ISSN 0910-500X

英文學思潮

THOUGHT CURRENTS IN ENGLISH LITERATURE

VOLUME LXXX

2007

THE ENGLISH LITERARY SOCIETY
OF
AOYAMA GAKUIN UNIVERSITY

青山学院大学英文学会

Sites They

Strongly disagree	Response total
2 (0.3)	738 (100)
2 (0.3)	737 (100)
10 (1.3)	744 (100)
6 (0.8)	751 (100)
5 (0.7)	735 (100)
2 (0.3)	740 (100)

The effect of intentional reasoning demands on second language speech production

Tomohito ISHIKAWA

The notion that pedagogic tasks affect the way the learner speaks a second language (L2) in qualitative terms has been increasingly popular among L2 researchers and teachers especially in recent years (e.g., Garcia-Mayo, 2007; Robinson, 2007a, 2007b; Skehan, 1998). What is required, under this circumstance, are a guiding framework and subsequent systematic investigations into what task parameters, conditions, and other individual differences factors contribute to, or relate to, the learner's L2 performance. The purpose of this article is to examine the effect of intentional reasoning demands on L2 speech production, a *cognitive* task complexity parameter which is proposed by Robinson (2007a).

Robinson's triadic framework for TBLT

Robinson (2007a) proposes a framework for task-based language teaching (TBLT), which realizes a three-way distinction between task condition, task complexity, and task difficulty. 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). According to Robinson (2007a), task condition is largely constrained by the target task, identified by needs analysis, and is to be maintained throughout task sequences. Maintaining task condition across the sequence is important, for instance, for transfer of training, motivation. The important point here with respect to syllabus design is that whereas task condition is a crucial factor if a series of pedagogic tasks are to bear a close similarity to a target task, it is not considered as playing a central role when it comes to task-sequencing decisions.

Second, *task complexity* refers to pedagogic tasks' information processing or demands on memory, attention, and reasoning (Robinson, 2001a, 2001b). This category of task component is characterized as cognitive in nature, and so represents inherent and relatively fixed processing demands of pedagogic tasks, which are manipulable by syllabus designers in a prospective way. As Robinson states, "[t]ask complexity is therefore the *sole* basis of pedagogic task sequencing" (Robinson, 2007a, p. 22, emphasis original).

Finally, 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). Concerning prospective task sequencing decisions, the learner's perceptions are impossible to assess prior to task implementation; therefore, they cannot be considered in a prospective manner.

Dual aspects of task complexity

According to Robinson (2001a, 2001b, 2005, 2007a), task complexity comprises two major categories of dimensions, with each dimension containing task complexity features that are cognitive in nature (e.g., [(± intentional reasoning]). The conceptual distinction between the two categories of dimensions is important, and sets of proposed task complexity features are predicted to influence language learning and performance differentially. Those dimensions are: the resource-directing and the resource-dispersing dimensions.

The resource-directing dimensions. The resource-directing dimensions have their essential characteristics of directing the learner's attention to various language codes (Robinson, 2001a, b, 2003, 2005, 2007a). Those dimensions are expressed by "±" features although these can form a continuum. By manipulating these features, the learner's attention is 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. For instance, when the task goal is to identify certain objects among a set of similar distracters (i.e., [+ few elements]), the learner's attention may be directed toward a set of relevant linguistic codes in order to distinguish the intended objects from competing ones (e.g., attributive adjectives or locative expressions depending on the history of discourse and the nature of the situational context). The fundamental idea here is that conceptual/functional demands of tasks help the learner to attend to linguistic codes, which is predicted to produce potential impacts on learning and expansion of L2 linguistic resources and so guide L2 development.

The resource-dispersing dimensions. In contrast to the resource-directing dimensions, increasing task complexity along the resource-dispersing dimensions does not necessarily draw attention to the language code; rather, increasing task complexity along the resource-dispersing dimensions is predicted to dissipate available attentional resources (i.e., dimensions to affect accessibility to L2 knowledge base). As such, manipulating them is predicted to affect the learner's performance efficiency. For instance, when the leaner is asked to solve an unfamiliar problem-solving task (i.e., [− prior knowledge]), compared with a familiar one (i.e., [+ prior knowledge]), more time is likely to be needed to access the knowledge base since the general knowledge base for solution is not well-established. Consequently, lack of fluency and errors due to attention lapses may be expected to be observed in unfamiliar tasks relative to familiar tasks.

