Task Complexity and Linguistic Complexity: An Exploratory Study

K. Philip Choong[1]
Teachers College, Columbia University

ABSTRACT

Central to any task-based syllabus is the notion of complexity. Proponents of task-based language teaching (TBLT) have argued that tasks be sequenced according to their inherent cognitive complexity, partially because learner performance changes according to the complexity of the task. This exploratory study examines the effect of task complexity on the linguistic complexity of task performance. The participants in the study were a group of ten advanced-level second language (L2) English speakers, and two groups of native speakers of English. Task complexity was operationalized by manipulating two independent variables – reasoning demand and contextual support – in a series of picture narration tasks. The study thus had a 2 x 2 factorial design, with participants completing the tasks under four different sets of conditions. Each set provided the participants with different reasoning demands and/or contextual support. Repeated measures ANOVA were used to analyze the initial data. Following these analyses, separate ANOVAs were calculated to distinguish between the types of reasoning that may have contributed to differences in task performance. It was found that contextual support had little influence on the complexity of task performance, but that reasoning demands, specifically causal and spatial reasoning, may have contributed to differences in the linguistic complexity of participants' task performance.

INTRODUCTION

Task-based language teaching (TBLT) has gained favor over the last two decades, both in second language pedagogy and in studies on second language acquisition. Task-based approaches are motivated by ideas espoused by communicative language teaching, which, *inter alia*, calls for language teaching to make use of real-life situations that necessitate language use. Under TBLT, learners perform tasks that focus on meaning exchange and use language for real-world, non-linguistic purposes.

Task-based approaches seek to create conditions for L2 acquisition that conform to what is theorized and examined in SLA research. It is believed that second language acquisition is not linear, and that there are many individual differences that distinguish learners from one another so that what is taught in the classroom may not be what is learned. Task-based language teaching, through rich exposure to the target language, theoretically allows for acquisition to take place

[1] K. Philip Choong is a doctoral student in TESOL at Teachers College, Columbia University. His research interests include task-based language teaching and learning and L2 curriculum design. Correspondence should be sent to: K. Philip Choong, 455 Albany Ct. West New York, NJ 07093. Email: kpc2001@tc.columbia.edu

according to the learners' internal syllabi. Additionally, task-based language instruction can provide an ideal platform for a focus on linguistic form (Long, 2000).

It has been hypothesized that the intentional manipulation of task variables in the context of meaningful language use will likely result in learners' focusing on form. According to Skehan (1998) and Robinson (2001a), tasks can be designed in such a way that learners allocate more attention to language form while still primarily focusing on task completion. This is done through what Skehan and Robinson refer to as the manipulation of *task complexity*, which can be matched both to the learner's linguistic development and to the purpose of the lesson. While both the elements of task complexity and the effects of manipulating task complexity are contested in the literature, the generally accepted definition provided by Robinson (2001b) stipulates that task complexity is "the result of the attentional memory, reasoning, and other information processing demands imposed by the structure of the task on the language learner" (p. 29). Consequently, Robinson proposes that tasks can be graded and sequenced according to their cognitive complexity. That given, whether the manipulation of task complexity would affect task performance still remains an empirical question.

This study examines the relationship between the cognitive complexity of tasks and the linguistic complexity of language learner output in the performance of those tasks. In what follows, the theoretical basis of the study will be presented. The design of the study will subsequently be described, followed by the results and discussion.

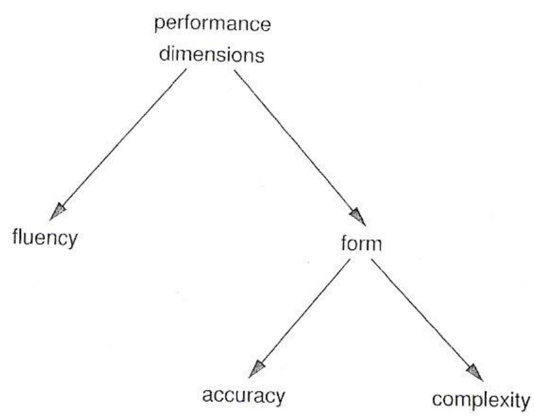
Task Complexity in TBLT Research

Notwithstanding the many definitions of *task* available in the literature, for the purposes of this paper, the definition provided by Ellis (2003) will be used. Ellis posits that the criterial features of a task are that (a) it is a workplan that includes a primary focus on meaning, (b) involves real-world processes of language use, (c) employs any of the four language skills, (d) engages cognitive processes, and (e) has a clearly defined communicative outcome.

While tasks may employ any of the four language skills, most of the research in tasks and task complexity has been conducted on output activities such as writing and speaking. Perhaps because of this, the goals of language learning have also been put in terms referring to speech. Skehan (1996, 1998) has divided these goals into three separate categories: fluency, accuracy, and complexity. Within this multicomponential view on L2 performance, *fluency* refers to a learner's ability to produce speech at a normal rate and without interruption; *accuracy* refers to a learner's ability to produce error-free language; *complexity* refers to the learner's use of more advanced language, with a wide range of syntactic structures. These constructs (cf. Housen & Kuiken, 2009) have been used not only to describe the goals of language acquisition, but also to measure progress in language learning. As such, much of the research conducted within the task-based framework has assessed the effects of task complexity in terms of these three language performance dimensions.

Theorizing about the different performance dimensions, Skehan (1996, 1998) proposes possible orientations that a learner may have in performing a task (see Figure 1). For instance, if the learner focuses primarily on task completion rather than language form, then the (s)he is oriented towards meaning, and thus would be more fluent. If, on the other hand, the learner focuses more on language and language development rather than on task completion, then the learner is oriented towards form, and his or her language would be more complex and/or accurate. These performance dimensions and the tensions between them are believed to mirror differential demands upon the attentional resources of language learners. In other words, shifting the orientation of attentional resources is believed to impact task performance along performance dimensions. It has been hypothesized, therefore, that task properties, such as task complexity, could be manipulated to shift learners' attentional resources. In the next section, the notion of task complexity will be examined in detail, and a review of research on the proposed dimensions of task complexity will follow.

FIGURE 1
Task Performance Dimensions



Skehan and Foster (2001)

Task Complexity and Language Performance

Properties of tasks are said to have an effect on learners' L2 performance. One of these inherent task properties is complexity. The use of the term *complexity*, however, is fraught with ambiguity in TBLT research, as it can refer both to properties of language performance, depicted as one of the performance dimensions in Figure 1, as well as to properties of tasks themselves. In what follows, the effects of the *complexity of tasks* (understood as a *task property*) on the *complexity of performance* (understood as a *property of language proficiency*), will be examined from a theoretical point of view.

