THE JOURNAL OF ASIA TEFL Vol. 5, No. 1, pp. 29-63, Spring 2008

The Effect of Task Demands of Intentional Reasoning on L2 Speech Performance

Tomohito Ishikawa
Aoyama Gakuin University/Soka Women's College, Japan

The present study examined effects of task demands on second language speech performance by manipulating the task complexity dimension of intentional reasoning, which was proposed as one of the task complexity features by Robinson (2007a). Twenty-four Japanese college students participated in the study and performed a control task and two intentional reasoning tasks. Six production measures of fluency, complexity and accuracy were employed in order to test whether intentional reasoning would lead to increases in accuracy and complexity at the cost of fluency, as predicted by the Cognition Hypothesis (e.g., Robinson, 2005, 2007a). The results showed that intentional reasoning produced 1) positive effects on syntactic as well as lexical complexity and accuracy, and 2) a negative effect on fluency (i.e., disfluency).

Key words: task complexity, the cognition hypothesis, intentional reasoning

THE EFFECT OF TASK DEMANDS OF INTENTIONAL REASONING ON L2 SPEECH PERFORMANCE

Task-based language teaching (TBLT) has recently attracted increased attention in the field of second language acquisition (SLA) research and pedagogy (e.g., Garcia-Mayo, 2007; Long & Crookes, 1992; Nunan, 2004;

Robinson, 2005, 2007a; Skehan, 1998, Van den Branden, 2006; Willis, 1996, etc.). Part of the major reasons for this is that task is a viable, and crucially manipulable and gradable, unit both for researching and language teaching (e.g., Long & Crookes, 1992). In short, tasks can be seen as a viable unit that researchers and language teachers can deploy for instructional purposes.

Emerging out of conceptions of tasks as instructional resources that can be deployed strategically by language teachers and syllabus designers are post-modernistic ideas of context-sensitive principles or "theories." Richards and Rodgers (2001) and Willis (1996), for instance, characterize task-based language teaching as "approach" as opposed to ready-made solutions or "method" (e.g., Kumaravadivelu, 1994), where the former requires the teacher to constantly take into account the learner and situational characteristics in relation to teaching and learning. Needless to say, such teacher theories (i.e., the backbone of what teachers think and believe about language teaching) are supported by views on what language is and how languages are learned and used, and so they affect teacher perceptions, actions and how language is taught (e.g., Wilkins, 1990). Furthermore, even the traditional distinction between syllabus and methodology also becomes blurred when tasks are considered as central to syllabus design (e.g., Long, 1985; Nunan, 2004). Within such conceptual contexts, part of the question of how languages are taught or learned needs to be replaced by how tasks are categorized and sequenced in principled ways.

One of the proposed theories is that L2 instruction should be based on particular needs as they are related to the learner's role defined within a specific communication context (e.g., Long, 2006). Long (1989) claims that a series of pedagogic tasks can be created or identified and be sequenced so that they may approximate demands of the target real-world task. From this perspective, maintaining and increasing relevance of pedagogic tasks to the target real-world tasks occupies the central place.

Complementary to Long's proposal, Robinson (2007a) points out the necessity to categorize pedagogic tasks from information processing perspectives, while maintaining behavioral descriptions of, and relevance to, the target tasks identified by needs analysis (e.g., one-way vs. two way tasks). Robinson also emphasizes the importance of respecting learner characteristics and considering the role they would play in TBLT. In order to accommodate those broad issues, Robinson (2001a, 2007a) proposes his framework for TBLT, where three major task components and their roles in syllabus design are identified.

ROBINSON'S TRIADIC FRAMEWORK FOR TBLT

One of the major advantages of having a conceptual framework is that it can provide a comprehensive research picture, where important factors and their inter-relations are specified. The present study adopts Robinson's triadic framework mainly 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 and relating the three task components: task condition, task difficulty, and task complexity, the last of which is the central concern of the present article.

The first component of *task condition* refers to the interactive characteristics of tasks, which consists of several descriptive, behavioral properties of the target as well as pedagogic tasks. Those interactional, descriptive task characteristics are identified by needs analysis and include participation variables (e.g., open- vs. closed-tasks) and participant variables (e.g., participants familiarity).

Second, the task component of *task difficulty* refers to the L2 learner's "perceptions" of difficulty that are dependent on the level of task complexity and L2 learner's individual differences such as ability (e.g., working memory capacity) as well as affective variables (e.g., anxiety). Some variables within this category may be subject to temporal change (e.g., task motivation, interest) in contrast to relatively stable ability variables (aptitude, intelligence, proficiency, etc.).

Finally, *task complexity* refers to pedagogic tasks' information processing demands on memory, attention, and reasoning (Robinson, 2001a). This category of task component is characterized as cognitive in nature, and so represents inherent and relatively fixed processing demands of pedagogic tasks. Thus, they are manipulable by teachers and syllabus designers based on cognitive criteria in a prospective way. According to Robinson, task classification based on cognitive-theoretic criteria plays central roles in decision-making processes about task sequencing.

According to Robinson (2001a, 2001b, 2005, 2007a), task complexity can be defined partly based on what he calls "resource-directing" dimensions. The resource-directing dimensions of task complexity have their essential characteristics of "directing" the learner's cognitive resources to various language codes (Robinson, 2001a, 2001b, 2003, 2005, 2007a). Those dimensions are expressed by "±" features although these can form a continuum. It is claimed that by manipulating these features, the learner's cognitive resources are predicted to be drawn to specific task-relevant language forms, promoting learning and retention of new L2 items and analysis of existing L2 knowledge base.

INTENTIONAL REASONING AND PREDICTIONS OF THE COGNITION HYPOTHESIS

The purpose of the present study is to examine the effect of manipulating one of the proposed resource-directing dimensions, [± intentional reasoning], on L2 monologic *performance* rather than on learning. According to Robinson, intentional reasoning concerns "reasoning about other people intentions, beliefs and desires and relationships between them" (Robinson, 2007a, p. 18). This means that intentional reasoning is specific to human social/psychological domain. In fact, intentional reasoning is often described as everyday psychological or social reasoning.

Robinson's Cognition Hypothesis (Robinson, 2005) predicts that increasing task complexity along the [± intentional reasoning] dimension leads to increases in accuracy and complexity at the cost of fluency. This is partly because manipulating resource-directing dimensions is predicted to affect allocation of cognitive resources to language form.

In cognitive terms, intentional reasoning would create greater conceptual gaps that needs to be filled with additional thoughts and inferences, which are to be expressed verbally. In performing the intentional reasoning task then, the learner needs to rely on inferences based on mindreading, or thinking about intentions and desires of others, in order to cross conceptual gaps, relative to the task without such demands, where simple information transmission is required.

In linguistic terms, according to Givon (1998), thinking imposes discourse-pragmatic demands on the part of the speaker, where multi-propositional considerations are required. Thus, thinking about what to say in what order invites top-down hierarchical organization of texts (e.g., Berman & Slobin, 1994) and bottom-up attention to language form (e.g., Swain, 1995), or enhanced interactions of both (e.g., Celce-Murcia & Olshtain, 2005), and these effects would be reflected in overall accuracy as well as complexity.

Such coherence-establishing reasoning demands would decrease L2 fluency as well. Delays in producing speech plans should lead to delays in speaking (e.g., Levelt, 1989). The speaker may hesitate to a greater extent or speak slowly in order to create time. Intentional reasoning would also require incessant decision-making. Furthermore, inferences, confirmations, doubts, and reconfirmation are likely products. Thus, it is expected that intentional reasoning would impede efficient scheduling and smooth execution of L2 speech performance relative to no intentional reasoning counterparts.

