially disconfirm Robinson's cognition hypothesis.

INTRODUCTION

Literature Review

Task difficulty and its effects on language production. Use of tasks in second language (L2) classrooms has been gaining enormous popularity. Unlike traditional exercises which focus exclusively on the manipulation of language forms, tasks are designed to encourage learners to pay primary attention to meaning and simultaneously attend to the form that is necessary to

[1] In Robinson's (e.g., 2005) term, *task difficulty* indicates learners' perceptions of how difficult a task is. In the present study, however, the term refers to the level of cognitive demands of a task, which is independent of learners' perceptions.

Second Language Studies, 29(2), Spring 2011, pp. 107-129.

convey the meaning. Based on several empirical studies, Ortega (2007) argues that language learning is fostered when learners concurrently attend to meaning and form, and claims the importance of tasks in language acquisition. However, it is still an open question what kinds of tasks can effectively trigger such learning processes. Many second language acquisition (SLA) researchers (e.g., Bygate, Skehan, & Swain, 2001; Ellis, 2003, 2005; Ortega, 1999, 2007; Robinson, 2001a, 2001b, 2005; Skehan, 1998) acknowledge that various factors mediate the learning processes. One such factor that has received considerable attention in SLA research is the role of pre-task planning. The question of interest is whether and how giving learners time to plan their utterances before engaging in the task affects their language use in terms of complexity, accuracy, and fluency. Another factor that is hypothesized to mediate the effects of task lies within the task itself, that is, the question of how the difficulty inherent to the task itself affects the learners' language use. The current study focuses on the latter factor, inherent task characteristics, in an effort to further our understanding of the processes involved in task-based L2 use and learning.

Peter Robinson (1995, 2001a, 2001b, 2003b, 2005, 2007; Robinson & Gilabert, 2007) is one of the researchers who has a strong interest in the effects of inherent task characteristics on learners' language production. He argues that when tasks are cognitively and functionally demanding or difficult, learners will be encouraged to produce more complex and more accurate language production. This is known as the *cognition hypothesis*. This hypothesis is partially based on Cromer's (1974) cognition hypothesis for first language acquisition (FLA). Cromer (1974) argues that in FLA "our cognitive abilities at different stages of development make certain meanings *available* for expression" (italics in the original) and therefore, "it is cognition which determines language acquisition" (p. 246). Robinson (2003b) claims that this argument is applicable to adult language learning. Although adults have already developed complex notions of the world, they have not yet acquired enough linguistic knowledge to express them early in their L2 learning. Therefore, Robinson (1995, 2005, 2007) claims that complex notions and high functional demands will lead adult language learners to develop or stretch their interlanguage so that they can meet the increased demands of the task and express elaborated ideas.

As for accuracy development, Robinson (2005) argues that cognitively difficult tasks are "likely to draw learner attention to the ways in which the L1 and the L2 may differentially grammaticize conceptual notions . . ., and so have positive effects on L2 accuracy of production"

(p. 9). For example, learners will be more likely to notice how to form past tense sentences in their L2 in engaging in a difficult *there-and-then* task (as opposed to a simple *here-and-now* task), which requires task participants to produce the past tense for successful task completion. Another example would be relative pronouns: a difficult narrative task with a number of characters will encourage learners to pay attention to how similar objects can be discriminated and eventually to use noun modifiers such as relative clauses accurately. Difficult tasks, therefore, are hypothesized to offer better chances for accuracy development than are simpler tasks.

Robinson, therefore, views syntactic complexity and accuracy as arising from functional complexity in discourse and, hence, increased functional demands imposed by the task should have detectable linguistic consequences. In this model, concurrent attention to different aspects of L2 use is considered not just possible, but natural.

This claim is underpinned by the idea that individuals have multiple-resource pools of attention and their attentional capacity is unlimited (Robinson, 2003a). There are two separate models on attentional capacity and attention allocation. The first model is proposed by Wickens (1992). He argues that individuals draw their attention from different resource pools when completing different tasks. That is, we use different pools of attention, depending on processing mechanisms (i.e., encoding or responding), codes (i.e., spatial or verbal), modalities (i.e., visual or auditory), or responses (i.e., manual or vocal) that each task requires. It is claimed that a competition for attention occurs not between pools but within them. For example, it is usually easier to talk to somebody sitting in the passenger seat than to talk to him/her on the phone while driving. When one drives a car and uses a cell phone at the same time, one is drawing attention from the same resource pool (i.e., manual). On the other hand, one can easily talk to somebody in the passenger seat while driving, because talking (i.e., vocal) and driving (i.e., manual) require attention from different resource pools. The other model is called *the interference model*. Researchers who advocate this model, such as Gopher (1993) and Sanders (1998), argue that it is not limitation of capacity but the limited time available to complete a task or confusion that hinders processing of multiple sets of information. In other words, the central executive of working memory, which is responsible for attention allocation, easily loses its control when it has to deal with multiple stimuli within the limited amount of time. Robinson (2003a) combines these two models and argues that individuals have multiple-resource pools of attention and the amount of attention within each pool is unlimited.

Robinson's cognition hypothesis, however, has been criticized by scholars, such as Peter Skehan, who argue for limited attentional capacity. Skehan (1998; Foster & Skehan, 1996; Skehan & Foster, 2001) argues that human attentional capacity is limited and therefore, learners are unable to pay concurrent attention to accuracy and complexity. Foster and Skehan (1996) cite VanPatten's (1990) study to advocate the limited attentional capacity model. VanPatten (1990) conducted a research study to examine if learners were able to pay attention to both form and content simultaneously. He asked his participants to listen to a short passage. Group 1 was asked to listen exclusively for the content. Group 2 was asked to listen for an important lexical item while comprehending the passage content. Groups 3 and 4 were asked to listen for non-communicative grammatico-morphological forms (i.e., a definite article and a verb morpheme, respectively) while comprehending the passage content. The results indicated that when learners had to pay attention to language forms (i.e., Groups 3 & 4), their comprehension of the passage was negatively affected; however, learners in Group 2 could pay attention to the lexical item and comprehend the passage at the same time. VanPatten concluded that learners have difficulty in paying simultaneous attention to informational content and meaningless forms. Based on these findings, Foster and Skehan (1996) argue that human attentional capacity is limited and learners are not capable of paying concurrent attention to multiple aspects of the language (e.g., content and form, complexity and accuracy), especially when the task content itself is cognitively difficult and demanding.