The Trade-Off Hypothesis

It has been assumed in the TBLT literature that tasks can be designed to direct learners' attentional resources, influencing their output. According to Skehan's (1996, 1998) *single resource model of attention*, tasks need to be constructed in such a way that learners' attentional resources are not exhausted by the performance of these tasks, and that there are resources remaining for learners to focus on language form. In specific, Skehan works from a dyadic framework, following the work of VanPatten (1990), who makes the argument that learners have limited attentional resources such that they cannot pay attention to language forms without some sacrifice of attention to language meaning. In Skehan's view, the more difficult a task, the less likely a learner will have attentional resources to devote to form over meaning, making language acquisition more difficult. As a result, tasks that are perceived as difficult by the learners, under Skehan's framework, generally result in less fluency, accuracy, and complexity in L2 performance. However, Skehan does hold out the possibility that perhaps "tasks with 'difficult' content will prompt learners into trying to express themselves through complex language forms because it makes them realize that simple language will not do" (Skehan & Foster, 2001, p. 197).

In order to assess task difficulty, Skehan (1998) analyzes tasks in terms of "the language required, the thinking required, and the performance conditions for a task" (p. 99). He designates these, respectively, as code complexity, cognitive complexity, and communicative stress, as shown in Figure 2. *Code complexity* refers to the linguistic demands of the task. Regarding *cognitive complexity* of tasks, Skehan distinguishes between two aspects of cognition: the capacity to access familiar solutions to tasks (referred to as *cognitive familiarity*) and the need to work out solutions to novel problems (referred to as *cognitive processing*). Finally, *communicative stress* reflects the performance conditions under which tasks are performed. Examples include the urgency with which a task must be completed, or the perception on the part of the student of how much pressure there is to complete a task under difficult conditions (e.g., time limits, speed of presentation, and opportunities to control interaction).

The factors described in Table 1 have been proposed to have a systematic effect on second language performance. In general, Skehan argues that more cognitively complex tasks will force second language users to divert most of their attentional resources to meaning, and less to form. This is referred to as the *Trade-Off Hypothesis*. As a corollary, decreasing task complexity will allow learners to increase their attention to form, as fewer resources will be required to process for meaning.

TABLE 1
Task Analysis Scheme

1 *Code complexity*

- linguistic complexity and variety
- vocabulary load and variety
- redundancy and density

2 *Cognitive complexity*

Cognitive familiarity
- familiarity of topic and its predictability
- familiarity of discourse genre
- familiarity of task

Cognitive processing
- information organization
- amount of 'computation'
- clarity and sufficiency of information given
- information type

3 *Communicative stress*
- time limits and time pressure
- speed of presentation
- number of participants
- length of texts used
- type of response
- opportunities to control interaction.

Skehan (1998)

The Cognition Hypothesis

The *Cognition Hypothesis*, proposed by Robinson (2001b, 2005, 2007a), posits that pedagogic tasks should be sequenced incrementally for learners based on the cognitive complexity of the tasks. Robinson's hypothesis also assumes that as tasks increase in the conceptual/communicative demands placed on learners, learner attention to aspects of the L2 system that attempt to meet those demands may also be increased. In contrast to Skehan (1998), Robinson adopts a *multiple resource view of attention*, where attentional resources employed in the completion of a task are drawn from multiple pools. This view holds that tasks are made more difficult only if there is interference, or cross-talk, within distinct resource pools. Robinson hypothesizes that attention can be directed towards a specific resource pool, or dispersed among many. Task complexity may be used as a modulator of attention between and amongst those pools.

With regard to task complexity, Robinson (2001b, 2005, 2007a), questions the way Skehan (1998) and others have described it. Previously (e.g., Skehan & Foster, 2001), in relation to the inherent task properties the terms *difficulty* and *complexity* were used interchangeably, encompassing a wide array of factors, including, but not limited to, cognitive, affective, linguistic, interactional, and experiential influences (Robinson, 2001b). Taking cues from research in psychology, Robinson instead proposes a triadic componential framework that organizes the factors influencing task performance into three groups: (a) *cognitive factors*, which comprise *task complexity* (b) *interactional factors*, which make up *task conditions*, and (c) *learner factors*, which constitute *task difficulty* (summarized in Table 2).

TABLE 2
The Triadic Componential Framework for Task Classification

Task Complexity (Cognitive factors)	*Task Condition* (Interactive factors)	*Task Difficulty* (Learner factors)
(Classification criteria: cognitive demands) (Classification procedure: information-theoretic analyses)	(Classification criteria: interactional demands) (Classification procedure: behavior-descriptive analyses)	(Classification criteria: ability requirements) (Classification procedure: ability assessment analyses)
(a) Resource-directing variables making cognitive/conceptual demands	*(a) Participation variables* making interactional demands	*(a) Ability variables* and task-relevant resource differentials
+/– here and now	+/– open solution	h/l working memory
+/– few elements	+/– one-way flow	h/l reasoning
–/+ spatial reasoning	+/– convergent solution	h/l task-switching
–/+ causal reasoning	+/– few participants	h/l aptitude
–/+ intentional reasoning	+/– few contributions needed	h/l field independence
–/+ perspective-taking	+/– negotiation not needed	h/l mind/intention-reading
(b) Resource-dispersing variables making performative/procedural demands	*(b) Participant variables* making interactant demands	*(b) Affective variables* and task-relevant state-trait differentials
+/– planning time	+/– same proficiency	h/l opennes to experience
+/– single task	+/– same gender	h/l control of emotion
+/– task structure	+/– familiar	h/l task motivation
+/– few steps	+/– shared content knowledge	h/l processing anxiety
+/– independency of steps	+/– equal status and role	h/l willingness to communicate
+/– prior knowledge	+/– shared cultural knowledge	h/l self-efficacy

Robinson and Gilabert (2007)

According to Robinson (2001b, 2005, 2007a), task complexity is the sole basis of design decisions for task sequencing within the TBLT framework. It must be noted that, in contrast to the Trade-Off Hypothesis (Skehan, 1996, 1998), the Cognition Hypothesis differentiates between task complexity and task difficulty. *Task complexity* accounts for *intra-learner variability*, or the variability which would be evident if the same learner were to perform different tasks. In other words, task complexity is external to the learner. Those factors which are internal to the learner, such as aptitude, anxiety, and intelligence, relate to *task difficulty*. Therefore, task difficulty accounts for *inter-learner variability*. Because decisions cannot usually be made according to task difficulty or, according to Robinson, task conditions, it has become *de rigueur* for researchers working with the Cognition Hypothesis to exclude learner internal factors from discussions about the *a priori* sequencing of tasks in a language syllabus (Robinson, 2001b). Consequently, research has examined the impact of manipulating cognitive factors of task complexity identified by Robinson as resource-dispersing or resource-directing dimensions of task complexity on L2 development.