Furthermore, from a pedagogical perspective, if task sequencing is a pedagogical expedient means to guide the learner's language development and if intentional reasoning tasks have greater potential to expedite such developments than their non-reasoning counterparts, we would naturally expect that intentional reasoning tasks should possess greater guiding potential than their no reasoning counterparts. Thus, in all three respects, intentional reasoning should elicit more advanced language performance, leading to greater accuracy and complexity at the cost of fluency, as the Cognition Hypothesis predicts.

RESEARCH QUESTION AND HYPOTHESES

Based on the above cognitive and linguistic rationales and pedagogical considerations, the present study formulated the following research question: *Does intentional reasoning lead to greater accuracy and complexity at the cost of fluency?* In order to answer the research question, the following methodology was employed. First of all, the present study used six descriptive production measures to be described in the Methodology section: fluency measures (un-pruned and pruned speech rate, and disfluency); structural complexity measure (S-nodes per T-unit); lexical complexity measure (Guiraud 2000); accuracy measure (percentage of error-free T-units). Secondly, three tasks were created: the control task and the two intentional reasoning tasks. The control task will be referred to as “No Reasoning Task” and each of the two intentional reasoning tasks will be referred to as “Simple Reasoning Task” and “Complex Reasoning Task” respectively. Thirdly, in statistical terms, planned comparisons were conducted in order to contrast the to-be-stated predicted effects on L2 speech production of the No Reasoning Tasks versus the two intentional reasoning tasks. Planned comparisons were used because the main purpose of the present study was to examine the effect of [± intentional reasoning] and additionally because given the lack of current empirical evidence it was felt that it would be premature to predict specific differential effects of the two reasoning tasks in a priori ways. Whether or not such fined-grained continuum could be established depends on future investigations and subsequent theorizing.

As stated above, the present study formulated six hypotheses. Hypotheses 1 to 3 were formulated as follows:

Hypothesis 1: Intentional reasoning will reduce L2 fluency relative to the No Reasoning Task, as reflected in the measure of un-pruned speech rate.

Hypothesis 2: Intentional reasoning will reduce L2 fluency relative to the No Reasoning Task, as reflected in the measure of pruned speech rate.

Hypothesis 3: Intentional reasoning will reduce L2 fluency relative to the No Reasoning Task, as reflected in the measure of disfluency.

Because specific operationalizations are given in the Methodology section, here only brief sketches of the production measures are provided in conceptual terms. First, Hypotheses 1 to 3 concerns different aspects of speaking fluency. The measure of un-pruned speech rate concerns overall measure of L2 production speed whereas the measure of pruned speech rate is a more selective one in that it considers only those utterances that are based on intended messages of the speaker. From a lexical perspective, those pruned words can be considered as constituting intended words produced by the speaker (e.g., Shriberg, 1994). Speech rate is dependent partly on pause length and partly on articulation rate (e.g., Goldman-Eisler, 1968). Another aspect of fluency is disfluency rather than production speed. As such, this measure can reflect the degree of disfluency (i.e., hesitations). In accordance with the predictions of the Cognition Hypothesis and the speculations developed above, Hypotheses 1 to 3 predicted that aspects of fluency, if not all, would be impaired due to intentional reasoning demands.

The rest of the hypotheses are as follows:

Hypothesis 4: Intentional reasoning will enhance L2 complexity relative to the No Reasoning Task, as reflected in the measure of S-nodes per T-unit.

Hypothesis 5: Intentional reasoning will enhance L2lexical complexity relative to the No Reasoning Task, as reflected in the measure of Guiraud 2000, a type-token ratio.

Hypothesis 6: Intentional reasoning will enhance L2accuracy relative to the No Reasoning Task, as reflected in the measure of percentage of error-free T-unit.

Hypotheses 4 to 6 concerns aspects of complexity and accuracy. Regarding Hypothesis 4, the syntactic complexity measure of S-nodes per T-unit is a good indicator of examining the aforementioned hierarchical organization of texts and has been employed by several TBLT studies (e.g., Gilabert, 2007; Ishikawa, 2006; Robinson, 1995). Hypothesis 5 concerns the lexical complexity measure of what I call "Guiraud 2000," which is a type-token measure and can be an indicator of activations of wider semantic fields due to intentional reasoning. The traditional Guiraud Index is known as an alternative to the simple type-token measure, because the simple type-token measure has been notorious for being influenced by text lengths. The improved type-token ratio measure is the Guiraud Index and was used by Gilabert (2007) in TBLT. Guiraud Index is defined as total number of words (i.e., tokens) divided by square root of total number of tokens.

The difference between the traditional Guiraud Index and "Guiraud 2000" as defined in the present study is that the latter considers only those words whose frequencies are equivalent to, and lower than, 2000 word frequency level, which is reported to be more sensitive than the traditional Guiraud Index (Daller, van Hout, & Treffers-Daller, 2003). Finally, Hypothesis 6 is based on the accuracy measure of percentage of error-free T-units. This global measure of accuracy has been used by a number of TBLT studies (e.g., Gilabert, 2007; Ishikawa, 2007; Robinson, 1995) and can capture the effect of intentional reasoning on accuracy. Specific operationalizations of these production measures are described in the next section.

METHODOLOGY FOR THE PRESENT STUDY

Participants

The participants of the present study were college students, whose L1 was all Japanese (N = 24; male = 2, female = 22). They were either English majors or English for Academic Purposes (EAP) students. The participants' proficiency levels ranged between low to high intermediate and their ages ranged from 18 to 35 years old, with the average age of 22.2 years old. Their mean length of stay in English-speaking countries was approximately 8 months. They participated in the study voluntarily and were paid 1,000 yen (circa 10 dollars).

Materials

Three tasks were used in the present study. All tasks were one-way monologic tasks. First, two versions of reasoning tasks were created, the Simple Reasoning Task and the Complex Reasoning Task. Afterwards, a control task, No Reasoning Task, was created.

In all tasks, the learner played a role of a company's general manager, who was in charge of a new section that was established "a week ago." Then the manager (= learner) was asked to report to the president about hypothetical human relationships of his/her section members. At the time of reporting, the setting was that the president was not available, so the manager decided to report to the president by leaving a message on the president's answering machine.

In the Simple Reasoning Task, the learner was asked to report to the president about hypothetical human relationship changes. The learner as manager had only two section members (Appendix A) and first selected a "trigger" out of four choices (i.e., one of the section members either lost a floppy disk, showed up late for a meeting, lost an important document, or deleted data on a computer), which caused hypothetical "trouble" between the two section members. Each time the participant performed a task, they were asked to choose only one trouble trigger and assigned it to one of the two section members. The learner was then asked to think up the trouble trigged by a mistake and report to the president about what happened in the office. In the Simple Reasoning Task presented in Appendix A, the solid line (originally colored in blue) represented a good relationship, which turned into a bad one, represented by the broken line (originally colored in red). Section members' hypothetical names and positions in the office were also available.

The Complex Reasoning Task was performed under exactly the same conditions, except that the manager was in charge of four section members rather than two (Appendix B). Four out of the six possible human relationships were supposed to have changed after hypothetical office trouble. It was expected that increasing the number of human relationships to be verbalized would also require more elaborated rehearsal for L2 speech performance (i.e., speech planning) relative to that required by the Simple Reasoning Task.