In addition, when language learners, as opposed to native speakers of the language, are concerned, simultaneous attention to task content and language production becomes problematic (Skehan & Foster, 2001). Native speakers may be able to pay attention to both aspects simultaneously, as their language knowledge has already been proceduralized; and therefore, they have better controls over their language production. For language learners, however, their language knowledge has not yet been proceduralized or automatized and thus processing of the language consumes greater amount of attention. As a result, concurrent attention to task content and language forms becomes difficult to be achieved. In this view, when the task demands a considerable amount of attention to its content (i.e., the task is cognitively difficult and demanding), there will be only a little attention left for language forms. As a result, attention to complexity and accuracy becomes likely to compete with each other. (Skehan & Foster, 2001).

Empirical studies. Several scholars have empirically examined Robinson's cognition

hypothesis (e.g., Ishikawa, 2007; Kuiken & Vedder, 2007; Robinson, 1995, 2001b, 2007). Robinson (1995) himself, for example, used monologic narrative tasks (i.e., a simple *here-and-now* task and a difficult *there-and-then* task) to investigate the effects of increased task difficulty on L2 task performance. Participants were given a picture strip for each task and asked to narrate a story based on a given set of pictures. In the simple here-and-now task condition, they were allowed to view the picture strip while writing a story, whereas for the difficult there-and-then task, participants were asked to return the strip before they started their story telling. The results indicated that the difficult task promoted accuracy (as measured by target-like use of articles), but not at a significant level. No statistically significant differences were observed for complexity (as measured by S-nodes per T-unit) in this study.

In a more recent study, Robinson (2007) used interactive narrative tasks: a simple task with a small degree of reasoning demands, a neutral task with moderate reasoning demands, and a difficult task with high reasoning demands. Participants (speakers) were given three sets of pictures (one at a time) and asked to sequence the pictures in one minute and then tell a story based on these sequenced pictures. Their partner's (hearer) job was to sequence the same set of pictures according to the speaker's description. The results indicated no statistically significant differences among the three different tasks in either accuracy (as measured by error-free C-units) or complexity (as measured by clauses per C-unit).

Similar to Robinson (1995), Gilabert (2007) used a simple here-and-now task and a difficult there-and-then task and found that the complex task elicited significantly more accurate language production (as measured by percentage of self-repairs); however, increased task complexity did not have a significant effect on language complexity (as measured by S-nodes per T-unit). Tasks used in Robinson (1995), Robinson (2007), and Gilabert (2007) were all *oral* tasks and the results of these three previous studies are in contradiction to the cognition hypothesis.

Some other studies investigated the effects of increased task difficulty on L2 task performance in the modality of *writing*. Ishikawa (2006, 2007) used narrative writing tasks (i.e., a simple here-and-now task and a difficult there-and-then task). Ishikawa (2006) found that complexity (as measured by S-nodes per T-unit) and accuracy (as measured by target-like use of articles) were higher for the difficult task at a significant level. He found the same trends in his 2007 study: The difficult there-and-then task elicited significantly more complex (as measured by S-nodes per T-unit, clauses per T-unit, S-nodes per clause, and dependent clauses per clause)

and accurate (as measured by target-like use of articles) language production than did the simple task.

Kuiken and Vedder (2007) also used writing tasks and asked their participants to write two letters to discuss a holiday destination. In one task, they had to take into account three requirements to choose a holiday resort from five countries, whereas in the other task, they had to take into account as many as six requirements to choose a Bed & Breakfast in France from five options. The first task was considered to be cognitively simpler and the second one to be more difficult (based on Robinson's cognition hypothesis). The results indicated that accuracy (as measured by total error per T-unit) was significantly higher for the difficult task, whereas both simple and difficult tasks achieved approximately the same level of complexity (as measured by clauses per T-unit and dependent clauses per clause). These three studies, hence, lend stronger support to Robinson's cognition hypothesis.

A summary of the previous studies on the effects of increased task difficulty on L2 task performance is presented in Table 1. There are some trends observed. When these results are examined in terms of modality, difficult *writing* tasks seem to be more successful in simultaneously eliciting complex and accurate language production than are difficult *speaking* tasks. No studies to date, however, have used both writing and speaking tasks within one study design, which makes the direct comparison of written and oral task performance difficult. The present study, thus, attempts to fill this gap and investigates the effects of increased task difficulty on both written and oral task performance.

Table 1
Summary of the Previous Studies on Cognitive Task Difficulty and Task Performance

	Robinson (1995)	Robinson (2007)	Gilabert (2007)	Ishikawa (2006)	Ishikawa (2007)	Kuiken & Vedder (2007)
Modality	Spoken	Spoken	Spoken	Written	Written	Written
Complexity	+–	+–	+–	++	++	+–
Accuracy	+	+–	++	++	++	++

Note. ++ = significantly greater for a difficult task; + = greater for a difficult task; +– = little difference between simple and difficult task

Research Questions and Hypotheses

The present study poses the following four research questions (RQs):

1. What effects does increased task difficulty have on complexity and accuracy of learners' written task performance?
2. What effects does increased task difficulty have on complexity and accuracy of learners' oral task performance?
3. Is there a correlation between written and oral task performance in terms of its complexity?
4. Is there a correlation between written and oral task performance in terms of its accuracy?

Based on the previous literatures as outlined above, the following hypotheses were formulated:

Hypothesis 1: A difficult writing task will elicit more complex and more accurate language production than will a simple writing task (RQ 1).

Hypothesis 2: A difficult speaking task will fail to elicit complex and accurate language production (RQ 2).

Hypothesis 3: Therefore, no correlation will be observed between written and oral task performance in terms of both complexity and accuracy (RQs 3 & 4).

METHOD

Participants

Ten undergraduate students (eight females and two males) participated in this study. Their age ranged from 21 to 23 (M=21.4). All participants were studying at the University of Hawai'i at Manoa as exchange students from their home universities at the time of the experiment. Five of them were from Korea and the rest were from Japan. Three students had been living in Hawai'i for four months and the rest for eight to nine months at the time of the experiment. As for their English proficiency level, their iBT TOEFL scores ranged from 68 to 98 out of 120 (M=84.5)[2]. All participants were recruited on a voluntary basis and received five dollars as compensation.