Research questions

The purpose of the present study is to examine the effect of manipulating one of the proposed resource-directing dimensions on L2 monologic performance. 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 the psychological literature, intentional reasoning is also described as everyday psychological reasoning.

Robinson's Cognition Hypothesis (Robinson, 2005) predicts that intentional reasoning demands leads to increases in accuracy and complexity at the cost of fluency. This is partly because manipulating resource-directing dimensions is predicted to affect the allocation of attentional resources to language forms.

Intentional reasoning would create greater conceptual gaps that needed to be filled with additional thoughts and inferences. In the intentional reasoning task then, the learner needs to rely on inferences based on mindreading in order to cross conceptual gaps, relative to the task without such demands, where simple information transmission is required.

In linguistic terms, intentional reasoning imposes discourse-pragmatic demands on the part of the speaker, where multi-propositional considerations are required (Givon, 1998). For instance, 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). In contrast, in tasks without intentional reasoning demands, where transmission of simple propositional information is required and therefore is free from discourse-pragmatic demands.

Such coherence-establishing reasoning demands would decrease L2 fluency, as well, as Robinson claims. Delays in producing speech plans should lead to delays in speaking (e.g., Levelt, 1989). The speaker may hesitate or speak slowly. Intentional reasoning would also require incessant decision-making. Thus, intentional reasoning would impede efficient scheduling and smooth execution of L2 speaking.

In addition to the effect of intentional reasoning demands on L2 speech performance, the present study also examines the effect of intentional reasoning demands on the learner's perceptions of task difficulty and the relationship between the L2 speech performance and the learner's perceptions of task difficulty. According to Robinson (2005), increasing task complexity affects certain aspects of task difficulty (e.g., anxiety) and also produces stronger correlations between L2 performance and the learner's perceptions of task difficulty.

Based on the above considerations, the present study formulated the following research questions: (1) Does increasing intentional reasoning demands lead to increases in accuracy and complexity at the cost of fluency?; (2) What are the effects of intentional reasoning demands on the learner's perceptions of task difficulty?; (3) Do increasing task complexity lead to increases in the number of significant correlations between L2 task performance indices and the learner's perceptions of task difficulty?

Methodology for the present study

Participants

The participants of the present study were college students, whose L1 were 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.

Task 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).

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.

Finally, the No Reasoning Task (Appendix C) asked the manager simply to report 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 need to think up a trouble situation. In order to elicit sufficient amount of L2 data, a setting was chosen, where four section members were present 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 B) and the other for the Simple and the Complex Reasoning Tasks (Appendix B).

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 research 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.

After completing each task, the learner responded to brief written questionnaires, which were created to assess the learner's subjective perceptions of task difficulty. Shortly after the learner completed each task, he/she was asked to mark on those eight questionnaires in the form of ten-point Likert scales (from 0 to 9). The original questionnaires were written in the learners' L1 (i.e., Japanese).

The questionnaire items were of eight kinds: (1) *Concentration* ("This task required concentration----This task did not require concentration."); (2) *Time pressure* ("I did not feel time pressure during task performance----I felt time pressure during task performance."); (3) *Anxiety* ("This task made me anxious----This task did not make me anxious."); (4) *Frustration* ("I felt frustrated during task performance----I did not feel frustrated during task performance."); (5) *Difficulty* ("This task was easy----This task was difficult."); (6) *Interest* ("This task was interesting----This task was not interesting."); (7) *Ability* ("I did not do this task well----I did this task very well."); (8) *Motivation* ("I want to do tasks like this----I don't want to do tasks like this.").

Experimental design for the present study

The experimental design had two features. First, a within-subject experimental design was used. Second, the orders of task performance were counterbalanced by means of a complete Latin-Square experimental control.