Finally, the No Reasoning Task (Appendix C) asked the manager simply to describe the current human relationships between the section members. In the No Reasoning Task, necessary information to convey was present (i.e., current human relationships between the members) and there was no explicit need to think up a trouble situation. In order to elicit sufficient L2 speech data, a setting was chosen, where the manager was in charge of four members rather than two or three members.

Procedures

The procedures of the experiment were as follows. The research was conducted on an individual basis. When the participant came to the room, the researcher had a rapport with him/her for them to be relaxed. No other people, apart from each participant and me, were present. Afterwards, the researcher told the participant that she/he would perform three tasks. The researcher then passed the task-instructions sheet to the participant and told that she/he could take as long time as possible to read the instructions and were also allowed to ask questions but not during task performance. Task-instructions sheets were two kinds: one for the No Reasoning Task (Appendix D) and the other for the Simple and the Complex Reasoning Tasks (Appendix E).

Two pieces of information were intentionally left out from the task instructions sheet: the number of section members involved and trouble triggers (the latter were only for the intentional reasoning tasks). This was done in order to double-check whether the participant read and understood the instructions. Almost all participants asked questions about trouble triggers (i.e., job mistakes) and the number of section members involved. The participant's asking questions naturally provided the researcher with opportunities to re-describe task instructions. The researcher then explained how many section members were involved without giving specific information regarding the patterns of their human relationships.

Following that, the researcher showed the trouble-trigger list and the participant chose just one trouble trigger out of four choices and assigned it to one of the section members. The participant was also asked to select a different trouble trigger each time they performed a reasoning task. Prior to task performance, three-minute planning time was given to the learner. Additionally, prior to task performance, the participant was also informed that the answering machine would stop in three minutes and that they did not have to keep talking for three minutes. After each planning session, the researcher read out a message of the answering machine ("Hi, you have reached xxx-xxx-xxxx. I cannot come to the phone right now. Please leave a message after the beep. Thanks"). A beep sound followed and the recording started. In order to reduce the carry-over effect, the participant took a short break between task performance sessions.

Experimental Design for the Present Study

The experiment used a within-subject experimental design (i.e., all participants performed all tasks). The orders of task performance were also counterbalanced by means of a complete Latin-Square experimental control (i.e., ABC, ACB, BAC, BCA, CAB, and CBA, where each letter represents each of the three tasks).

L2 Production Measures

Fluency measures. As mentioned briefly, the present study employed three fluency measures: un-pruned and pruned speech rate and disfluency. The un-pruned speech rate measure was defined as total number of words produced divided by total speaking time in second multiplied by 10. That is, the measure is equivalent to un-pruned speech rate per 10 seconds. The use of 10 seconds rather than a minute was motivated by the fact that some learners completed tasks within one minute. In counting the number of words, hesitation markers such as “uhm” or “um” were also counted as words (Clark & Fox Tree, 2002). Often the participant produced word fragments during hesitations (e.g., “sec-” for “section”) and in those cases, word fragments were coded as .5, following Lickley (1998). With respect to the total speaking time, the beginning of the speech was defined as the end of the beep sound of the answering machine. The end of the speech was most often signaled by closing markers such as “thank you,” “see you,” “Good-bye” but a few participants could not finish speaking within the pre-specified three minutes. In such cases, the end of the T-unit closest to the three-minute boundary was identified as the end of the speech. The total speaking time was measured using a software program, Speech Analyzer. Identical coding procedures were used for pruned speech rate as well (i.e., the number of pruned words per 10 seconds), with the exception being that, needless to say, pruned words did not contain word fragments.

With respect to the measure of disfluency, it was operationalized as total number of disfluency episodes divided by total number of pruned words multiplied by 10. Thus, this is equivalent to total number of disfluency episodes per pruned 10 words (Ishikawa, submitted). To illustrate, if the learner said, “Bob forgot to tell Sue that he *losed* lost her document,” the italicized word (i.e., "losed") was coded as a disfluent word. In this particular example, because the number of fluent words was 10 and there was only a single disfluency episode, the specific value of the measure of disfluency was coded as .10 (= 1/10). Complete descriptions of the coding steps are beyond the scope of this paper but are available in Ishikawa (2008).

Complexity and accuracy measures. The present study also used three descriptive measures of syntactic as well as lexical complexity and accuracy: S-nodes per T-unit, Guiraud 2000, and percentage of error-free T-unit (EFT). The formula for the syntactic complexity measure of S-nodes per T-unit was: total number of S-nodes divided by total number of T-units. The definition of a T-unit given by Hunt was "one main clause with all the subordinate clauses attached to it" and "[t]he number of subordinate clauses can, of course, be none" (Hunt, 1965, p. 20). Conventional phrases were excluded from the analysis (e.g., "Hello" "thank you"). T-unit was used in the present study because the tasks employed were non-interactive monologic tasks. S-nodes are equivalent to finite and infinite verb phrases (VPs). Modals (e.g., "can") were not considered as constituting S-nodes whereas their phrasal counterparts were considered as S-nodes (e.g., "be able to"). Relevant contracted forms were also considered as S-nodes (e.g., "wanna"). Also considered as S-nodes were modal-like elements such as "would like," "(had) better." With respect to sentence fragments, they were considered as T-units without S-nodes.

Next, the formula for Guiraud 2000 was total number of words (i.e., tokens), whose frequency ranks were higher than, or equivalent to, the frequency level of 2000, divided by square root of total number of tokens. "Guiraud 2000" in the present study is based on Daller, Van Hout, & Treffers-Daller (2003), who developed a measure named "Guiraud Advanced Index." Guiraud Advanced Index was defined based on the frequency of lexical entry. In this sense, the measure is a combination of qualitative (i.e., frequency) and quantitative measures. In setting the frequency criteria, Daller et al. (2003) referred to Oehler's basic word list for German as a foreign language (Oehler, 1983) and counted word types and tokens, whose frequencies were ranked higher than, or equivalent to, the 2000-frequency level. In the present study, "JACET List of 8000 Basic Words" was used (JACET is the abbreviation for the Japan Association of College English Teachers), which was developed by the Committee of Revising JACET List of Basic Words (2003) and described by Uemura and S. Ishikawa (2004). Analyses were conducted by means of Shimizu's (2004) software program, which automatically attaches frequency levels defined by JACET 8000 to the inputted words. Contracted words were counted as two words (e.g., "wanna" was processed as "want" and "to"), and proper nouns (e.g., participants' names and the office members' names) were excluded from the analyses.

Finally, in the present study, one accuracy measure was employed; namely, the percentage of error-free T-unit (EFT). The measure was used by previous studies on task complexity effects (e.g., Gilabert, 2007; Ishikawa, 2007; Robinson, 1995). The formula for EFT was: total number of error-free T-units divided by total number of T-units multiplied by 100. This is a global measure of accuracy.

The coding of the speech data was conducted by two English native-speakers with MAs in TESOL and me. The inter-rater reliabilities of the measures, using 25% of the samples, reached above 90% consistencies.

RESULTS

Descriptive statistics for the means and standard deviations of the six dependent variables are given in Table 1. In order to confirm that the scores were distributed normally Shapiro-Wilk tests were applied to the six dependent variables. The test results showed that the distributions were not deviant from normal distributions ($p > .05$).