Materials

Writing and speaking tasks. Participants engaged in four tasks: a simple writing task, a difficult writing task, a simple speaking task, and a difficult speaking task. The speaking tasks

[2] The score of 550 in ITP or PBT was converted to 80 in iBT scores.

were all monologic. Monologic tasks were chosen over interactive tasks for the following two reasons. Robinson's (2001a, 2001b, 2003b) cognition hypothesis (i.e., cognitively difficult tasks elicit greater language complexity and accuracy than do simple tasks) is more applicable to monologic tasks than to interactive tasks. Robinson (2001a, 2001b, 2003b) argues that complex interactive tasks will trigger more confirmation checks and clarification requests from a hearer and as a result, they will reduce complexity of speaker's each utterance. In addition, Ortega (1999) found that in engaging in an interactive task, some participants intentionally avoided complex structures for the sake of the hearer. I concluded, therefore, that monologic tasks would be more suitable for the present study.

As for the task type, narrative tasks were chosen for the following two reasons. Firstly, as Ortega (1999) argues, the narrative or the story-telling task type can be naturally conducted as a monologic task. Secondly, narrative tasks are more likely to elicit complex language production than other types of tasks (Foster & Skehan, 1996). Robinson (2001) used direction-giving tasks and found that both simple and difficult tasks elicited low language complexity. No significant differences were observed between the two tasks in terms of complexity of learners' L2 production; that is, the difficult task did not elicit complex language production as predicted by the cognition hypothesis. Robinson (2001) argues that in this particular case, the effects of cognitive task difficulty on task performance may have been minimized as a consequence of both tasks having elicited low language complexity. For this reason, the narrative task type was chosen for the present study.

One set of simple and difficult tasks was adapted from the Picture Arrangement subset of the Wechsler Adult Intelligence Scale-Revised, Japanese version[3] (Shiagawa, Kobayashi, Fujita & Maekawa, 1990) and from Elder and Iwashita (2005), respectively (see Appendix A & B). The first task was considered to be less demanding or less difficult because it contains fewer characters involved in the story than does the second task. The simple task included two characters of different genders (i.e., a woman and a man). The original picture included two women; however, I changed a woman to a man to reduce the cognitive demands of distinguishing two people of the same gender. When there is a woman and a man, task participants can simply refer to them as *a woman* and *a man* or *she* and *he*. On the other hand, when the pictures include two women, participants will have to distinguish them by the use of

[3] The chosen task was at the third level of complexity (difficulty) out of nine.

relative clauses or pre- and post-modification of noun phrases to add some extra information. This, in turn, makes the task cognitively more challenging. The difficult task, on the other hand, contained four men, a woman, a policeman, two ambulance attendants, and a dog. Participants were required to differentiate these similar characters, particularly the four men. Consequently, the second task was considered to be cognitively more difficult and demanding than the first task. The simple task describes a situation where a lady is trying to open the door by pulling it when she is supposed to push it (the *Door* task) (see Appendix A). The difficult task is about a dog which escapes from his owner, runs around the town, and causes a car accident (the *Dog* task) (see Appendix B).

The other set of simple and difficult tasks was both adapted from Tavakoli and Foster (2008). One picture set is about a girl and a boy going on a picnic (the *Picnic* task) and the other is about four boys trying to get their ball out of a hole (the *Football* task) (see Appendix C & D). The *Picnic* task was considered to be cognitively and linguistically more demanding, as it requires task participants to connect the foreground and the background events of the story for its successful completion. On the other hand, the *Football* task can be successfully completed simply by explaining the foreground information. Here, foreground means "the material that supplies the main points of discourse" and background is "the part that merely assists, amplifies, or comments on [foreground events]" (Tavakoli & Foster, 2008, p. 444). Based on several studies, Tavakoli and Foster (2008) hypothesized that task participants would produce more subordination in connecting foreground and background events because in English background information is typically described by using syntactic subordination, such as *while*, *when*, and *because*. This hypothesis was indeed supported by their findings. Consequently, the *Picnic* task was labeled as a difficult task and the *Football* task as simple.

Questionnaire. In order to learn the participants' background information, a questionnaire was administered. The questionnaire asked the participants' name, gender, age, educational background, first language, most recent TOEFL scores, ELI (i.e., English Language Institute) courses at the University of Hawai'i at Manoa, age of onset of English learning, total length of English learning, and length of stay in Hawai'i and other countries.

Design

The present study employed an experimental design with four independent variables: a

simple writing task, a difficult writing task, a simple speaking task, and a difficult speaking task. Dependent variables were complexity and accuracy of written and oral task performance. It is typically the case to include fluency as a dependent variable when investigating the effects of cognitive task difficulty on language production. However, it was eliminated from data analysis of the present study because the results for the variable of fluency have been relatively consistent: Fluency decreases as cognitive task difficulty increases. Fluency was not included also because it simplifies the research design and decreases the critical *F* value that the observed *F* value should exceed in order for the results to be statistically significant. There were two control variables. First, the pictures given to the participants were already sequenced to avoid any confounding variables, such as the order of the pictures (c.f., Robinson, 2007). Second, all participants received one minute pre-task planning time before engaging in each task.

Procedure

The participants first read a consent form and agreed to participate in the study. They, then, engaged in the four tasks (i.e., two writing and two speaking tasks with two levels of difficulty) and answered the questionnaire. In the writing task phase, the participants were given one minute to plan their story and typed it in an on-line survey program, *Survey Monkey*. They followed the same procedure twice, once for the simple writing task and the other for the difficult writing task. As for the speaking tasks, the participants were asked to plan their story for one minute and then narrate a story orally. They were given an audio recorder and recorded their own story in a sound booth. Again, they followed the same procedure twice, once for the simple speaking task and the other for the difficult speaking task. At the beginning of the first task, the participants were told that the length of their story would not matter and they could write or speak as long as they would like. The order of the simple-complex tasks and the writing-speaking tasks was counterbalanced. Furthermore, in order to minimize the influence of having different picture sets as writing or speaking tasks, six participants performed the *Football* and the *Picnic* tasks in writing and the *Door* and the *Dog* tasks in speaking. The rest of the participants performed the former two in speaking and the latter two in writing. After they finished the four tasks, they answered the questionnaire and provided their personal information, English proficiency level, and English learning experiences.