L2 production measures

Fluency measures. The present study employed four fluency measures: un-pruned and pruned speech rate, disfluency, and hesitation length. First, the un-pruned speech rate measure was defined as total number of words produced divided by total speaking time in second multiplied by 10. In counting the number of words, hesitation markers such as "uhm" or "um" were also counted as words. Often the participant produced word fragments during hesitations (e.g., "sec-" for "section") and in those cases, word fragments were coded as .5. Regarding the measure of disfluency, it was operationalized as the total number of disfluency episodes divided by the total number of pruned words multiplied by 10. Thus, this is equivalent to total number of disfluency episodes (e.g., repetitions and false starts) per pruned 10 words. Finally, the measure of hesitation length was also used. The measure was defined as the number of disfluent words per disfluency episode.

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.

Next, the formula for the measure of 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. In this sense, the measure is a combination of qualitative (i.e., frequency) and quantitative measures.

Finally, in the present study, one accuracy measure was employed; namely, the percentage of error-free T-unit (EFT). 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.

Results

The effect of intentional reasoning demands on L2 performance

Table 1 presents the descriptive statistics for the eight L2 production measures as a function of task complexity. The means and standard deviations of the No, Simple, and Complex Reasoning Tasks are presented.

Table 1

Descriptive Statistics for the Eight L2 Production Measures as a Function of Task Complexity

	Task complexity					
	NR		SR		CR	
Production measure	*M*	*SD*	*M*	*SD*	*M*	*SD*
Fluency measure						
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	0.628	1.313	0.759	1.440	0.863
Hesitation length	1.514	0.437	1.488	0.408	1.570	0.461
Complexity measure						
S-nodes per T-unit	1.230	0.216	1.382	0.222	1.334	0.212
Guiraud 2000	1.836	0.470	2.605	0.404	2.465	0.574
Accuracy measure						
Error-free T-unit	25.932	17.659	39.012	14.688	35.824	18.492

Note. NR = No Reasoning Task; SR = Simple Reasoning Task; CR = Complex Reasoning Task

In order to test the effect of intentional reasoning demands on L2 performance, planned comparisons were employed, with the contrast being between the no versus intentional reasoning demands. The ANOVA planned comparison results are summarized in Table 2. Table 2 shows that the effect of intentional reasoning demands on L2 production was significant in the measures of Disfluency ($p<.05$), S-nodes per T-unit ($p<.01$), Guiraud 2000 ($p<.01$), and Error-free T-unit ($p<.01$), but not in the three fluency measures. These results showed that intentional reasoning demands led to increases in accuracy and complexity at the cost of speech disfluency.

Table 2

ANOVA Planned Comparison Summary for the Effect of Intentional Reasoning Demands on the Eight L2 Production Measures

Production measure	*df*	F	□^{2}
Fluency measure			
Un-pruned speech rate	(1, 23)	0.988	ns
Pruned speech rate	(1, 23)	1.929	ns
Disfluency	(1, 23)	4.336*	0.159
Hesitation length	(1, 23)	0.028	ns
Complexity measure			
S-nodes per T-unit	(1, 23)	8.555**	0.271
Guiraud 2000	(1, 23)	59.737**	0.722
Accuracy measure			
Error-free T-unit	(1, 23)	12.168**	0.346

Note. *p < .05, **p < .01

The effect of intentional reasoning demands on the learner perceptions of task difficulty

Table 3 presents the descriptive statistics of the eight task difficulty measures as a function of task complexity. In order to test the effect of intentional reasoning demands on the learner perceptions of task difficulty, the distributions of the eight task difficulty measures was considered under each task complexity condition separately. It was revealed that distributions of the task difficulty measures were non-normal; consequently, non-parametric, post hoc comparisons (i.e., Friedman tests) were applied to the data set.