TABLE 1
Descriptive Statistics of the L2 Production Measures of Fluency, Complexity, and Accuracy of the Three Task Conditions

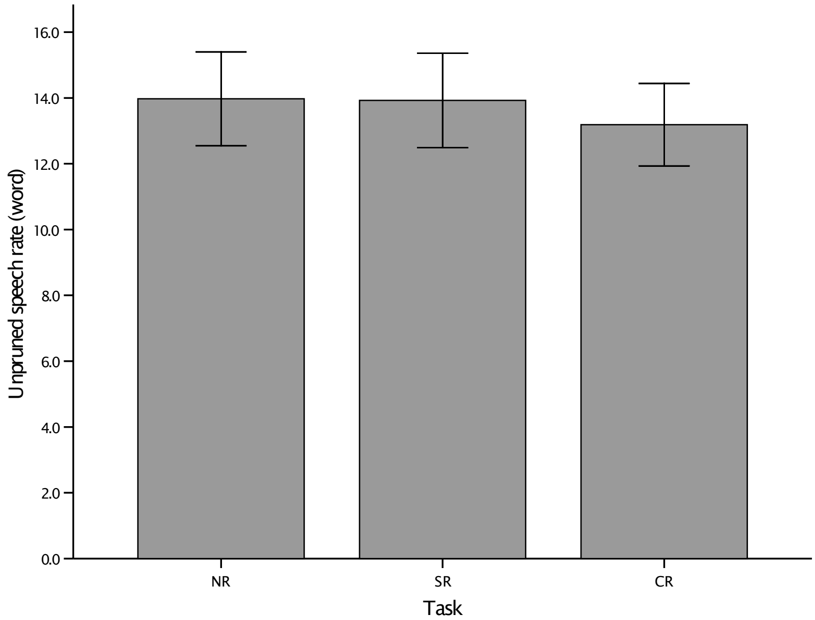
	No Reasoning		Simple Reasoning		Complex Reasoning	
Production measures	*M*	*SD*	*M*	*SD*	*M*	*SD*
Un-pruned speech rate	13.974	3.495	13.927	3.517	13.188	3.073
Pruned speech rate	11.832	3.245	11.654	3.176	10.859	2.825
Disfluency	1.209	.628	1.313	.759	1.440	.863
S-nodes per T-unit	1.230	.216	1.382	.222	1.334	.212
Guiraud 2000	1.836	.470	2.605	.404	2.465	.574
Error-free T-unit	25.932	17.659	39.012	14.688	35.824	18.492

Subsequently, an analysis based on a repeated measures multivariate analysis of variance (MANOVA) was conducted in order to test the general effect of intentional reasoning demands on L2 production. The within-subject variable was Task Complexity and the dependent variables were the six L2 production measures. The results of the repeated measure MANOVA indicated that the main effect for Task Complexity was statistically significant ($F(14, 82) = 4.681$, $p = .001$, $\eta^2 =.343$). The sections to follow focus on the seven hypotheses based on planned comparisons, using repeated measures of analyses of variance (ANOVAs).

Hypothesis 1: The Effect of Intentional Reasoning Demands on Un-Pruned Speech Rate

Hypothesis 1 stated that intentional reasoning demands would reduce L2 fluency relative to the No Reasoning Task, as reflected in the measure of un-pruned speech rate. Figure 1 presents the means of un-pruned rate across the three task conditions.

FIGURE 1
Means of un-pruned speech rate in the No Reasoning (NR) Task, the Simple Reasoning (SR) Task, and the Complex Reasoning (CR) Task.

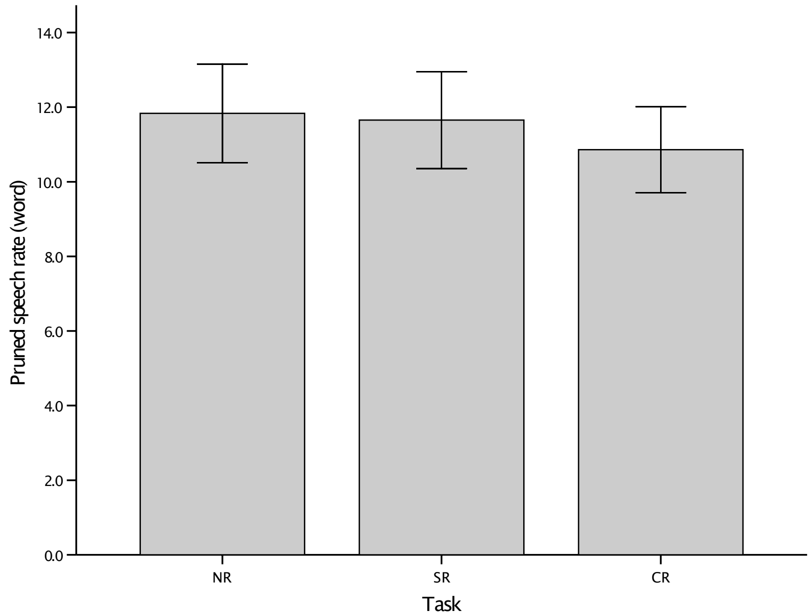


In order to test Hypothesis 1, a planned comparison between the No Reasoning Task and the two reasoning tasks was conducted. The results were not statistically significant ($F(1, 23) = 988$, $p = .330$), indicating that Hypothesis 1 was disconfirmed.

Hypothesis 2: The Effect of Intentional Reasoning Demands on Pruned Speech Rate

Hypothesis 2 stated that intentional reasoning demands would reduce L2 fluency relative to the No Reasoning Task, as reflected in the measure of pruned speech rate. Figure 2 shows the means of the three task conditions.

FIGURE 2
Means of pruned speech rate in the No Reasoning (NR) Task, the Simple Reasoning (SR) Task, and the Complex Reasoning (CR) Task.

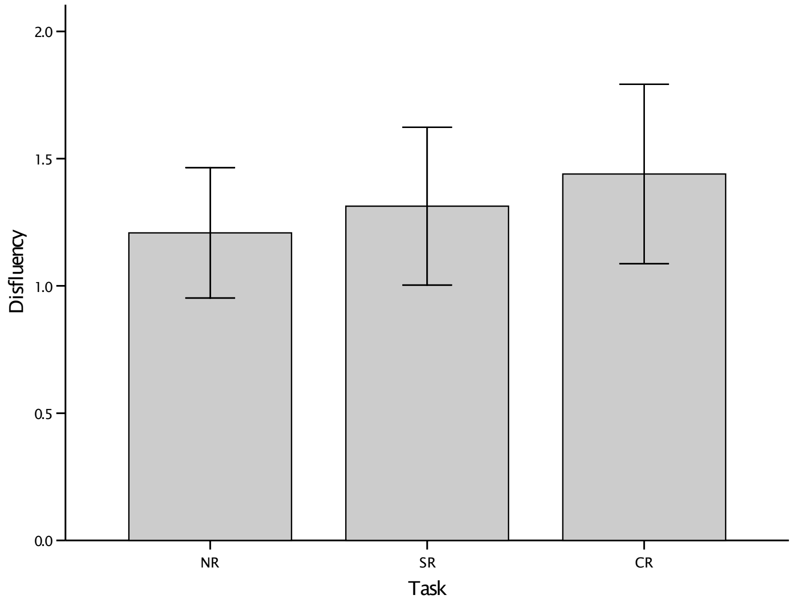


A planned comparison between the No Reasoning Task and the two intentional reasoning tasks was conducted in order to test Hypothesis 2. The result of the planned comparison was not significant ($F(1, 23) = 1.929$, $p = .178$), and Hypothesis 2 was disconfirmed.