Task complexity has been characterized by Robinson as "the intrinsic cognitive demands of a task which contribute to between task variation in spoken and other kinds of performance for any one learner performing a simple and a more complex version [of a task]" (Robinson, Cardierno, & Shirai, 2009, p. 535). Task complexity is conceptualized as the result of information processing demands imposed by the structure of the task on the language learner. For example, Robinson *et al*. claim that a task requiring learners to exchange business cards in an L2 is less complex than conducting a business meeting in an L2, predicting that performative success is more likely on the less complex task than the more complex task. Following the multiple resource view of attention, Robinson distinguishes between two task complexity variables, categorized as resource-directing and resource-dispersing (see Table 3). *Resource-directing variables* account for, *inter alia*, whether there are few or many elements to be compared, whether contextual support is available to the learner, and whether there are reasoning demands imposed on the learner. *Resource-dispersing variables* include the presence or absence of planning time, whether the task is single or multiple in nature, and whether the learner has prior knowledge that could aid in the completion of the task.

By increasing task complexity along resource-dispersing dimensions, (e.g., decreasing planning time or assigning a novel task), attentional resources are predicted to be depleted and dispersed among many non-specific language areas. This may in turn lead to tradeoffs in terms of the complexity, accuracy, and fluency of the language used to complete the task. However, increasing task complexity along resource-directing dimensions is expected to focus the attention of learners on specific, language-relevant features of the task and thus improve the accuracy and complexity, though not fluency, of the learner's linguistic code. These hypotheses form the crux of what Robinson calls the *Cognition Hypothesis* (2001a, 2001b, 2005, 2007a).

TABLE 3
Resource-directing and Resource-dispersing Dimensions of Task Complexity

Task Complexity (Cognitive Factors)	
resource-directing	***resource-dispersing***
e.g.,	e.g.,
+/- few elements	+/- planning time
+/- contextual support (here-and-now)	+/- single task
+/- reasoning demands	+/- prior knowledge
+/- intentional reasoning	
+/- causal reasoning	
+/- spatial reasoning	

Robinson (2001b, 2005, 2007a)

Cognitive Factors in Task Complexity

Empirical studies examining the effects of task complexity provide some support for the impact of cognitive factors in task complexity on L2 language performance. Measures of complexity, accuracy, and fluency have been used to capture these effects. Less research, however, has been conducted on the resource-directing dimensions of task complexity than on the resource dispersing-dimensions. Recently, two of Robinson's resource-directing dimension, *the here-and-now versus the there-and-then dimension* (sometimes referred to as *+/- contextual support*) and *the +/- reasoning demands dimension*, have drawn the attention of researchers testing the Cognition Hypothesis. According to Robinson (2001b, 2001b, 2005, 2007a), increases in these two dimensions should improve the complexity of learner output. In contrast, Skehan (1996, 1998) predicts that these same increases will decrease the complexity of learner output. A review of research on these two dimensions of task complexity follows.

Contextual Support and Language Performance

Many research and pedagogic tasks used in TBLT research involve the transfer of information (Ellis, 2003). These tasks often involve the use of non-verbal input, such as a picture, movie, or cartoon. The learner or participant performing these tasks must then translate these images into a verbal or written narrative or description. Such tasks can be designed in such a way that the language learners have contextual support (+ here-and-now, or – there-and-then) in which they are able to see the pictures or movie while narrating them. The tasks can also be designed so that there is no contextual support (- here-and-now, or + there-and-then), in which the learners are unable to see the visual input as they narrate.

It has been asserted that in first language acquisition, the ability to refer to events displaced in time and place emerges later than the ability to refer to events occurring in the here-and-now frame (Brown & Bellugi, 1964). Bardovi-Harlig (1994) makes the same claim for second language acquisition. Givon (1985) asserts that in cognitively undemanding contexts, in which contextual support is available to reduce the demands on language to convey the message, language learners resort to a pragmatic mtode of communication. *The pragmatic mode of communication* is usually characterized by a topic-comment structure, little use of grammatical morphology, and simple syntax. In contrast, in cognitively demanding contexts where contextual support is not available, speakers use a syntactic mode of communication. The *syntactic mode of communication* is characterized by subject-predicate structure, greater use of grammatical morphology, and complex syntax and subordination. Following these arguments, Robinson (1995, 2001ba, 2001b, 2005, 2007a) places this dimension of task complexity in the resource-directing category of task variables, and predicts that "where context support is not available … the language user has to ensure that all the necessary presuppositions are coded within the message … [and] needs to develop or exploit a greater range of syntactic resources" (Robinson, 1995, p. 104).

Several studies have investigated the effects of here-and-now/there-and-then dimension on the syntactic complexity of language learners. For the purpose of this paper, only those that investigate oral production will be reviewed. Robinson (1995) used picture stories to elicit narratives from L2 participants of diverse language backgrounds. The participants were asked to narrate while viewing the pictures (here-and-now condition) as well as from memory, after the pictures were removed from sight (there-and-then condition). The results revealed that while the there-and-then condition improved accuracy and lexical complexity, it had no statistically significant effect on syntactic complexity, as measured by the *index of subordination*, defined as a number of s-nodes (i.e., verb phrases) per t-unit (i.e., a main clause plus any subordinating clauses). Rahimpour (1997, as cited in Robinson, 2005), in a partial replication study, found similar results.

Skehan and Foster (1999) examined contextual support by using video instead of picture series. Forty-seven lower-intermediate-level participants were asked to describe the story told by the video. Skehan and Foster operationalized contextual support as a watch-and-tell condition, where the participants narrated the story as they watched the video. Under the there-and-then condition, the participants narrated the story after they watched the video. Syntactic complexity was again indexed by a measure of subordination, in this case, the number of clauses per c-unit (i.e., conversation unit). Skehan and Foster found that the there-and-then condition led to more syntactic complexity, but improved neither accuracy nor fluency.

In a larger study, Iwashita, Elder, and McNamara (2001) also examined contextual support, but within a testing context. One hundred and ninety-three pre-university students were asked to narrate a story based on a set of picture prompts. The prompts were available to the participants under the +contextual support condition but were unavailable to them under the –contextual support condition (the authors referred to these conditions as +/- immediacy). Iwashita *et al.* found that the there-and-then condition produced more accurate language, but there was no effect on fluency or syntactic complexity as measured by the number of clauses per c-unit.

Another study that examined the effects of contextual support was conducted by Gilabert (2007). Forty-eight participants were asked to narrate picture stories under four different conditions, as planning time was also being examined in the study. Like the previous studies reviewed above, contextual support was operationalized by whether or not the pictures were visible to the participants. Gilabert found that there was an effect on contextual support for the accuracy of speech, but not on syntactic complexity, as measured by s-nodes per t-unit.