Table 3

Descriptive Statistics for the Eight Measures of the Learner Perceptions of Task Difficulty as a Function of Task Complexity

	Task complexity					
	NR		SR		CR	
Task difficulty measure	*M*	*SD*	*M*	*SD*	*M*	*SD*
Concentration	7.000	2.449	7.833	1.404	8.250	1.567
Time Pressure	3.792	3.120	4.750	2.878	5.167	2.869
Anxiety	6.542	2.536	7.000	1.956	7.333	1.971
Frustration	6.875	2.112	7.042	2.010	7.875	1.329
Difficulty	6.208	2.167	6.542	2.043	7.667	1.308
Interest	6.375	2.516	6.750	1.939	7.167	1.926
Ability	2.167	2.160	1.917	2.412	1.333	1.606
Motivation	6.333	2.854	6.375	2.841	6.333	2.839
Concentration	7.000	2.449	7.833	1.404	8.250	1.567
Time Pressure	3.792	3.120	4.750	2.878	5.167	2.869

Table 4 presents the rank order means of the task difficulty measures as a function of task complexity. As Table 4 shows, raw mean orders of most of the task difficulty measures increased as task complexity increased. The exceptions were found in the cases of Ability and Motivation, but Ability rank order means decreased as task complexity increased, which was expected. The results of Motivation were also expected since Robinson (2001a) also reports similar tendency. These tendencies are also presented in Figure 1.

Table 4

Mean Rank orders of the Task Difficulty Measures as a Function of Task Complexity

Measure	Mean rank in NR	Mean rank in SR	Mean rank in CR
CON	1.71	1.94	2.35
TMP	1.75	1.98	2.27
ANX	1.90	1.96	2.15
FRU	1.77	1.92	2.31
DIF	1.75	1.79	2.46
INT	1.81	1.83	2.35
ABL	2.25	1.98	1.77
MTV	1.98	1.92	2.10

Note. NR = No Reasoning Task; SR = Simple Reasoning Task; CR = Complex Reasoning Task.

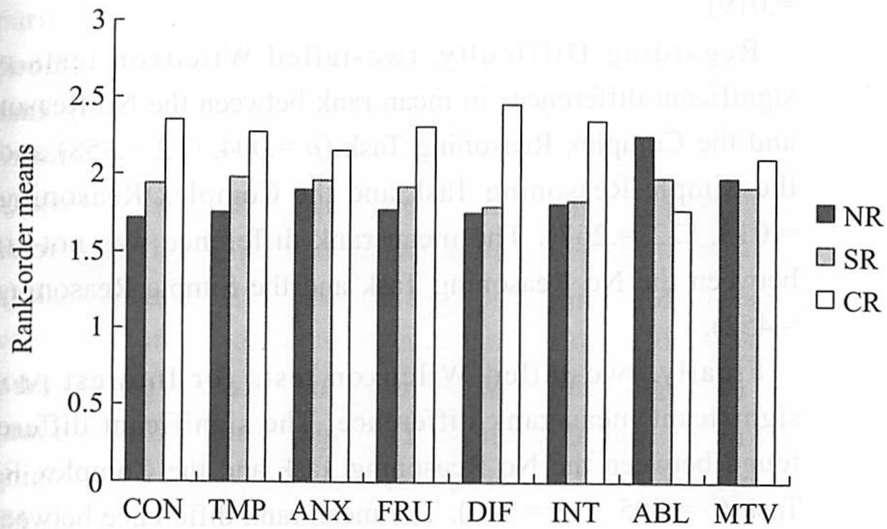


Figure1. Rank-order means of the eight task difficulty indices across the three task conditions. NR = No Reasoning Task; SR = Simple Reasoning Task; CR = Complex Reasoning Task; CON = Concentration; TMP = Time Pressure; ANX = Anxiety; FRU = Frustration; DIF = Difficulty; INT = Interest; ABL = Ability; MTV = Motivation.

Friedman tests were conducted to assess if there were differences in the mean ranks of the task difficulty responses across the three task conditions. As Table 5 shows, the Friedman tests revealed that significant results were found in the cases of Concentration (p = .011), Difficulty (p = .007), and Interest (p = .035). The mean rank differences of the other task difficulty indices did not reach significance although the result of Frustration was marginally significant (p = .075).