Data Analysis

Participants' utterances for the speaking tasks were all transcribed. The transcribed utterances and their writing texts were divided into T-units and dependent clauses. The T-unit was defined as "a main clause plus all subordinate clauses and non-clausal structures attached to or embedded in it" (Hunt, 1970, p. 4 cited in Foster, Tonkyn, & Wigglesworth, 2000). With reference to a grammar book (Celce-Murcia & Larsen-Freeman, 1999), a "dependent clause" was defined to include "subordination" (e.g., *When* I was in high school, I used to jog every morning) and "embedding" (e.g., The man *who* is standing over there is my dad.) (p. 20). To address research hypotheses 1 and 2, the data were submitted to a one-way repeated-measures analysis of variance (ANOVA) using SPSS. In addition, the Pearson correlation was performed to address research hypothesis 3. The dependent variables in the current study were analysed by two measures: the number of clauses per T-unit (complexity) and the percentage of error-free clauses (accuracy).

Complexity Measure. Complexity of language production was analysed by the number of clauses per T-unit. It was calculated following the formula of the total number of clauses divided by the total number of T-units. The same measure has been used by other researchers such as Foster and Skehan (1996), Mehnert (1998), and Robinson (2001b, 2007) for L2 spoken corpora[4] and Ishikawa (2007) for L2 writing corpora.

Accuracy Measure. Accuracy of language production was analysed by the percentage of error-free clauses per T-unit. Errors were defined at a sentence level rather than a discourse or pragmatic level. For this reason, it was not counted as an error when different verb tenses were used within a single task performance, when common nouns substituted pronouns, or when the use of conjunctions (e.g., *and*, *but*, and *when*) was not appropriate. Following Ishikawa's (2007) study, a/an distinction was not included in the analysis. To calculate the percentage of error-free clauses per T-unit, the number of error-free clauses was divided by the total number of T-units and multiplied by 100. Several researchers, such as Foster and Skehan (1996) and Mehnert (1998), have used the same accuracy measure for the analysis of oral task performance.

RESULTS

Tables 2 and 3 present the descriptive statistics for complexity and accuracy of written and

[4] Note, however, that these researchers, except for Mehnert (1998), used *C-unit* as a unit of analysis.

oral task performance. First, let's look at the effects of increased task difficulty on written and oral language production. Then, the results of a correlation between written and oral task performance will be reported.

Table 2

Descriptive Statistics for Complexity of Written and Oral Production

Variable	*n*	*M*	*SD*	Median	Minimum	Maximum	Skewness	Kurtosis
Writing								
Simple	10	1.13	0.11	1.11	1.00	1.38	1.40	3.22
Difficult	10	1.40	0.21	1.36	1.15	1.83	1.23	0.92
Speaking								
Simple	10	1.08	0.09	1.05	1.00	1.25	0.76	-0.78
Difficult	10	1.22	0.17	1.18	1.00	1.50	0.43	-0.87

Table 3

Descriptive Statistics for Accuracy of Written and Oral Production

Variable	*n*	*M*	*SD*	Median	Minimum	Maximum	Skewness	Kurtosis
Writing								
Simple	10	55.97	29.19	56.95	12.50	90.00	-0.32	-1.36
Difficult	10	45.61	17.62	50.00	21.43	75.00	0.12	-1.01
Speaking								
Simple	10	36.13	24.38	31.67	12.50	80.00	0.72	-0.81
Difficult	10	30.20	17.22	26.67	10.00	66.67	1.00	0.85

Descriptive Statistics and ANOVA

By looking at the descriptive statistics for complexity, it appears that the difficult writing task elicited most complex language production among the four tasks. Investigating the results in terms of different modalities, the difficult writing task seems to have elicited more complex language production than did the simple writing task. The same trend can be observed for the speaking tasks: It appears that the difficult speaking task elicited more complex language production than did the simple speaking task.

In order to test for the statistical significance of the observed differences, the complexity scores of both written and oral language production were submitted to one-way repeated-measures ANOVA. Mauchly's test of sphericity revealed that the present data met the

assumption of sphericity. The observed *F* value was 7.82. One-way repeated-measures ANOVA revealed that the *F* value exceeded the critical value of *F*, indicating that at least one of the mean complexity scores for the four tasks (i.e., a simple writing, a difficult writing, a simple speaking, and a difficult speaking) is different from at least one other mean complexity score ($p = 0.001$). As for the strength of association, a partial eta-square was 0.47. This means that approximately 47 % of the variability in the complexity scores was associated with variability due to engagement in different tasks. This value of 0.47 indicates a large effect size (Cohen, 1988). Table 4 presents a summary of the results of complexity obtained by one-way ANOVA with repeated-measures.

Table 4
One-way ANOVA with Repeated-Measures Summary Table for Complexity

Source	*df*	*SS*	*MS*	*F*
Test	3	0.60	0.20	7.82*
Subjects	9	0.05	0.005	
Test x Subjects	27	7.02	0.03	
Total	39	7.67		

Note. *$p < .025$ [5]

A post hoc comparison with Bonferoni revealed that the exact differences lie between the simple writing task and the difficult writing task as well as between the difficult writing task and the simple speaking task. The mean complexity scores for the simple and the difficult writing tasks were significantly different at the probability level of 0.003; and the mean complexity scores for the difficult writing and the simple speaking tasks were significantly different at the probability level of 0.02. Statistical significance was also confirmed by checking the 95% confidence intervals for difference. Zero did not fall between the lower and the upper bounds for either contrast: A null hypothesis that there is no difference in the mean complexity scores among the four different tasks can be rejected.