Overall, it remains unclear what effects contextual support has exactly on syntactic complexity. In general, it seems that language users lose fluency, but gain accuracy under there-and-then conditions ([-contextual support]). It should be noted that no study has found a simultaneous gain for accuracy and complexity.

Reasoning Demands and Task Performance

As noted previously, Robinson (2001a, 2001b, 2005, 2007a) states that reasoning demands are a part of the resource-directing dimension of task complexity. Three aspects of reasoning have been identified by Robinson, taking cues from work conducted in first language acquisition studies and psychological research: spatial, intentional, and causal reasoning. He cites *spatial reasoning* as a type of reasoning demand that could be a resource-directing component of task complexity. The support that spatial reasoning leads to greater syntactic complexity comes from both first and second language acquisition research. It has been shown that L1 learners first use ground level route maps for spatial location, in which a given landmark is connected only by other landmarks which are visible from it. As these language users develop, they use survey maps to navigate, allowing for the use of multiple landmarks and also for the use of axis-based relations of betweeness and back/frontness (e.g., Becker & Carroll, 1997; Cornell, Heth, & Alberts, 1994).The claim about intentional reasoning is also echoed in first language acquisition research. For example, Lee and Rescola (2002) found that cognitive state terms (e.g., *think, know*) emerged later in children than physiological (e.g., *sleepy*), emotional (e.g., *happy*), and desire terms (e.g., *want*). In other words, intentional reasoning entails "reasoning about, and successfully understanding (intention-reading) the motives, beliefs and thoughts which cause people to perform actions" (Robinson, 2007b, p. 194). Finally, causal reasoning entails reasoning about the cause and effect of events.

In predicting the impact of reasoning demands on task complexity, Robinson follows Brown, Anderson, Shillock, and Yule (1984) who state that tasks which, for example, require the transmission of information and reasoning to establish causality and justification of beliefs, are more complex than tasks that do not have these reasoning demands. More specifically, Robinson (2005) claims that:

> tasks which require no causal reasoning to establish event relations, and simple transmission of facts, compared to tasks which require the speaker to justify beliefs, and support interpretations of why events follow each other by giving reasons, also require, in the latter case, expressions, such as logical subordinators (*so, because, therefore*, etc.). In the case about reasoning about other people's intentions and beliefs, use of psychological, cognitive state verbs (e.g., *know, believe, suppose, think*) is required. Both of these introduce complex syntactic complementation. (p. 5)

Accordingly, Robinson (2005) contends that "increasingly complex demands that tasks impose along these dimensions can be met by the use of specific aspects of the L2 which code these familiar adult concepts" (p. 6). In other words, increasing the complexity of the tasks themselves provides what Loschky and Bley-Vroman (1993) refer to as *task-natural conditions for language use*.

To date, few studies have examined the effects of reasoning demands on learner production. Robinson (2000, as cited in Robinson, 2005) operationalized reasoning demands through a series of one-way, interactive, and closed picture arrangement tasks. In this particular study, one participant was asked to view a randomly ordered set of pictures and determine the correct order. Then, he/she would narrate the story to another participant, who could ask questions, so that the second participant could arrange his/her own set of randomly ordered pictures in the correct order. To differentiate the reasoning demands, Robinson used sets of pictures from the Wechsler Intelligence Scale, ranging from the least to most complex sequences. The least complex set of pictures depicted a simple sequence of a house being built, which did not require any reasoning about motives or intentions of other people. The most complex set of pictures required that the participants infer the motives and intentions of the characters in the story in order for the pictures to be sequenced. Linguistic input, in the form of phrases that would be helpful but not essential for completing the tasks, were provided along with the pictures. Robinson found that participants were more likely to incorporate and attend to linguistic input during more complex tasks. Although this study did not examine the syntactic complexity of the participants' production directly, it did show that the more complex tasks were, the more likely participants were to "incorporate and practice aspects of the L2 over which they may have low control, or grammatical and lexical forms they may not have yet acquired" (Robinson, 2005, p.18). Therefore, the results would indicate that given relevant input, learners could increase the syntactic complexity of their performance by completing more complex tasks.

Niwa (2000, as cited in Robinson, 2005), used the same instrument as Robinson (2000, as cited in Robinson, 2005). The participants in this study were also asked to order and narrate the picture stories, but were not required to interact with another participant. Niwa examined the effects of individual differences in intelligence, aptitude, and working memory capacity, as well as task complexity, on task performance. This study provided evidence that increasing task complexity through reasoning demands also increases the role that individual differences play in differentiating task performance. For example, those with higher working memory capacity and aptitude were less fluent on the more complex tasks, suggesting that they were making greater efforts to meet the reasoning and accompanying linguistic demands of the tasks compared with those with lower aptitudes and working memory capacities, who were less able to meet those reasoning demands (Robinson, 2005). However, there was no direct evidence that syntactic complexity increased as tasks became more complex. The findings on accuracy were inconclusive.

In a partial replication of Robinson (2000, as cited in Robinson, 2005), Robinson (2007b) conducted a study examining the effects of intentional reasoning on task performance. Identical procedures and materials were used, that is, pairs performing picture arrangement and narration. Unlike the previous study, Robinson (2007b) examined the syntactic complexity of L2 production. Both general (i.e., the number of clauses per c-unit) and specific measures of syntactic complexity (i.e., number of psychological state terms, infinitival phrases, and conjoined and wh- clauses) were used. The results showed no impact of manipulation in the domain of intentional reasoning on the general measures of syntactic complexity. However, on specific measures, the results were mixed in terms of linking intentional reasoning demands with syntactic complexity. Robinson pinpointed two factors that could explain the results of the study. The first is that the tasks may have been complex along the resource-dispersing dimensions (i.e., insufficient planning time may have been given). The second explanation is that dyadic tasks, because of interlocutor interaction, could affect the complexity of output in response to task demands. It appears, therefore, that just as with contextual support, there is mixed evidence regarding the effects of reasoning demands on syntactic complexity.

As mentioned before, there is a need for more research that examines the effects of resource-directing cognitive factors in task complexity on L2 language performance. Since these factors are the major source of contention between the Trade-off Hypothesis and the Cognition Hypothesis, they warrant further scrutiny. It was against this theoretical background that Robinson's Cognition Hypothesis framework was used for the current study. In the handful of previous studies available, monologic tasks were used in the context of researching resource-directing cognitive factors (e.g., Gilabert, 2007). In contrast to dyadic tasks (e.g., Robinson 2005, 2007b), monologic tasks may facilitate the use of more complex syntax, as it is believed that interlocutor interaction could potentially "mitigate attempts to produce complex syntax in response to conceptual and functional demands of the task" (Robinson, 2005, p.10). Therefore, the present study set out to investigate the effects of task complexity on L2 performance in monologic tasks. This study also sought to broaden the empirical base related to the effects of reasoning demands on L2 performance in relation to contextual support.