Subsequently, Friedman tests were followed up with two-tailed Wilcoxon tests. The alpha was set at the probability level of p = .0167 as a result of Bonferroni correction (i.e., .05 divided by 3). The results of the two-tailed Wilcoxon test for Concentration, however, did not detect significant differences: the No Reasoning Task and the Simple Reasoning Task (z = -1.391, p =.164); the Simple Reasoning Task and the Complex Reasoning Task (z = -1.876, p =.061); the No Reasoning Task and the Complex Reasoning Task (z = -2.344, p =.019).

Regarding Difficulty, two-tailed Wilcoxon tests produced significant differences in mean rank between the No Reasoning Task and the Complex Reasoning Task (p =.004, □2 =.358) and between the Simple Reasoning Task and the Complex Reasoning Task (p =.013, □2 =.267). The mean rank difference was not significant between the No Reasoning Task and the Simple Reasoning Task (p =.452).

Finally, two-tailed Wilcoxon tests for Interest produced a significant mean rank difference. The significant difference was found between the No Reasoning Task and the Complex Reasoning Task (p =.005, □2 =.342). The mean rank difference between the No Reasoning Task and the Simple Reasoning Task was not significant (p =.942), nor was the mean rank difference between the Simple Reasoning Task and the Complex Reasoning Task (p =.228).

Table 5
Summary Table of the Friedman Test Results

Measure	df	$\square^2$	$\square^2$	Post-hoc result
CON	2	8.982*	.127	CR = SR = NR
TMP	2	3.877	ns	--
ANX	2	1.099	ns	--
FRU	2	5.171	ns	--
DIF	2	9.838**	.122	CR > SR, NR
INT	2	6.677*	.094	CR > NR; CR = SR; SR = NR
ABL	2	4.586	ns	--
MTV	2	1.235	ns	--

Note. * p<.05. ** p<.01. NR = No Reasoning Task; SR = Simple Reasoning Task; CR = Complex Reasoning Task; CON = Concentration; TMP = Time Pressure; ANX = Anxiety; FRU = Frustration; DIF = Difficulty; INT = Interest; ABL = Ability; MTV = Motivation.

The main findings were as follows: (1) Generally speaking, the participants rated the Complex Reasoning Task more stressful and poorly performed than the Simple Reasoning Task and, similarly, rated the Simple Reasoning Task more stressful and poorly performed than the No Reasoning Task; (2) Those tendencies were most clearly represented by the perceptions of Difficulty, where the Complex Reasoning Task was rated as more difficult than the No Reasoning and the Simple Reasoning Tasks; (3) The Complex Reasoning Task was also rated more interesting than the other two tasks; (4) The presence/absence of reasoning demands did not affect the participants' task motivation.

The number of significant correlations between the L2 production measures and the task difficulty measures

The results of the correlational analyses provided a relatively straightforward answer. Table 6 present the correlational table of the three task conditions. In Table 6, significant correlations are indicated by dark-colored cells. As can be seen, there is a remarkable increase in the number of significant correlations in the Complex Reasoning Task. Thus, the No Reasoning Task produced five significant negative correlations for Anxiety, Frustration, Difficulty, Interest, and Motivation; the Simple Reasoning Task produced five significant correlations, negative for Ability and Motivation, and positive for Concentration, Frustration, and Difficulty; and the Complex Reasoning Task produced 15 significant correlations. In addition, potentially interesting patterns were that in the No Reasoning Task, most of the significant correlations correlated with fluency aspects, in the Simple Reasoning Task, those were found mainly with the measure of lexical complexity, and in the Complex Reasoning Task, significant correlations were found in all aspects of L2 production except for the measure of syntactic complexity.

Discussion and implications

The present study was motivated by the following three research inquiries: (1) Does increasing intentional reasoning demands lead to increases in accuracy and complexity at the cost of fluency?; (2) What are the effects of intentional reasoning demands on the learner's perceptions of task difficulty?; (3) Do increasing task complexity lead to increases in the number of significant correlations between L2 task performance indices and the learner's perceptions of task difficulty? Based on the results of the present study, some tentative answers to each of those questions can be given.

Does increasing reasoning demands lead to increase in accuracy and complexity at the cost of fluency during L2 oral production?