Hypothesis 3: The Effect of Intentional Reasoning Demands on Disfluency

Hypothesis 3 stated that intentional reasoning demands would reduce L2 fluency relative to the No Reasoning Task, as reflected in the measure of disfluency. As described previously, the measure of disfluency indicates the number of disfluency episodes per pruned 10 words. Figure 3 presents the means of the three task conditions.

FIGURE 3
Means of disfluency in the No Reasoning (NR) Task, the Simple Reasoning (SR) Task, and the Complex Reasoning (CR) Task.

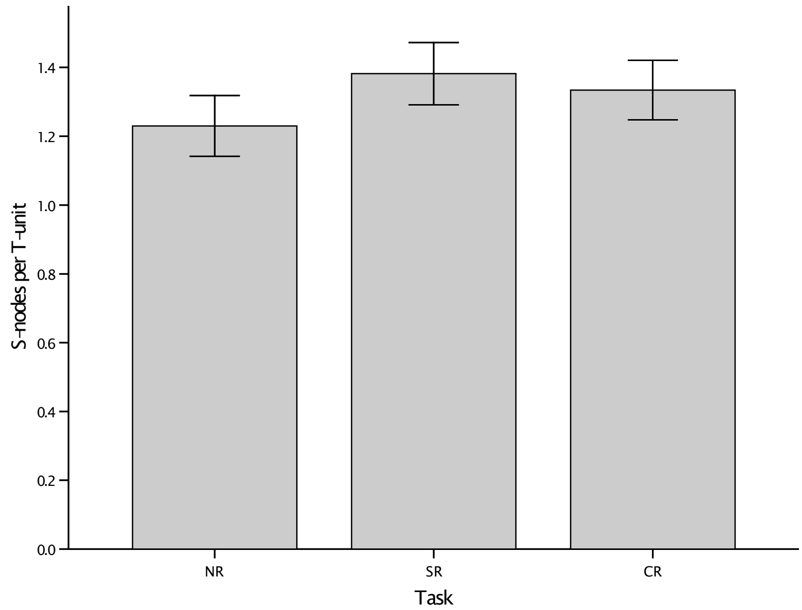


A planned comparison between the No Reasoning Task and the two reasoning tasks showed that the contrast was significant (F(1, 23)=3.949, p=.049, η^2 = .159). Thus, Hypothesis 3 was confirmed.

Hypothesis 4: The Effect of Intentional Reasoning Demands on Syntactic Complexity

Hypothesis 4 predicted that intentional reasoning demands would enhance L2 complexity relative to the No Reasoning Task, as reflected in the measure of S-nodes per T-unit. Figure 4 presents the means of the three task conditions.

FIGURE 4
Means of S-nodes per T-unit in the No Reasoning (NR) Task, the Simple Reasoning (SR) Task, and the Complex Reasoning (CR) Task.

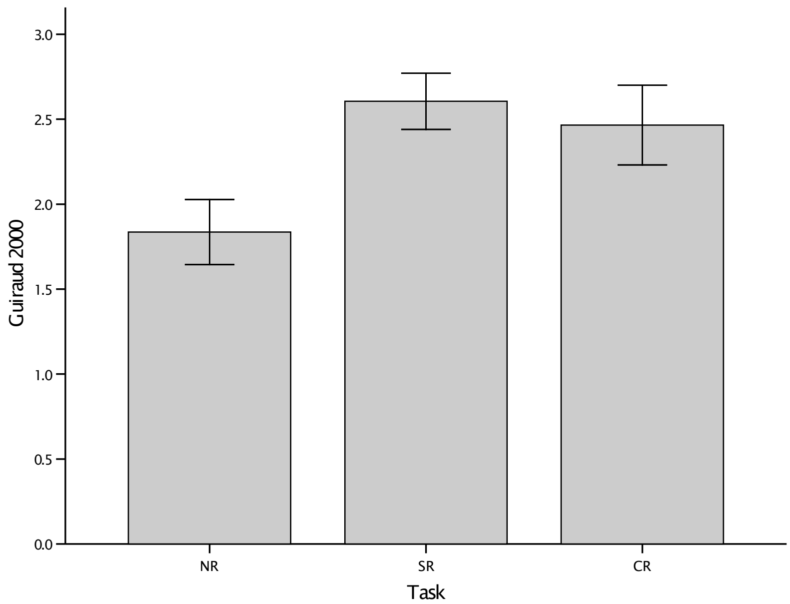


A planned comparison indicated that intentional reasoning enhanced L2 complexity relative to the No Reasoning Task statistically significantly (F(1, 23)=8.555, p=.008, η^2=.271). The results thus supported Hypothesis 4.

Hypothesis 5: The Effect of Intentional Reasoning Demands on Lexical Complexity

Hypothesis 5 stated that reasoning demands would enhance L2 complexity relative to the No Reasoning Task, as reflected in the measure of Guiraud 2000, a type-token measure. Figure 6 shows the means of the three task conditions.

FIGURE 5
Means of Guiraud 2000 in the No Reasoning (NR) Task, the Simple Reasoning (SR) Task, and the Complex Reasoning (CR) Task.



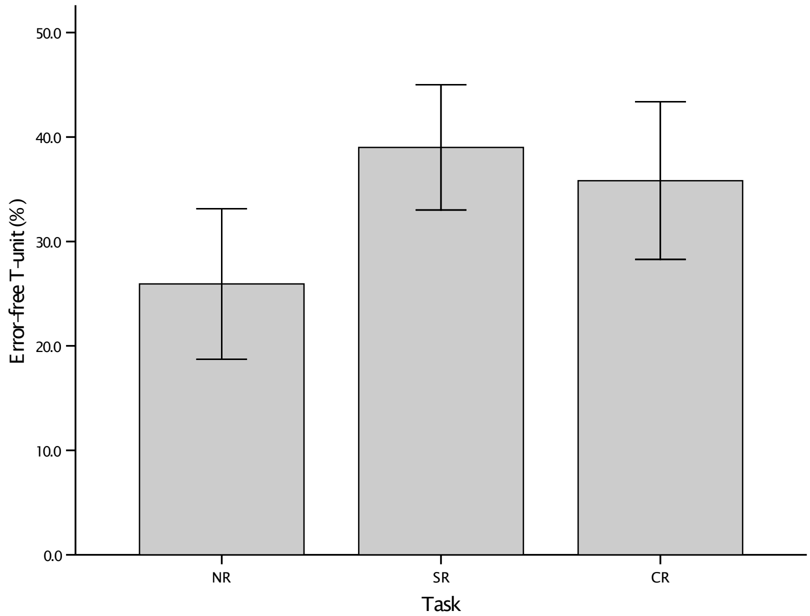
A planned comparison was conducted in order to test Hypothesis 5. The results were statistically significant ($F(1, 23) = 59.737$, $p < .01$, $\eta^2 = .722$), showing that intentional reasoning demands enhanced lexical complexity relative to the No Reasoning Task. Thus, Hypothesis 5 was confirmed.

Hypothesis 6: The Effect of Intentional Reasoning Demands on Percentage of Error-free T-units

Finally, Hypothesis 6 stated that intentional reasoning demands would enhance L2 accuracy relative to the No Reasoning Task, as reflected in the measure of Error-free T-unit. Figure 6 shows the means of the three task conditions.

FIGURE 6

Means of percentage of error-free T-unit in the No Reasoning (NR) Task, the Simple Reasoning (SR) Task, and the Complex Reasoning (CR) Task.