As for accuracy, it seems that regardless of task difficulty, the writing tasks elicited more accurate language production than did the speaking tasks. To test for statistical significance of the observed differences, the accuracy scores of both written and oral language production were submitted to one-way repeated-measures ANOVA. Mauchly's test of Sphericity revealed that

[5] With multiple performances of ANOVA (one for complexity and the other for accuracy), the alpha level was adjusted to 0.025 (i.e., 0.05/2) by using Bonferoni.

sphericity could be assumed for this set of data and one-way repeated-measures ANOVA found the *F* value of 5.50 to be significant ($p = 0.004$). This indicates that at least one of the mean accuracy scores for the four tasks is different from at least one another mean accuracy score. As for the strength of association, a partial eta-square was found to be 0.38. This means that approximately 38 % of the variability in the accuracy score was associated with variability due to engagement in different tasks. This value of 0.38 indicates a large effect size (Cohen, 1988). Table 5 presents a summary of the results of accuracy obtained by one-way ANOVA with repeated-measures.

Table 5
One-way ANOVA with Repeated-Measures Summary Table for Accuracy

Source	*df*	*SS*	*MS*	*F*
Test	3	382.98	1273.66	5.50*
Subjects	9	3055.86	339.54	
Test x Subjects	27	6258.25	231.79	
Total	39	9697.09		

Note. *$p < .025$ [6]

A post hoc comparison with Bonferoni, however, detected no significant differences among any of the mean accuracy scores of the four tasks. Figure 1 graphically presents the mean accuracy scores of the four tasks. As can be seen in Figure 1, the 95 % confidence intervals within the same modality greatly overlap; therefore, the difference may lie between different task modalities rather than the levels of cognitive task difficulty. In order to test for this assumption, two-way repeated-measures ANOVA was performed with modality (i.e., writing or speaking) and difficulty (i.e., simple or difficult) as the two factors, using SPSS. The results revealed a significant main effect for modality ($F = 9.06$, $p = 0.015$). As for the effect size, Cohen's *d* was calculated and was 0.74. This is considered to indicate a somewhat large effect size by Cohen (1988). Therefore, although the post hoc comparison with Bonferoni did not detect any significant differences between the four tasks in accuracy, two-way repeated-measures ANOVA revealed a significant difference between the mean accuracy scores of the written and oral language production.

[6] With multiple performances of ANOVA (one for complexity and the other for accuracy), the alpha level was adjusted to 0.025 (i.e., 0.05/2) by using the Bonferroni adjustment.

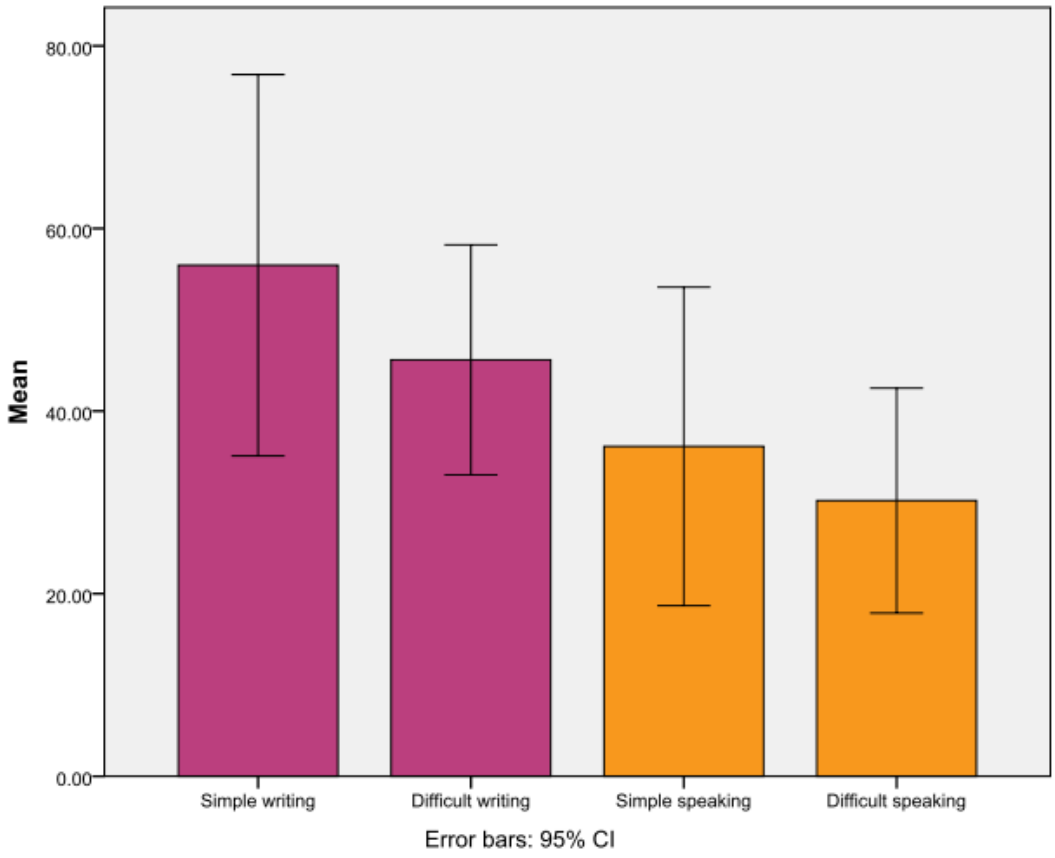

Figure 1. Mean accuracy scores for the four tasks with the 95% confidence intervals

Correlation

To investigate the relationship between written and oral task performance, the Pearson correlation was performed. It found no significant correlation between the two modalities in terms of complexity (r = 0.17, p = 0.47). The participants who produced complex sentences in engaging in the writing tasks, therefore, did not necessarily produce complex sentences for the speaking tasks and vice versa. Given the insignificant result, the complexity scores of the writing tasks do not predict the scores of the speaking tasks. On the other hand, a significant correlation was found between the writing and the speaking task performance in accuracy (r = 0.49, p = 0.02). A coefficient of determination for the writing and the speaking task performance for accuracy was 0.24: They share approximately 24% of variance. This value of the coefficient of determination or r^2 is considered to indicate a large effect size according to Cohen (1988). Figures 2 and 3 are the scatterplots showing the relationships between the two task modalities in

terms of complexity and accuracy, respectively.