Among the formulated research questions, this research question was the central concern of the present study. In order to answer the research question, seven planned-comparison hypotheses were formulated. As we have seen already, the results indicated that

Table 6

Pearson Correlations between the Eight L2 Production Measures and Eight Task Difficulty Measures as a Function of Task Complexity

Task Complexity	L2 production measure	Concentration	Time pressure	Anxiety	Frustration	Difficulty	Interest	Ability	Motivation
NR	Speech rate (un-pruned)	.069	-.213	-.325	-.357	-.377	.143	.258	-.032
	Speech rate (pruned)	.028	-.346	-.450*	-.481*	-.524**	.051	.29	.080
	Disfluency	-.064	.189	.284	.090	.110	-.216	.213	-.572**
	Hesitation length	.309	.319	.360	.255	.182	.329	-.107	-.020
	S-nodes per T-unit	.062	.403	.086	.057	.120	.166	-.062	.023
	Guiraud 2000	-.097	-.074	.216	-.009	-.104	-.524**	.166	-.185
	Error-free T-unit	-.394	.001	-.141	.019	-.152	-.152	.271	-.094
SR	Speech rate (un-pruned)	-.101	-.215	-.259	-.282	-.343	-.098	.205	-.063
	Speech rate (pruned)	-.153	-.371	-.178	-.242	-.372	-.047	.179	.144
	Disfluency	-.004	.271	-.043	.061	.071	-.331	.022	-.535**
	Hesitation length	-.152	.197	-.163	.065	.220	.12	-.067	-.037
	S-nodes per T-unit	.116	-.055	-.106	.107	-.194	-.171	.117	-.109
	Guiraud 2000	.477*	.386	.253	.511*	.579**	-.187	-.543**	-.026
	Error-free T-unit	.042	-.155	-.127	.005	.165	-.359	-.154	-.368
CR	Speech rate (un-pruned)	-.135	-.385	-.322	-.130	-.613**	.103	.424*	-.138
	Speech rate (pruned)	-.112	-.452*	-.373	-.213	-.648**	.098	.478*	.138
	Disfluency	.058	.095	.187	.169	.255	-.231	-.107	-.448*
	Hesitation length	.056	.582**	.032	.178	.311	-.146	-.178	-.496*
	S-nodes per T-unit	-.194	.374	-.309	-.123	.037	.234	.022	-.131
	Guiraud 2000	-.200	-.049	-.472*	-.540**	-.209	-.564**	.541*	-.237
	Error-free T-unit	.214	-.494*	-.478**	-.167	-.255	-.296	.423*	-.042

Note. NR = No Reasoning Task; SR = Simple Reasoning Task; CR = Complex Reasoning Task.

reasoning demands led to increases in accuracy and complexity at the cost of fluency in the sense of increased disfluency. Those results were compatible with the main prediction of the Cognition Hypothesis: thus, intentional reasoning demands lead to parallel attention to accuracy and complexity at the cost of fluency.

The two measures of L2 speech rate and hesitation length, however, did not show predicted negative effects of intentional reasoning. Several explanations of the unexpected results are possible, but perhaps the most likely account for the rejection of the hypotheses regarding L2 speech rate is related to the provision of planning time. With respect to hesitation length, the results of the present study could not find signs of the predicted negative effect of intentional reasoning demands. It may be that hesitation length is more sensitive to sustained attention. If this is the case, the measure may be more suitable for more extended task performance, where for instance time-series designs are possible. This construct validity issue should be further explored in future studies.