In order to test Hypothesis 6, a planned comparison was conducted between the No Reasoning Task and the two reasoning tasks. The results indicated that reasoning demands enhanced global accuracy relative to the No Reasoning Task statistically significantly ($F(1, 23)=12.168$, $p=.002$, $\eta^2=.346$). Hypothesis 6 was thus supported. The results reported in this section are summarized in Table 2.

TABLE 2
Summary of the Findings of the Present Study

L2 production category	Hypothesis	Result
Fluency	Hypothesis 1: *Intentional reasoning will reduce L2 fluency relative to the No Reasoning Task, as reflected in the measure of un-pruned speech rate.*	Disconfirmed
	Hypothesis 2: *Intentional reasoning will reduce L2 fluency relative to the No Reasoning Task, as reflected in the measure of pruned speech rate.*	Disconfirmed
	Hypothesis 3: *Intentional reasoning will reduce L2 fluency relative to the No Reasoning Task, as reflected in the measure of disfluency.*	Confirmed
Complexity	Hypothesis 4: *Intentional reasoning will enhance L2 complexity relative to the No Reasoning Task, as reflected in the measure of S-nodes per T-unit.*	Confirmed
	Hypothesis 5: *Intentional reasoning will enhance L2lexical complexity relative to the No Reasoning Task, as reflected in the measure of Guiraud 2000, a type-token ratio.*	Confirmed
Accuracy	Hypothesis 6: *Intentional reasoning will enhance L2 complexity relative to the No Reasoning Task, as reflected in the measure of error-free T-unit.*	Confirmed

DISCUSSION

In order to answer the research question, *Does intentional reasoning lead to greater accuracy and complexity at the cost of fluency?,* six hypotheses were formulated and tested in the present study. As we have seen above, the results indicated that intentional reasoning led to increases in accuracy and complexity at the cost of fluency in the sense of disfluency, but not in the sense of speech rate. Those results were largely compatible with the predictions of the Cognition Hypothesis.

The two measures of L2 speech rate, however, did not show predicted negative effects of intentional reasoning on L2 fluency. Several accounts may be possible, but perhaps the most likely account for the failures to confirm the hypotheses is related to the provision of planning time. Reduced planning time may produce a significant effect, especially in the case of pruned speech rate, given the observed low probability level (i.e., $p = .110$) and the direction of the effect.

Despite the lack of the predicted effects of intentional reasoning on the speech rate measures, the results were overall compatible with the predictions of the Cognition Hypothesis and they have several theoretical implications. For instance, some authors such as Skehan (1998) and VanPatten (1996) argue that attention to meaning leads to reduced attention to form due to attentional capacity limitation. This theoretical position assumes a friction between attention to meaning and form, where standing chances of the marriage of meaning and form are somewhat underestimated. There are of course other positions, however, one of which is the "focus on form" approach (e.g., Doughty & Williams, 1998; Long & Robinson, 1998; Muranoi, 2000; Williams, 2005). This position does not posit necessary trade-offs between attention to meaning and attention to form. Rather meaningful communication is seen as motivational sources for attention to form. This means that meaning is seen as potential for language development rather than hindrance to L2 processes. This claim of course does not mean that attention to form is possible under *any* condition. Besides potential developmental readiness and individual differences in cognitive abilities relevant to focus on form, interference and confusion are likely factors that would hinder efficient attention to form (e.g., Robinson, 2003). As we have already seen, potential task complexity dimensions that are predicted to cause interference and confusion are grouped into what Robinson calls resource-*dispersing* dimensions. Clearly, from both theoretical and practical perspectives, specifying under what *conditions* attention to form is facilitated or debilitated is of more importance (e.g., lack of planning time and prior knowledge, see Robinson 2007, for further specifications) than simply ascribing performance decrements to the *theoretical construct* of attentional capacity limitation. In other words, simply assuming inherent attentional capacity is nothing less than duplicating terminology without explaining *why* attention to form is debilitated, which, as Neumann (1996) points out, is tautological. The results of the present study were to a greater extent in line with the focus on form position and needless to say with the Cognition Hypothesis, supporting the view that attention through meaning (e.g., Samuda, 2001) is possible and also is a viable methodological principle of TBLT.

From a different perspective, attention to form through meaning also seems to be in line with disfluency results. Fluency and attention to form are at least conceptually distinct, but it is not clear whether they are independent at all. In other words, a question arises regarding whether fluency is just about "fluency." The agreed definition of what we mean by fluency is problematic (e.g., Riggenbach, 2000); however, Hieke's (1981) notion of hesitations as signs of "quality control" for one's output does suggest a link between disfluency and learner efforts toward better output (whatever means by "better" here) at the expense of articulatory fluency.

This interpretation is also compatible with the view that language use is characterized by incessant workings of strategic competence, i.e., assessment, planning, and execution, in order to attain task goals (Bachman, 1990). From this perspective, it can be said that the speaker's strategic competence was gauged towards attention to form in achieving more complex communicative goals. Thus, the workings of executive control over attentional allocation to form through meaning led to the simultaneous positive effects on accuracy and complexity at the cost of fluency in the sense of disfluency. The results of the present study suggested that it was the cognitively simple task that invited the participants to use less accurate and complex language in a relatively fluent manner presumably because communication was assessed to be possible by means of those less developed forms (i.e., Givon's notion of pragmatic mode).

Robinson (2005, 2007a) states that intentional reasoning prompts the learner to go beyond propositional information transmission, encouraging them to make inferences and produce utterances. Inferential demands invite discourse-pragmatic demands (e.g., Givon, 1993, 1998) because ordering of information needs to be taken into account by manipulating concepts in order to guide reasoning processes. Such discourse-pragmatic demands require the text to be organized hierarchically in a top-down fashion (e.g., Berman & Slobin, 1994), pushed output (e.g., Swain, 1995) motivates the learner to focus on crucial information in a bottom-up way, or allow enhanced interactions of both discourse processing modes. In the No Reasoning Task, in contrast, since the order of propositional information transmission was relatively unconstrained (e.g., the relationship between X and Y is good/bad and so on), propositional information could be transmitted without heavy discourse-pragmatic demands; accordingly, demands on discourse coherence were also reduced. The results of the present study showed that intentional reasoning led to greater syntactic complexity, indicating that discourse-functional complexity accompanied syntactic complexity, as claimed by Givon (1979, 1998), Berman and Slobin (1994), and Robinson (2005) and led to pushed output and attention allocation to language form as claimed by Swain (1995) at the cost of disfluency.

LIMITATIONS OF THE PRESENT STUDY

The present study has, as any study does, several limitations to be acknowledged. First, the L2 production measures used in the study may not be the "best" ones. This of course does not mean that the results obtained are insignificant, but the intended point is that the obtained results were just part of the whole picture. For one thing, the L2 production measures used in the present study were all general descriptive measures. This means that as Wolfe-Quintero *et al.* (1998) mention, while those general measures are useful, they do not provide specific information regarding the nature of the language produced by learners. The issue of content validity of the L2 production metric needs to be addressed in future analyses on the current production data.

Furthermore, even data examinations by means of specific measures may not address the whole issue. This point needless to say relates to Bley-Vroman's (1983) warning of "the comparative fallacy." As Larsen-Freeman and Long (1991) point out, even the notion of "errors" can be accused out of the exclusive reliance on the target-language norm rather than assessing from developmental perspectives and the accuracy level may draw well-known U-shaped behaviors during the course of restructuring. Once such developmental perspectives are taken, we also need to address the issue of cross-linguistic influences on the effect of task complexity as well as one developmental trend on another (e.g., relations between discourse-syntactic development and inflectional morphology development). Some of the interesting questions are beginning to be addressed, for instance, by Nakamura (2007) from a typological perspective and Robinson (2007b) by considering the use of psychological state terms within the context of intentional reasoning.