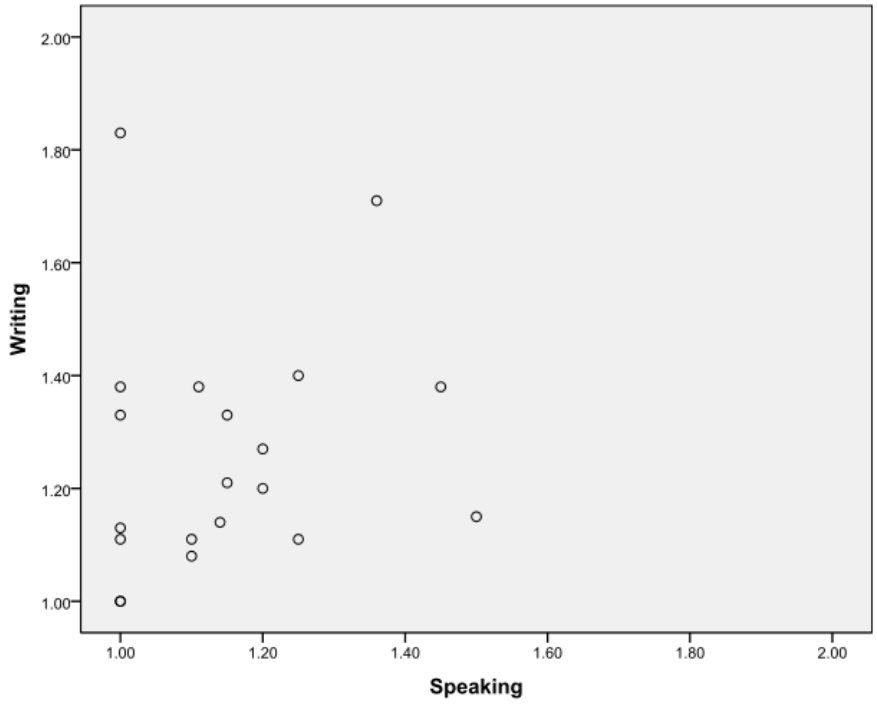

Figure 2. A scatterplot of the complexity scores for the writing and speaking tasks.

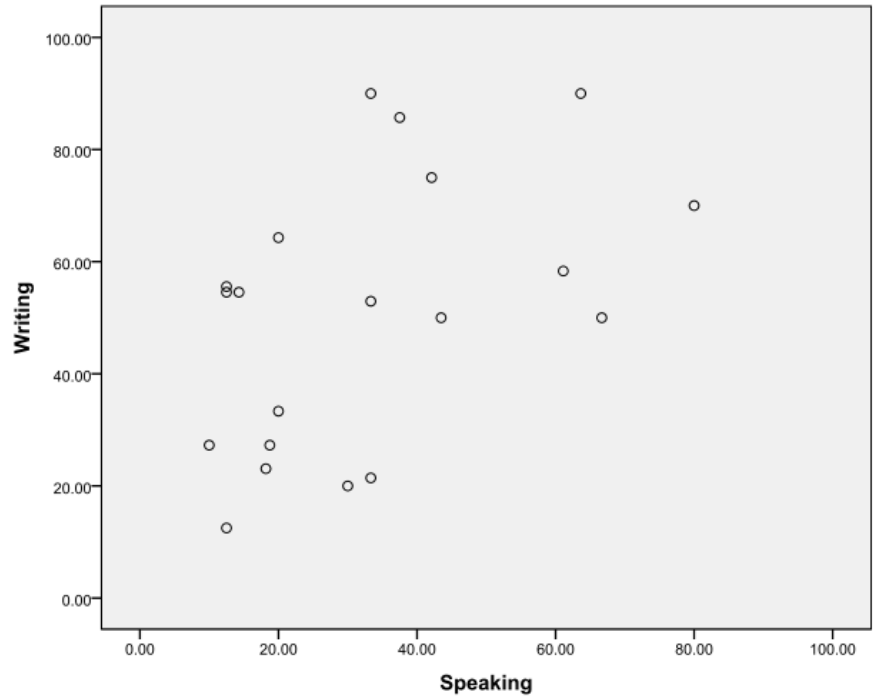

Figure 3. A scatterplot of the accuracy scores for the writing and speaking tasks.

DISCUSSION AND CONCLUSION

Task Difficulty and its Effects on Written and Oral Production

To summarize the results, the difficult writing task elicited significantly more complex language production than did the simple writing task and the simple speaking task, whereas accuracy remained at the similar level across all four tasks. These findings give full support to Hypothesis 2 (i.e., a difficult speaking task will fail to elicit complex and accurate language production) and only partial support to Hypothesis 1 (i.e., a difficult writing task will elicit more complex and more accurate language production than will a simple writing task). The results revealed that the difficult writing task, but not the difficult speaking task, elicited more complex language performance than did the simple tasks without deteriorating its accuracy. This may suggest that language learners are more likely to benefit from the idea of the cognition hypothesis when the task modality is writing rather than speaking. This seems to suggest that as Skehan (1998) argues, individuals' attentional capacity is indeed limited and it becomes more difficult to pay simultaneous attention to complexity and accuracy when the processing load is heavier as in speaking tasks.

Yet, this support for the limited attentional capacity model does not minimize the significance of Robinson's cognition hypothesis in task-based language research. Even though our attentional capacity seems to be limited, the results of the present study show that language learners are capable of paying a greater amount of attention to their language complexity without sacrificing their language accuracy when tasks are designed to pose higher cognitive, functional, and linguistic demands for their successful completion. The present study by no means lends full support to Robinson's cognition hypothesis, however. Although the participants of the present study were capable of producing more complex sentences in engaging in the difficult writing task without reducing the level of accuracy, their accuracy did not improve as the cognition hypothesis predicted. This seems to indicate that the cognition hypothesis may be more relevant to language complexity than accuracy: Language complexity may be more manipulable by inherent task characteristics or cognitive task difficulty than accuracy is.