Research Questions

Motivated by previous research, this study examined the effects of task complexity manipulated along the [+/- contextual support] and [+/- reasoning demands] components on L2 performance. The research questions investigated were:

1. Is there a relationship between task complexity and the complexity of oral L performance?
2. If so, what is the nature of that relationship?

METHODOLOGY

Design

The present study incorporated a repeated measures design, with the study participants performing picture narration tasks under four different conditions. The conditions were defined as [+/- contextual support] and [+/- reasoning demands]. Contextual support was operationalized as whether or not the participants could see the picture sets as they narrated. Those conditions which were +contextual support allowed the participants to look at the pictures as they narrated the story, while in the –contextual support conditions, the participants were not able to see the pictures as they narrated. As stated in the review of literature, this has often been referred to as a here-and-now versus a three-and-then condition. Reasoning demands were operationalized as whether or not the picture series were in sequence. In the –reasoning demands conditions, the pictures were sequenced when presented to the participants. Those conditions that were +reasoning demands required the participants to sequence the pictures prior to narrating the story. Because the two variables of reasoning demands and contextual support were counterbalanced in the study, one could compare the effects of either variable separately and also in combination. It must also be noted that each condition also included its own picture set that constituted the input in this design. Table 4 below summarizes the study design.

TABLE 4
Study Design

Picture set	Condition	Contextual Support	Reasoning Demands
1	1	+	-
2	2	-	-
3	3	+	+
4	4	-	+

In this design, the independent variables were contextual support and reasoning demands. The dependent variable was the linguistic complexity of the participants' output, understood as content complexity and syntactic complexity. In keeping with the repeated measures design, only intra-participant variance, and not inter-participant variance was examined. This is in line with the Cognition Hypothesis, as it predicts differences within learners, not across learners.

The research questions were addressed by comparing the output produced by the participants under conditions/picture sets 1 and 3 with the output elicited by conditions/picture sets 2 and 4 (+/-contextual support), and by comparing the output elicited under the condition/picture sets of 1 and 2 with the output elicited by the conditions/picture sets 3 and 4 (+/-reasoning demands). As mentioned before, in this design, each picture set represented one of the four experimental conditions.

Participants

The participants in this study included an experimental group and two comparison groups. The experimental group comprised ten native speakers of Japanese who were studying ESL in New York City. They were students in advanced ESL classes, having been placed there by their performance on an institutional placement test. The classes met four times a week, for two and a half hours each class. The average age of the experimental group was 38, and there were nine females and one male. Their length of residence in the United States varied from two months to several years. All of them had at least five years of EFL instruction, both in secondary school and university settings.

The first comparison group was made up of five native speakers of English, all females, with an average age of 33. The second comparison group consisted of ten adult native speakers of English of mixed gender, with an average age of 40. None of the participants in this study were told the purpose of the study beforehand.

Instruments

With the main focus of the investigation being on the effects of task complexity on learner spoken production, the study used a series of picture narration tasks. Picture narration has been used in previous studies investigating the contextual support and reasoning demands conditions (e.g., Gilabert, 2007; Robinson, 2001a, 2001b, 2005, 2007b). In the present study, four sets of pictures were used, each telling a unique story. Each sequence of ten pictures was taken from the Mr. Bean television program. While all the participants knew the character of Mr. Bean, none were familiar with the particular stories portrayed in the picture sets used in this study.

The first picture set illustrated a sequence in which Mr. Bean was taking a test. In the act of cheating on the test, he caused interactions between himself and his neighbor. The second picture set was situated at a swimming pool, where Mr. Bean had a confrontation with other divers. The third picture set depicted Mr. Bean waiting in line at a hospital, where he employed ruses to jump to the front of the line. The fourth picture set was situated in a waiting room, where Mr. Bean again employed some scheme in order to read a comic book belonging to another person. Each of these picture sets depicted confrontations between Mr. Bean and others. Therefore, each of the picture sets required the use of reasoning demands in order to narrate them. The picture sets can be found in Appendix A.

Procedures

As stated above, the present study incorporated a repeated measures design, which required each member of the experimental group and first comparison group to narrate the four picture sets under their corresponding conditions (see Table 4). The participants in these groups were always given the first two pictures in each set as well as the last two pictures. Under the +reasoning demands conditions, the intervening six pictures were scrambled so that the participants had to sequence them before narrating, whereas under the –reasoning demands conditions, the pictures were presented to the participants in the correct sequence. To account for the second experimental variable, the pictures were made available to the participants who were narrating under the +contextual support conditions, while the pictures were unavailable to the participants who were narrating under the –contextual support conditions.

The participants met with the researcher individually. The average time taken to perform the tasks was approximately 40 minutes. To ensure that the participants understood the task, they were given a simple set of practice pictures to narrate. These practice sets showed simple, everyday actions, like checking out at a grocery store, frying an egg, or doing school homework. The practice sets which preceded each experimental narration were performed under the same conditions as the experimental narration (e.g., +contextual support / + reasoning demands). Instructions for each experimental condition were given in English as follows:

> *I am going to show you a set of pictures that tells11 a story. Please take as long as you like to look over the pictures, then tell me the story as if I cannot see the pictures. We will do this twice. The first set will be practice, just to make sure you understand the instructions. There is only one "correct" story for these pictures.*

Following the above statement, a brief description of the condition under which the picture sets were to be narrated ensued, so the participants knew whether they would have contextual support while narrating and whether they would need to sequence the pictures before narrating. Following these instructions, there was no interaction between the researcher and participants, making these tasks monologic.

The second comparison group examined each of the four picture sets and ranked them according to what they perceived as the most "difficult" to narrate or comprehend. They were not required to narrate the stories, though some did.

Data Analysis

The narrations produced by the experimental group and the first comparison group were captured in their entirety using a digital recorder . The recordings were then sent to a professional transcription service to be transcribed. Afterwards, the transcripts were checked by the researcher and modified when needed.

After transcripts were created, they were examined and coded for linguistic complexity, understood as content and syntactic complexity. For a measure of content complexity, understood as the meaningfulness and elaboration of propositional information within the parameters of the task, idea units were used. *Idea units* are said to "contain all the information a speaker can handle in a single focus of consciousness" (Olson, Torrance, & Hildyard, 1985, p. 106). It would stand to reason then that the more complex the content of the story to be narrated, the more idea units it would contain. For example, telling a simple story about going to the grocery store and buying a carton of milk may result in the use of fewer idea units than telling a complex story about the events of a movie.