Finally, the present study did not consider potential influences of the learner's individual differences on task complexity effects. This limitation is related to the view that the learner's individual differences factors are likely to mediate the effect of task complexity, which could be either facilitative or debilitative. For instance, Niwa (2000) showed that the learner's working memory capacity played important roles as task complexity was increased. Similarly, Robinson (2007b) also showed that the effect of task complexity was in part dependent on the level of output anxiety (MacIntyre & Gardner, 1994) in that facilitative task complexity effects were present only in less anxious learners but not in anxious learners. Investigations into the role played by the learner's individual differences will be of crucial importance, not only because the effect of task complexity may be dependent on them, but also because interactions between task complexity and learner characteristics can define teacher roles in classroom practice.

ACKNOWLEDGMENTS

I thank Professor Peter Robinson, Matsuo Kimura, Harumi Akimoto at Aoyama Gakuin University, Tokyo, and Fujiko Sano at Yokohama National University, Kanagawa, for their very advisable comments on the original study conducted for my Ph.D. study. Thanks are also due to Dr. Tomoko

Takahashi, Mr. Jonathan Epstein, Dr. Terrie Mathis, Dr. Robert Agajeenian, my M.A. teachers at Soka University of America, CA, for their consistent support and encouragement. I also thank twenty-four anonymous participants for their participations to the study. Thanks are also due to Maiko Katherine Nakano and Keiko Samantha Shimizu at Soka Women's College, Tokyo, for their transcript-checking and data-coding assistance, Dr. Elizabeth Shriberg at Speech Technology & Research (STAR) Laboratory, CA, who gave me useful advice on disfluency analysis and coding, and Dr. Francesc Tarres at the Universitat Politècnica de Catalunya, Spain, who allowed me to use his face database. All errors remain my responsibility. Finally, I would also thank Dr. Daisaku Ikeda, Founder of Soka Women's College and Soka University, for his warm constant encouragement.

THE AUTHOR

Tomohito Ishikawa graduated from Aoyama Gakuin University in March 2008 and is to be lecturer in the Dept. of English Communication at Soka Women's College, Tokyo, in April 2008. His current research interests cover task-based language teaching. His recent publications include The Effect of Increasing Task Complexity along the [±Here-and-Now] Dimension on L2 Written Narrative Discourse with Ishikawa (2007) and The Effect of Task Complexity and Language Proficiency on Task-Based Language Performance with Ishikawa (2006).
Email: itomohitos@yahoo.co.jp

REFERENCES

Bachman, L. (1990). *Fundamental considerations in language testing.* Oxford: Oxford University Press.

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

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

Celce-Murcia, M., & Olshtain, E. (2005). Discourse-based approaches: A new framework for second language teaching and learning. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (pp. 729-741). Mahwah, NJ: Erlbaum.

Clark, H., & Fox Tree, J. (2002). Using uh and um in spontaneous speaking. *Cognition, 84,* 73-111.

Daller, H., van Hout, R., & Treffers-Daller, J. (2003). Lexical richness in the spontaneous speech of bilinguals. *Applied Linguistics, 24,* 197-222.

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

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

Gilabert, R. (2007). The simultaneous manipulation of task complexity along planning time and [+/- Here-and-Now]. In Maria del Pilar Garcia Mayo (Ed.), *Investigating tasks in formal language learning* (pp. 44-68). Philadelphia, PA: Multilingual Matters.

Givon, T. (1979). *On understanding grammar.* NY: Academic Press.

Givon, T. (1985). Function, structure, and language acquisition. In D. Slobin (Ed.), *The crosslinguistic study of language acquisition: Vol. 2* (pp. 1008-1025). Hillsdale, NJ: Erlbaum.

Givon, T. (1998). The functional approach to grammar. In M. Tomasello (Ed.), *The new psychology of language: Cognitive functional approaches to language structure* (pp. 41-66). Hilsdale, NJ: Erlbaum.

Goldman-Eisler, F. (1968). *Psycholinguistics: Experiments in spontaneous speech.* London: Academic Press.

Hieke, A. (1981). A content-processing view of hesitation phenomena. *Language and Speech, 24,* 147-160.

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

Ishikawa, T. (2006). The effect of task complexity and language proficiency on task-based language performance. *The Journal of Asia TEFL, 3,* 193-225.

Ishikawa, T. (2007). The effect of increasing task complexity along the [±Here-and-Now] dimension on L2 written narrative discourse. In Maria del Pilar Garcia Mayo (Ed.), *Investigating tasks in formal language learning* (pp. 136-156). Philadelphia, PA: Multilingual Matters.

Ishikawa, T. (2008). *Investigating the effect of intentional reasoning demands on L2* *speech production.* Doctoral dissertation, Aoyama Gakuin University, Tokyo.

Kumaravadivelu, B. (1994). The postmethod condition: Emerging strategies for second/foreign language teaching. *TESOL Quarterly, 28,* 27-48.

Larsen-Freeman, D., & Long, M. (1991). *An introduction to second language acquisition research.* London: Longman.

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

Lickley, R. (1998). *HCRC Disfluency coding manual.* HCRC Technical Report. HCRC/TR-100.

Long, M. (1985). A role for instruction in second language acquisition. In K. Hyltenstam & M. Pienemann (Eds.), *Modeling and assessing second language acquisition* (pp. 77-99). Clevedon, Avon: Multilingual Matters.

Long, M. (1989). Task, group and task-group interactions. *University of Hawaii Working Papers in ESL, 8,* 1-26.

Long, M. (Ed.) (2006a). *Second language needs analysis.* Cambridge: Cambridge University Press.

Long, M., & Crookes, G. (1992). Three approaches to task-based syllabus design. *TESOL Quarterly, 26,* 27-47.

Long, M., & Robinson, P. (1998). Focus on form: Theory, research, and practice. In C. Doughty & J. Williams (Eds.), *Focus on form in classroom second language acquisition* (pp. 15-41). Cambridge: Cambridge University Press.

MacIntyre, P., & Gardner, R. (1994). The subtle effects of language anxiety on cognitive processing in the second language. *Language Learning, 44,* 283-305.

Muranoi, H. (2000). Focus on form through interaction enhancement: Integrating formal instruction into a communicative task in EFL classrooms. *Language Learning, 50,* 352-60.

Nakamura, D. (2007). Motion events, linguistic typology and task complexity in L2 narrative production. In P. Robinson (Chair), *Task complexity, language learning and language performance: Current research issues.* Symposium conducted at the meeting of the AAAL Annual Conference, Costa Mesa, CA.

Neumann, O. (1996). Theories of attention. In O. Neumann & A. F. Sanders (Eds.), *Handbook of perception and action, Vol. 3: Attention* (pp. 389-346).

Niwa, Y. (2000). *Reasoning demands of L2 tasks and L2 narrative production: Effects of individual differences in working memory, intelligence and aptitude*. Unpublished master's thesis, Aoyama Gakuin University, Tokyo.

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

Oehler, H. (1983). Grund- und aufbauwortschatz Deutsch (Deutsch-Türkisch). München:

Klett.