In order to lend more support to this argument, the Pearson correlation was performed between the simple and difficult tasks regardless of task modality. The results revealed that there

was no significant correlation between the two for complexity (r = 0.36, p = 0.12). This seems to indicate that at least for the participants of the present study, complexity of language production was more influenced by task difficulty than by individual learner's ability to produce complex sentences. That is, even when a learner was capable of producing complex sentences in the complex tasks, they did not do so when they could successfully complete the simple tasks without producing complex sentences (see Figure 4). This trend, however, was not observed when the data were analyzed in terms of accuracy. There was a significant correlation between the simple and the complex tasks (r = 0.77, p = 0.000). This means that the task participants who scored low in accuracy for the simple tasks also scored poorly in accuracy for the complex tasks and vice versa (see Figure 5). A coefficient of determination was 0.59: The simple and the complex tasks share approximately 59% of variance in terms of accuracy. It can be argued from these results that complexity of language production is more strongly influenced by task difficulty than individuals' ability to produce complex sentences, whereas accuracy is more susceptible to individual learner's ability than cognitive difficulty of the tasks, at least when the participants of the present study are concerned.

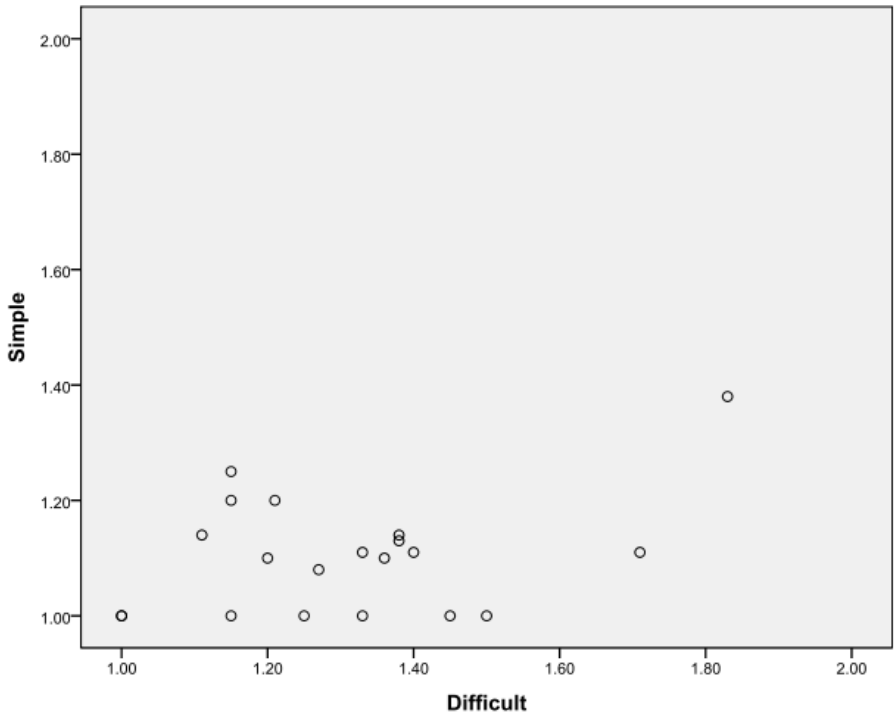

Figure 4. A scatterplot of the complexity scores for the simple and difficult tasks.

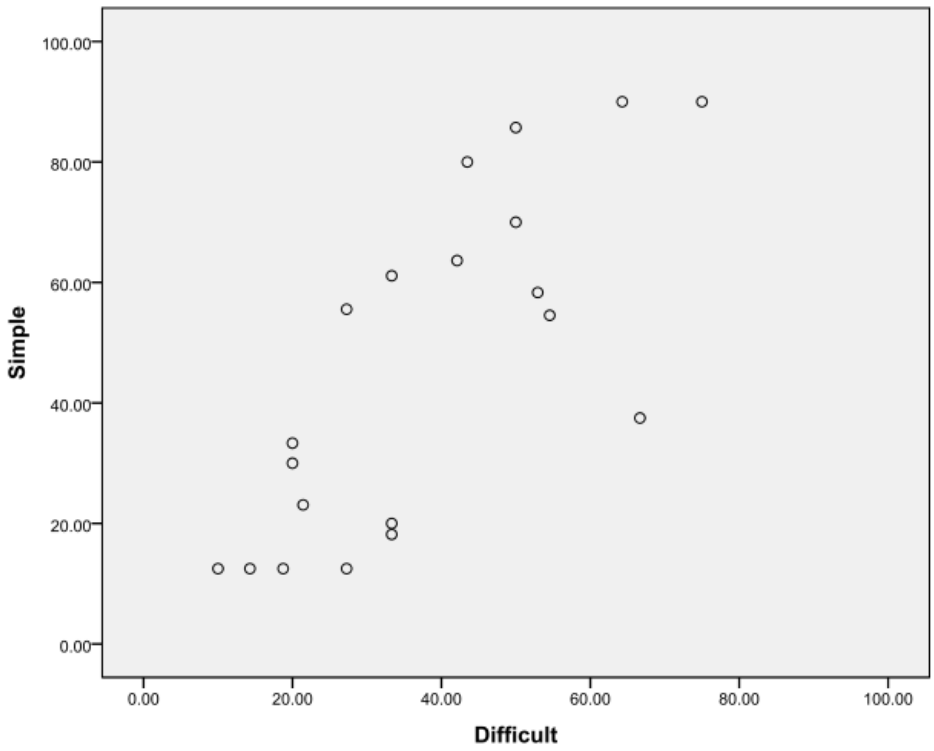

Figure 5. A scatterplot of the accuracy scores for the simple and complex tasks.

Correlation Between Written and Oral Task Performance

To summarize the research results, no correlation was found between the complexity scores of written and oral task performance, whereas the accuracy scores of written and oral task performance were significantly correlated with each other. Therefore, Hypothesis 3, which predicted that no correlation will be observed between written and oral task performance in terms of both complexity and accuracy, was only partially supported.

By looking at Figure 2, it can be observed that the complexity scores of oral production are low, relative to the ones of written production. This brings us back to the issue of attentional capacity limits and task performance: With limited attentional capacity, the participants must have had a harder time paying attention to complexity of their language production and producing complex sentences in speaking, as speaking tasks typically pose heavier processing load than do writing tasks.

The significant correlation between the two modalities in accuracy indicates that participants who could produce accurate sentences orally were capable of doing so in writing and vice versa.