Carrell (1985) provided an operational definition of an idea unit: "each idea unit consisted of a single clause (main or subordinate, including adverbial and relative clauses). Each infinitival construction, gerundive, nominalized verb phrase, or conjunct was also identified as a separate idea unit. In addition, optional and/or heavy prepositional phrases were also designated as separate idea units" (p. 737). An example of a sentence divided into idea units is also provided by Carrell, "serious nuclear accidents…have lead/ in the past/ to the implementation of strict safety rules/ regarding the construction/ and operation of nuclear power stations" (p. 752). In the present study, Carrell's operationalization of idea units was used, and a frequency count of the number of idea units in each narration was used as the measure of content complexity.

For a measure of syntactic complexity, words per t-unit were used, following Bygate (2001). A *t-unit* is defined as a finite clause together with any subordinate clauses dependent on it. This measure was chosen because it "reflects an ability to combine a number of lexical items around syntactic structures … it involves the extent to which lexical processing can be managed according to basic syntactic parameters" (Bygate, 2001, p. 34). *Words per t-unit* were calculated by the formula: *number of words in each narration divided by the number of t-units in each narration.* Table 5 summarizes how linguistic complexity was defined and measured in this study.

TABLE 5
Linguistic Complexity in this study

Operationalization	Content Complexity	Syntactic Complexity
Measure	No. of i-units	No. of words / t-unit

After the transcripts were coded by the researcher, entailing both the identification and frequency counts of idea units and t-units and a count of the number of words used per narration, half the transcripts were then coded by another applied linguist for the same measures. Overall, the inter-coder reliability was high, at .87. Any disagreements or discrepancies were discussed until a uniform coding was agreed upon. The variables were subsequently tested for normality using measures of skewness and kurtosis. The data were then entered into a statistical program, and repeated measures ANOVA was employed to check for statistical significance. To check if the assumption of sphericity was violated, Mauchly's tests of sphericity were performed for each analysis.

Results

Descriptive statistics for the measures of content and syntactic complexity are presented in Table 6. As demonstrated, both content and syntactic complexity varied in similar ways among the participants in the experimental group. In both types of complexity, condition 3 (+contextual support, +reasoning) produced the most complex utterances, followed by condition 4, condition 1, and condition 2 (3>4>2>1 for syntactic complexity vs. 3>4>1>2 for content complexity). On the surface, it seems that reasoning demands played a larger role in determining the complexity of the output than contextual support.

TABLE 6
Descriptive Statistics for Experimental Group (n=10)

Condition/Picture Set	*Linguistic Complexity*	
	Syntactic Complexity Mean/SD	Content Complexity Mean/SD
1: +contextual support, - reasoning demands	7.92/1.20	13.00/5.58
2: - contextual support, - reasoning demands	8.14/ .80	12.20/3.49
3: +contextual support, +reasoning demands	8.50/1.10	20.00/3.62
4: - contextual support, +reasoning demands	8.46/1.50	16.70/5.20

After checking that all the assumptions were met, a repeated measures ANOVA was performed on the data to determine whether there were intra-participant differences in linguistic complexity under the four conditions. Table 7 shows the results. Contextual support did not have a statistically significant effect on the syntactic or content complexity of the participants' narrations. Reasoning demands did not have a statistically significant effect on syntactic complexity, but did have a significant effect on content complexity. There were no statistically significant differences between the two +reasoning conditions, further suggesting that contextual support played a minor role in task performance, if any. Ostensibly, it seems that reasoning demands may have prompted the participants in the experimental group to speak more about the picture strips.

TABLE 7
Repeated Measures Analysis for Experimental Group (n=10)

Condition Comparison	*Linguistic Complexity*	
	Syntactic Complexity F (p-value)	Content Complexity F (p-value)
+contextual support vs. –contextual support	.039 (p=.848)	2.927 (p=.121)
+reasoning demands vs. –reasoning demands	1.271 (p=.289)	23.500 (p<.001)*
+reasoning +support vs. +reasoning –support	.004 (p=.954)	2.440 (p=.157)

To further establish the effects of contextual support and reasoning demands on linguistic complexity, the results for the experimental group were next compared to the results found for the first comparison group. Overall, the descriptive statistics for the first comparison group mirror what was found in the experimental group (see Table 8).

TABLE 8
Descriptive Statistics for First Comparison Group (n=5)

Picture Set/Condition	*Linguistic Complexity*	
	Syntactic Complexity Mean/SD	Content Complexity Mean/SD
1: +contextual support, - reasoning demands	9.66/1.82	13.40/3.29
2: - contextual support, - reasoning demands	8.89/ .99	11.40/2.07
3: +contextual support, +reasoning demands	10.40/ .94	20.80/2.86
4: - contextual support, +reasoning demands	9.80/1.35	19.40/4.51

Again, conditions three and four elicited more complex syntax and content than conditions one and two (3>4>2>1 for both syntactic and content complexity). The only exception to this pattern was for the syntactic complexity for the first comparison group. For this measure, conditions three and four, respectively, elicited the most complex output, but condition one elicited slightly more syntactically complex output than condition two. The repeated measures ANOVA showed similar results to the experimental group, with a significant effect for the reasoning demands conditions as measured by content complexity. These data are presented in Table 9.

TABLE 9
Repeated Measures Analysis for First Comparison Group (n=5)

Condition Comparison	*Linguistic Complexity*	
	Syntactic Complexity F (p-value)	Content Complexity F (p-value)
+contextual support vs. –contextual support	.953 (p=.384)	4.042 (p=.115)
+reasoning demands vs. –reasoning demands	1.340 (p=.312)	20.340 (p<.05)*
+reasoning +support vs. +reasoning –support	.390 (p=.566)	2.580 (p=184)

So far the results are similar for both the experimental and the first comparison group. The task complexity dimension of contextual support did not seem to have an effect on either the syntactic or content complexity of the output. The task complexity dimension of reasoning demands seemed to have an effect for the content complexity of the output, but not on the syntactic complexity of the participants' output.

It must be noted that in line with previous research on the resource-directing dimensions of task complexity, reasoning demands in this study were operationalized as a unitary construct. However, Robinson's (2007a, 2007b) recent conceptualization of reasoning within the Cognition Hypothesis offers a finer-grained understanding of reasoning demands, as he contends that there are many distinct components of cognitive reasoning. This argument may shed light on the results presented thus far. Therefore, the data in the present study are further examined thorough the lens of the finer-grained construct of reasoning.