Despite the lack of the predicted effects of intentional reasoning demands on some of the fluency aspects of L2 production, overall the results were compatible with the prediction 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 can trigger reduced attention to form due to attentional capacity limitation. This theoretical position assumes a tension in language production between attention to meaning and form, where standing chances of the marriage of meaning and form are somewhat underestimated. There are of course other theoretical positions, however, one of which is the focus-on-form position (e.g., Doughty & Long, 2003; Doughty & Williams, 1998). Proponents of focus on form are careful enough not to posit necessary trade-offs between attention to meaning and attention to form. Rather communicative needs are seen as motivational sources for attention to form. This means that communicative meaningfulness is seen as cognitive potential for language development rather than a hindrance to L2 developmental processes such as form-meaning mappings. Simultaneously, focus-on-form proponents, notably Furthermore, besides potential developmental readiness, rule complexity, and individual differences in cognitive abilities, interference and confusion are likely to interrupt attention to L2 form (e.g., during dual-task L2 performance, see Robinson, 2003). As we have already seen, potential task complexity dimensions that are predicted to cause interference and confusion between information processing codes 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 hindered is of more importance than simply ascribing performance decrements to the theoretical construct of attentional capacity limitation. Doing so is simply duplicating terminology without explaining why attention to form is hindered, which is tautological, as described in the section 3.5.2.2. The results of the present study were to a greater extent in line with the focus-on-form position and the Cognition Hypothesis, supporting the view that attention through meaning or task-induced attention to form is attainable and is also a viable methodological principle of TBLT.

From a different perspective, attention to form through meaning also seems to be in line with the disfluency results. The notions of attention to fluency and attention to form are at least conceptually distinct, but it is not clear whether they are independent or not. Thus, we may also need to ask 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 hesitation as a sign of "quality control" for one's output does suggest an intimate 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. From this perspective, the results of the present study can be rephrased as follows: 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 less developed linguistic forms, such as those characterizing Givon's (1985) pragmatic mode for communication.

Robinson (2005, 2007a, 2007b) states that intentional reasoning prompts the learner to go beyond propositional information transmission, encouraging him/her to make inferences and produce utterances guided by questions about the mental states and intentions of others that lead them to behave in certain ways and not others. Answering such questions makes discourse-pragmatic demands on production (e.g., Givon, 1998) since ordering of information needs to be reasonable to the potential hearer and relevant to the communicative goal and so the text is encouraged to be organized more hierarchically (e.g., Berman & Slobin, 1994). In contrast, in the No Reasoning Task, 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 discoursal requirements; accordingly, demands on discourse coherence were somewhat reduced. The results of the present study showed that intentional reasoning led to greater syntactic complexity, indicating that discourse-functional complexity was accompanied by syntactic complexity, as is claimed by Givon (1979, 1998), Berman and Slobin (1994), and Robinson (2005), and so led to pushed output and enhanced attention to language form.

Does task complexity affect learner perceptions of task difficulty?

The results based on the Wilcoxon Test showed that the Complex Reasoning Task was rated as more difficult and interesting than the No Reasoning and the Simple Reasoning Task. With respect to Concentration, there was an overall significant result; however, the post-hoc tests did not detect significant differences. Some of the answers to the research question were thus positive, but clearly we need more substantial evidence in future studies. In fact, the overall trends were compatible with the prediction of the Cognition Hypothesis.

As Duran and Ramaut (2006) states, Robinson's distinction between task complexity and task difficulty is important, where the former can explain intra-individual variation and the latter inter-individual variation in successful task performance and language learning. What we have examined so far, however, was something in between; namely, the effect of task complexity on task perceptions within individuals. In short, the overall orders of perceived task difficulty seem to provide positive support for the anticipatory validity of a task designer's or teacher's decisions about material design and task sequencing based on degrees of cognitive task complexity.

Within this picture, and also given that task complexity is the sole prospective and stable hence reliable criteria of sequencing pedagogic tasks, what becomes important then is the teacher's and the learner's increasing awareness of the importance of better self-management and self-regulation (see Dörnyei, 2007) during task performance.

Clearly, ignoring task difficulty gives us a lopsided image of learner task performance (e.g., Corno et al., 2002).

Do reasoning demands increase the number of significant correlations between task difficulty and task performance?

The results of the number of significant correlations provided a positive answer to this research question. As we have seen, five significant correlations were found both in the No Reasoning Task and in the Simple Reasoning Task and 15 significant correlations in the Complex Reasoning Task. These results indicated that the role of learner perceptions was more important when reasoning demands were high. The fact that those correlational patterns were obtained from the same participants in a repeated measure design provide direct evidence for the greater role of individual differences in complex tasks. Similar results were also obtained in Robinson (2007b), using the measure of output anxiety.