Ortega, L. (2005). What do learners do?: Learner-driven attention to form during pre-task planning. In R. Ellis (Ed.), *Planning and Task Performance in a Second Language* (pp. 77-109). Amsterdam: Benjamins.

Richards, J., & Rodgers, T. (2001). *Approaches and methods in language teaching*. Cambridge: Cambridge University Press.

Riggenbach, H. (2000). *Perspectives on fluency.* Ann Arbor, MI: University of Michigan Press.

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

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. (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. (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. (2007a). Criteria for classifying and sequencing pedagogic tasks. In Maria del Pilar Garcia Mayo (Ed.), *Investigating tasks in formal language learning* (pp. 7-26). Philadelphia, PA: Multilingual Matters.

Robinson, P. (2007b). Giving reasons and directions in a second language: The influence of task sequence and complexity on uptake of premodified input. In P. Robinson (Chair), *Task complexity, language learning and language performance: Current research issues.* Symposium conducted at the meeting of the American Associations of Applied Linguistics 2007 Annual Conference. Costa Mesa, CA.

Samuda, V. (2001). Guiding relationships between form 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, UK: Longman.

Shimizu, S. (2004). *V8an - revised web edition [computer program].* Retrieved February 7, 2007, from http://jacetvoc.kl.dendai.ac.jp/~shimizu/j8web/j8web.cgi

Shriberg, E. (1994). *Preliminaries to a theory of speech disfluencies.* Unpublished doctoral dissertation, University of California, Berkeley.

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

Swain, M. (1995). Three functions of output in second language acquisition. In G. Cook & B. Seidlhoffer (Eds.), *Principle and practice in applied linguistics: Studies in honour of H. G. Widdowson* (pp. 125-144). Oxford: Oxford University Press.

Uemura, T., & Ishikawa, S. (2004). JACET 8000 and Asia TEFL vocabulary initiative. *The Journal of Asia TEFL, 1*, 334-347.

Van den Branden, K. (Ed.) (2006a). *Task-based language education: From theory to practice.* Cambridge: Cambridge University Press.

VanPatten, B. (1996). *Input processing and grammar instruction in second language acquisition.* Norwood, NJ: Ablex.

Wilkins, D. (1990). Second languages: How they are learned and taught. In N. E. Collinge (Ed.), *An encyclopaedia of language* (pp. 518-550). London: Routledge.

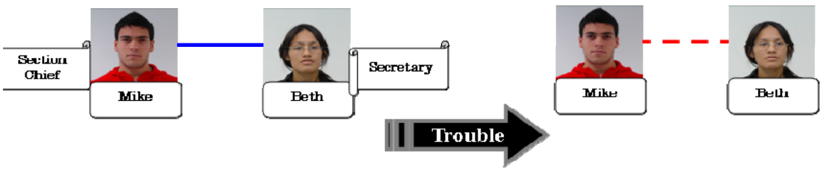
Williams, J. (2005). Form-focused instruction. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (pp. 671-691). Mahwah, NJ: Erlbaum.

Willis, J. (1996). *A Framework for task-based learning*. Harlow, Essex: Pearson.

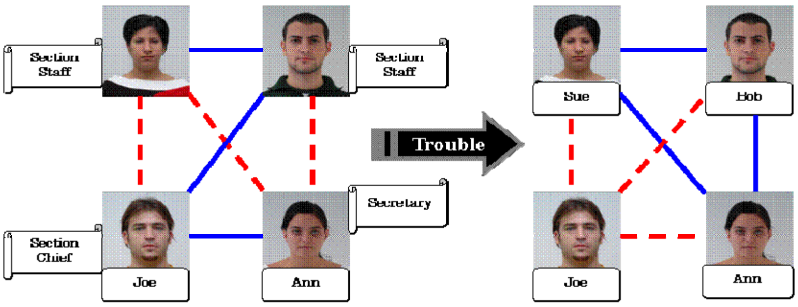
Wolfe-Quintero, K., Inagaki, S., & Kim, H. (1998). *Second language development in writing: Measures of fluency, accuracy, and complexity.* University of Hawaii: Second Language Teaching and Curriculum Center.

APPENDICES

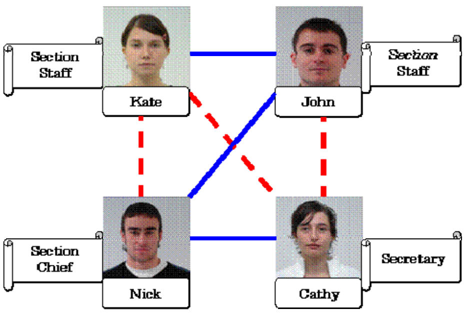
Appendix A. Simple Reasoning Task



Appendix B. Complex Reasoning Task



Appendix C. No Reasoning Task



Appendix D. Instructions for the No Reasoning Task

In this task, you will play a company's general manager's role. Today you are supposed to report to the president about the human relations of the section, which was established a week ago. In order to report to the president, you decided to make a phone call but the president was out of home. Now, you decided to leave a message on his answer machine.

Performance goal: As general manager, report to the president in English about the human relations of the new section by leaving a massage on the president's answer machine.

Below is the explanation of the task sheet.

You will be given a task sheet. The task sheet contains the following information:

- Persons' names, pictures, and positions in the section.
- Links between the section members' pictures: Good human relations are represented in blue solid lines and non-good human relations in red broken lines.

human relations,
names, pictures,
and positions

Information on the task sheet

When you report to the president, please keep in mind the following points

- Describe the current human relationships of the office members to the president. Assume that the president does not have any information about the section.

If you have questions, you can ask now. If no questions, please read "Performance goal" one more time. You have three-minute planning time. Also note that the maximum time length for recording of the answering machine is three minutes.

Appendix E. Instructions for the Simple and Complex Reasoning Task

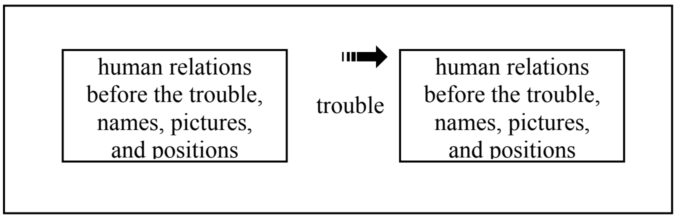
In this task, you will play a company's general manager's role. Today you are supposed to report to the president about the human relations of the section, which was established a week ago. But yesterday trouble happened and human relationships of the section members changed. In order to report to the president, you decided to make a phone call but the president was out of home. Now, you decided to leave a message on his answer machine.

Performance goal: As general manager, report to the president in English about the trouble and changes in human relations by leaving a massage on the president's answer machine.

Below is the explanation of the task sheet.

You will be given a task sheet. The task sheet contains the following information:

- Arrow (→), which represents unspecified processes of human relationship changes.
- Persons' names, pictures, and positions in the section.
- Links between the section members' pictures: Good human relations are represented in blue solid lines and non-good human relations in red broken lines.



Information on the task sheet

When you report to the president, please keep in mind the following points

- You will receive a list of job mistakes. Choose just one mistake, which is a trouble trigger not the direct cause of the human relationship changes. Assign the selected trouble trigger to one person.
- Please explain processes of human relationship changes that you come up with. Assume that the president does not have any information about the section.

If you have questions, you can ask now. If no questions, please read "Performance goal" one more time. When you finish double-checking the performance goal, you will receive the task sheet and a list of mistakes. You have three-minute planning time. Also note that the maximum time length for recording of the answering machine is three minutes.