As mentioned in the literature review, Robinson (2007a), taking cues from the field of cognitive psychology, identified three components of reasoning demands:intentional reasoning, spatial reasoning, and causal reasoning. Therefore, the first step towards explaining the role of reasoning demands in the present study was to catalog evidence that those demands were employed during the completion of the task. First, for evidence of intentional reasoning, a coding system was adopted from Lee and Rescola (2002). In this system, certain lexical items serve as indicators of intentional reasoning. These words include psychological state words (*happy, sad, angry*), desire words (*want, need, like*), and cognitive terms (*think*, *know*, *pretend*). Following this system, a frequency count was made of these terms and compared across all four conditions. The descriptive statistics for the frequency counts are given in Table 10 but, as stated previously, the Cognition Hypothesis only makes predictions about intra-learner variation, and descriptive statistics necessarily discount individual differences. The results are presented for the effects of picture set/conditions on L2 production. As mentioned in the design section, each picture set represented one of the four experimental conditions; for example, picture set 1 was performed under the +contextual support, –reasoning demands condition. Mean values displayed in Table 8 suggest that picture sets 3 (mean of 3.8) and 4 (mean of 3.9), both performed under the +reasoning conditions, elicited more indicators of intentional reasoning than picture sets 1 (mean of 2.9) and 2 (mean of 2.8), which were performed under the –reasoning conditions (3 > 4 > 2 > 1).

TABLE 10
Intentional Reasoning Measures (n=10)

Picture Set/Condition	*Mean*	*SD*
1: +contextual support, - reasoning demands	2.90	1.91
2: - contextual support, - reasoning demands	2.80	1.23
3: +contextual support, +reasoning demands	3.80	1.69
4: - contextual support, +reasoning demands	3.90	.74

Again, repeated measures ANOVA were used for analysis and significance tests. The results for evidence of intentional reasoning across the four experimental conditions are presented in Table 11. There were no differences in intentional reasoning among the four experimental conditions. The lack or presence of contextual support had no statistically significant effect on intentional reasoning with an F value of 0 and a p-value of 1. The results also show that whether the participants had to sequence the pictures or not had no impact on the intentional reasoning used in the narrations, though values of the F-test results may be approaching significance ($F=4.19$, $p=.071$). This may be limited evidence that the picture sets used for conditions 3 and 4 required more intentional reasoning than the sets used in conditions 1 and 2.

TABLE 11
Repeated Measures Analysis for Intentional Reasoning (n=10)

Condition Comparison	*Intentional Reasoning F (p-value)*
+contextual support vs. –contextual support	0 (p=1.00)
+reasoning demands vs. –reasoning demands	4.190 (p=.071)
+reasoning +support vs. +reasoning –support	.033 (p=.859)

Another component of reasoning demands analyzed in this study has been identified by Robinson (2007a) as spatial reasoning. Since spatial reasoning entails reference to spatial locations (Robinson, 2007a, 2007b), the measure employed in this study was the number of prepositions of location used in the narrating of the picture sets (e.g., *in, on, behind*, etc.). It would stand to reason that tasks requiring more references to location, and/or relative location between objects or people, would require greater use of these prepositions. The descriptive statistics for the frequency of these prepositions are given in Table 12. The picture set used in condition 3 (mean of 6.1) elicited the most evidence of spatial reasoning, followed by the pictures sets used in conditions 2 (mean of 4), then in 4 (mean of 3.4), then in 1 (mean of 2.5), suggesting that the picture sets used for the four experimental conditions may have differed according to the spatial reasoning demands (3>2>4>1) imposed on the participants' performance.

TABLE 12
Spatial Reasoning Measures (n=10)

Picture Set/Condition	*Mean*	*SD*
1: +contextual support, - reasoning demands	2.50	1.08
2: - contextual support, - reasoning demands	4.00	1.24
3: +contextual support, +reasoning demands	6.10	1.79
4: - contextual support, +reasoning demands	3.40	1.07

Again, repeated measures ANOVA were employed to examine evidence of spatial reasoning across the four experimental conditions. The results are shown in Table 13. It appears that the picture sets used to contrast contextual support (sets used in conditions 1 and 3 vs. 2 and 4) did not have a significant effect on spatial reasoning (F=1.97, p=.19). However, statistically significant effects for spatial reasoning were found for the picture sets used in conditions 3 and 4, which required sequencing pictures before narrating, when compared to picture sets used in conditions 1 and 2, which did not require sequencing (F=12.27, p<.01). Furthermore, it was found that the presence of contextual support led to more spatial reasoning (mean of 6.1 vs. mean of 3.4; F=23.35, p<.01) when the participants were narrating the picture sets used in conditions 3 and 4.

TABLE 13
Repeated Measures Analysis for Spatial Reasoning (n=10)

Condition Comparison	*Intentional Reasoning F (p-value)*
+contextual support vs. –contextual support	1.976 (p=.193)
+reasoning demands vs. –reasoning demands	12.270 (p<.01)*
+reasoning +support vs. +reasoning –support	23.350 (p<.001)*

The third component of reasoning demands analyzed in this study is identified in the Cognition Hypothesis as causal reasoning. Following suggestions provided by Robinson (2005), logical subordinators (e.g., *so, therefore, because*) were counted as evidence of causal reasoning. The descriptive statistics are provided in Table 14. With just a cursory glance at these numbers, picture sets used in conditions 3 (mean of 4.4) and 4 (mean of 3.3) yielded more evidence of spatial reasoning than pictures sets used in conditions 1 (mean of 2) and 2 (mean of 1.6), yielding a sequence 3>4>1>2.

TABLE 14
Causal Reasoning Measures (n=10)

Picture Set/Condition	*Mean*	*SD*
1: +contextual support, - reasoning demands	2.00	1.25
2: - contextual support, - reasoning demands	1.60	1.07
3: +contextual support, +reasoning demands	4.40	2.17
4: - contextual support, +reasoning demands	3.30	2.45

Again, repeated measures ANOVA were employed to examine evidence of causal reasoning across the four experimental conditions. The results are shown in Table 15. Mirroring the descriptive statistics, statistically significant effects for spatial reasoning were found for picture sets used in conditions 3 and 4, which required sequencing pictures before narrating, when compared to picture sets used in conditions 1 and 2, which did not require sequencing. (F=12.27, p<.01).

TABLE 15
Repeated Measures Analysis for Causal Reasoning (n=10)

Condition Comparison	*Intentional Reasoning F (p-value)*
+contextual support vs. –contextual support	3.86 (p=.081)
+reasoning demands vs. –reasoning demands	8.85 (p<.05)*
+reasoning +support vs. +reasoning –support	4.05 (p=.075)

In sum, the results show that there were no statistically significant differences among the experimental conditions in terms of the number of tokens of intentional reasoning. Those picture sets used for the conditions which required the participants to sequence pictures (conditions 3 and 4), however, elicited more tokens of both spatial and causal reasoning. There was also more evidence of spatial reasoning in picture set 3, which required participants to sequence pictures and then narrate with the pictures available to them, than in any other picture set.