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 are not the only one that could have been used. 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 adopting more specific measures may not result in stronger confirmation of the claim of the Cognition Hypothesis than that provided in the present study. This point 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 addressed by Nakamura (2007) from a typological perspective within the context of here-and-now versus there-and-then narrative production and Robinson (2007a, 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 idea 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, using reasoning tasks, Niwa (2000) showed that the learner's working memory capacity played an important role, correlating negatively and positively with some aspects of accuracy, fluency, and complexity, as task complexity was increased. Similarly, Robinson (2007b) also showed that the effect of task complexity was dependent on the level of output anxiety (MacIntyre & Gardner, 1994) in the sense that task complexity effects were present in less anxious learners but not in anxious learners. Investigations into the role played by the learner's individual differences will be of more crucial importance in future studies not only because the effect of task complexity can be dependent on them, but also because interactions between task complexity and learner characteristics can define teacher roles in the classroom in supporting and motivating language learners.

References

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

Bley-Vroman, R. (1983). The comparative fallacy in interlanguage studies: The case of systematicity. *Language Learning*, 33, 1-17.

Crono, L., Cronbach, L., Kupermitntz, H., Lohman, D., Mandinach, E., Porteus, A., & Talbert, J. (2002). *Remaking the concept of aptitude: Extending the legacy of Richard E. Snow*. Mahwah, NJ. Erlbaum.

Dörnyei, Z. (2007). *The psychology of the language learner: Individual differences in second language acquisition*. London: Mahwah, NJ.

Doughty, C., & Long, M. (2003). Optimal psycholinguistic environments for distance foreign language learning. *Language, Learning & Technology, 7*, 55-80.

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

Duran, G., & Ramaut, G. (2006). Tasks for absolute beginners and beyond: Developing and sequencing tasks at basic proficiency levels. In K. Van den Branden (Ed.), *Task-based language education: From theory to practice* (pp. 47-75). Cambridge: Cambridge University Press.

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

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.

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

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.

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

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.

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.

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

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

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. (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.

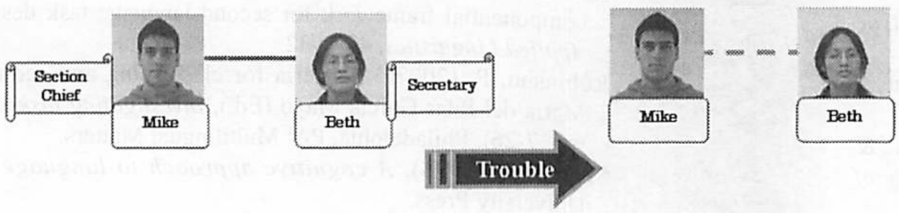
Skehan P. (1998). *A cognitive approach to language learning*. Oxford: Oxford University Press.

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.

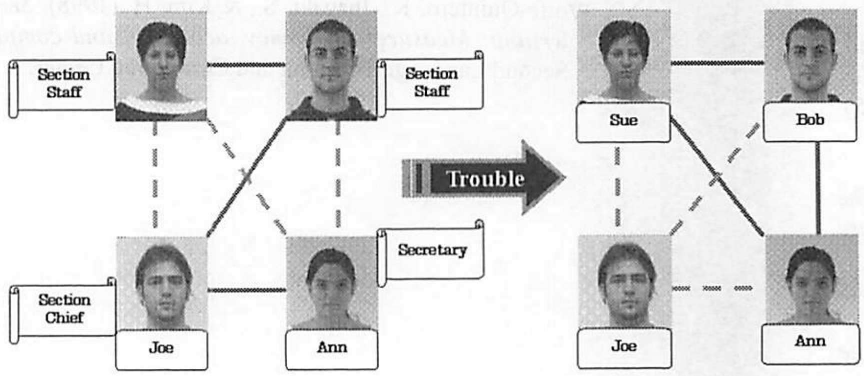
VanPatten, B. (1996). *Input processing and grammar instruction in second language acquisition*. Norwood, NJ: Ablex.

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.

Appendix A. Simple Reasoning Task



Appendix B. Complex Reasoning Task



Appendix C. No Reasoning Task