This supports the earlier argument that accuracy is more susceptible to individuals' ability to produce accurate sentences and is rarely influenced by inherent task characteristics such as cognitive task difficulty, at least when the participants of the present study are concerned.

In sum, based on the results that the difficult writing task in the present study was more successful than was the difficult speaking task in eliciting complex language production without deteriorating its accuracy, writing tasks may be more beneficial in improving learners' complexity and accuracy in a balanced manner. Furthermore, against Robinson's argument and in line with Skehan's claim, learners' attentional capacity seems to be limited. However, this does not entirely denigrate the significance of Robinson's cognition hypothesis. The cognition hypothesis is important in that cognitively and functionally demanding tasks have a power to encourage language learners to challenge the limit of their interlanguage and to produce more complex sentences, especially when the modality is writing and the processing load is eased. As for accuracy, whether learners can produce accurate sentences seems to depend more on their individual ability than cognitive, functional, and linguistic demands of the task.

SUGGESTIONS FOR FUTURE RESEARCH

This pilot study revealed some unexpected results regarding accuracy of language production: As long as the participants in the present study are concerned, their accuracy scores seem to be more correlated with their ability to produce accurate sentences than cognitive task difficulty. It would be interesting and beneficial to further investigate this result with a larger number of participants.

REFERENCES

Bygate, M., Skehan, P., & Swain, M. (Eds.). (2001). *Researching pedagogic tasks: Second language learning, teaching, and testing.* Harlow: Pearson Education.

Celce-Murcia, M., & Larsen-Freeman, D. (1999). *The Grammar Book: An ESL/EFL Teacher's Course.* Boston : Heinle & Heinle.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Cromer, R. (1974). The development of language and cognition: The cognition hypothesis. In B. Foss (Ed.), *New perspectives in child development*. Harmondsworth, UK: Penguin.

Elder, C., & Iwashita, N. (2005). Planning for test performance: Does it make a difference? In R. Ellis (Ed.), *Planning and Task Performance in a Second Language* (pp. 219-238). Amsterdam: John Benjamins.

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

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

Foster, P., & Skehan, P. (1996). The influence of planning time on performance in task-based learning. *Studies in Second Language Acquisition*, *18*, 299-234.

Foster, P., Tonkyn, A., & Wigglesworth, G. (2000). Measuring spoken language: A unit for all reasons. *Applied Linguistics*, *21*(3), 354-375.

Gilabert, P. (2007). The simultaneous manipulation of task complexity along planning time and [+/- Here-and-Now]: Effects on L2 oral production. In M. P. Garcia Mayo (Ed.), *Investigating tasks in formal language learning* (pp. 44-68). Clevedon: Multilingual Matters.

Gopher, D. (1993). The skill of attention control: Acquisition and execution of attention strategies. In D. Meyer & S. Kornblum (Eds.), *Attention and performance XIV: Synergies in experimental psychology, artificial intelligence, and cognitive neuroscience* (pp. 265-291). Cambridge, MA: MIT.

Ishiwaka, T. (2007). The effect of manipulating task complexity along the [+/- Here-and-Now] dimension on L2 written narrative discourse. In M. P. Garcia Mayo (Ed.), *Investigating tasks in formal language learning* (pp. 136-156). Clevedon, UK: Multilingual Matters.

Kuiken, F., & Vedder, I. (2007). Cognitive task complexity and linguistic performance in French

L2 writing. In M. P. Garcia Mayo (Ed.), *Investigating tasks in formal language learning* (pp. 117-135). Clevedon, UK: Multilingual Matters.

Mehnert, U. (1998). The effects of different lengths of time for planning on second language performance. *Studies in Second Language Acquisition*, *20*, 83–108.

Ortega, L. (1999). Planning and focus on form in L2 oral performance. *Studies in Second Language Acquisition*, *21*, 109-148.

Ortega, L. (2007). Meaningful L2 practice in foreign language classrooms: A cognitive-interactionist SLA perspective. In R. DeKeyser (Ed.), *Practice in a second language: Perspectives from applied linguistics and cognitive psychology* (pp. 180-207). Cambridge, UK: Cambridge University.

Robinson, P. (1995). Attention, memory and the 'noticing' hypothesis. *Language Learning, 45*, 283-331.

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, UK: Cambridge University.

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

Robinson, P. (2003a). Attention and memory in SLA. In C. Doughty & M. Long (Eds.), *Handbook of second language acquisition* (pp. 631-678). Oxford, UK: Blackwell.

Robinson, P. (2003b). The cognitive hypothesis of adult, task-based language learning. *Second Language Studies, 21*, 45-107. Retrieved May 8, 2008, from www.hawaii.edu/sls/uhwpesl/21(2)/Robinson.pdf

Robinson, P. (2005). Cognitive complexity and task sequencing: A review studies in a componential framework for second language task design. *International Review of Applied Linguistics in Language Teaching*, *43*(1), 1-33.

Robinson, P. (2007). Task complexity, theory of mind, and intentional reasoning: Effects on L2 speech production, interaction, uptake and perceptions of task difficulty. *International Review of Applied Linguistics in Language Teaching*, *45*(3), 193-213.

Robinson, P., & Gilabert, R. (2007). Task complexity, the cognitive hypothesis and second language learning and performance. *International Review of Applied Linguistics in Language Teaching*, *45*(3), 161-176.

Sanders, A. (1998). *Elements of human performance.* Mahwah, NJ: Lawrence Erlbaum Associates.

Shiagawa, F., Kobayashi, S., Fujita, K., & Maekawa, H. (1990). *Manual of the Japanese Wechsler Adult Intelligence Scale – Revised*. Tokyo: Psychological Corporation.

Skehan, P. (1996). A framework for the implementation of task-based instruction. *Applied Linguistics, 17*(1), 38-62.

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

Skehan, P., & Foster, P. (2001). Cognition and tasks. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 183-205). Cambridge, UK: Cambridge University.

Tavakoli, P., & Foster, P. (2008). Task design and second language performance: The effect of narrative type on learner output. *Language Learning, 58*(2), 439-473.

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

Wickens, C. D. (1992). *Engineering psychology and human performance* (2nd ed.). New York: Harper Collins.

Second Language Studies, *29*(2), Spring 2011, pp. 107-129.