So far the results have shown that reasoning demands had an effect on the content complexity of the participants' performance, but no effect on the syntactic complexity of their performance. Since picture sets were inseparable from conditions in this study, these results warranted a closer examination of the picture sets themselves. For example, picture set 1 was always performed under the + contextual support/-reasoning demands condition, just as picture set 2 was always performed under the -contextual support/-reasoning demands condition, and so on. As described in the methodology section, the picture sets were considered equal in terms of the content characteristics at the study design stage. Each of the picture sets had a single protagonist, who initiated situations in which conflict arose. Therefore, in the next step of analysis, the picture sets themselves were scrutinized for inherent characteristics, which, independent of experimental conditions, may have impacted the performance of the participants.

In order to examine the content complexity of the picture sets, data from the second comparison group were analyzed. The second comparison group, as described earlier, were a group of L1 English speakers who were asked to rate the "difficulty" of narrating the four picture sets. Therefore, a rank of 1 was given to what the participants perceived as the most difficult picture set, and a rank of 4 was given to what the participants perceived as the least difficult picture set. This group was not subjected to any sequencing conditions, and contextual support was available to them, as they could refer to all the pictures. The results are shown in Table 16.

TABLE 16
Ranking of Difficulty by the Second Comparison Group (n=10)

Picture set	*Condition*	*Mean*	*SD*
3	3	1.70	.675
4	4	2.10	.876
1	1	2.50	1.27
2	2	3.70	.483

Comparing the results from the second comparison group in Table 16 to the descriptive statistics from the experimental (Table 6) and first comparison groups (Table 8), similar patterns emerge. The rankings of difficulty by the second comparison group almost mirror the rankings of syntactic and content complexity of the other groups, suggesting that input (i.e., the picture sets) may have been a variable. The picture sets were ranked in difficulty in the order of sets 3>4>1>2. This is the exact ranking as seen in the results for content complexity, and close to the order found for syntactic complexity. It appears, therefore, that the participants in the second comparison group were able to predict the results of the study based on their own intuitions about the inherent complexity of the stories told by the picture sets. This also confirms the findings concerning picture sets 3 and 4, and the greater amount of spatial and causal reasoning required to narrate those tasks.

DISCUSSION AND CONCLUSIONS

The first research question in this study asked whether there was a relationship between task complexity, as operationalized by contextual support and reasoning demands, and linguistic complexity of L2 performance. The second research question concerned the nature of that relationship, if found to exist. Based on the results from the study, the answer is inconclusive. In this study, linguistic complexity comprised both content complexity and syntactic complexity. The findings show that there was a statistically significant effect for reasoning demands on content complexity, but not on syntactic complexity, as hypothesized by the Cognition Hypothesis. There were no statistically significant effects found for the role of contextual support, even when combined with reasoning demands on either of the components of linguistic complexity, that is, on syntactic or content complexity.

Nevertheless, the findings, viewed through the lens of the Cognition Hypothesis, shed light on one component of linguistic complexity targeted in this study, that is, content complexity. Using Robinson's (2007a) framework of reasoning demands, it was found that performances under the +reasoning demands conditions (i.e., conditions 3 and 4) were more demanding in terms of causal and spatial reasoning than the performances under –reasoning conditions (i.e., conditions 1 and 2), regardless of whether there was contextual support. These specific differential reasoning demands could explain the initial findings of higher content complexity of the narrations produced under the +reasoning demands conditions.

On the other hand, the Cognition Hypothesis does not make an explicit prediction of the effect of task complexity on content complexity, though it does make a prediction of the effect on syntactic complexity, which was not confirmed by the study. One possible explanation could be that the results are an artifact of the measure of syntactic complexity used in this study. This study utilized an index of syntactic complexity expressed as the number of words per t-unit, for the reason that the measure may capture sources of complexity that are rarely captured by other indices (Bygate, 2001). For example, using s-nodes per t-unit or c-unit, like others have (cf. Robinson, 2005, Skehan & Foster, 1999), may only account for complementation. As such, indices of complementation fail to capture syntactic elements like long noun phrases or infinitival constructions that may also be expressions of cognitive complexity. Therefore, this study followed Bygate (2001) in using words per t-unit in order to measure sources of syntactic complexity besides complementation. However, it seems that the measure employed in this study may have been too general, and, perhaps, somewhat insensitive to the complexity of the elicited narrations. Perhaps another measure of syntactic complexity would have yielded different results. However, even if that is the case, the Cognition Hypothesis is still not encompassing enough. Indeed, it must be noted that not all the putative resource-directing variables necessarily foster conditions in which syntactic complexity would be natural for the completion of a task. For example, the presence or absence of contextual support is theorized by Robinson (2001a, 2001b, 2005, 2007a) to influence the learners' attention to tense and aspect during task production, both notions being extraneous to syntactic complexity. Similarly, having more elements (another resource-directing variable, though not examined in this study) in the task does not necessarily introduce more syntactic complexity either. The sentence, "Jack went up the hill" would not be deemed more syntactically complex than "Jack and Jane went up the hill" using measure of complementation.

To mitigate the shortcomings of those general measures of syntactic complexity, Robinson (2005, 2007b) has called for more specific measures of complexity, such as the measures of intentional, spatial, and causal reasoning employed in this paper. However, measures such as these may not necessarily be able to capture a greater amount of syntactic complexity either, as shown in this paper. Instead, they may only serve as evidence that those reasoning faculties are used by a language learner and are encoded in their production.

In terms of increasing task complexity along the dimension of contextual support in the present study, there was no evidence of any effect of the variable by itself. However, some of the analyses have shown that under condition 3, which provided contextual support, more complex language was elicited than under condition 4, which was performed without contextual support. A possible interpretation of this phenomenon may be provided by Revesz (2009), who, in a study that examined the effects of contextual support and recasts on second language development, found that contextual support may have acted as a resource-dispersing factor rather than as a resource-directing one. Revesz argued that there may be a problem with equating +/- here-and-now with +/- contextual support. Indeed, the absence of contextual support could be viewed as imposing a dual task on the learner because "[this] absence requires an additional effort to memorize, and later retrieve, the details of the prompt employed in the task" (Revesz, 2009, p. 446). As such, this dual task of memorizing and retrieval would be classified as a resource-dispersing, and not a resource-directing, factor according to the Cognition Hypothesis. If this is the case, then in the present study, resource-directing and resource-dispersing factors may have been working in concert. In fact, Robinson (2005, 2007b) predicted that "the effects of increasing complexity along resource-directing dimensions should be greater when these are simultaneously simpler along the resource-depleting dimensions" (Robinson, 2001, p. 35). In other words, resource-directing factors have greater force if resource-dispersing factors are minimized. Therefore, if, in the present study, contextual support acted as a resource-dispersing factor, it would stand to reason that condition 3 (+reasoning, +contextual support) elicited more complex language than condition 4 (+reasoning, -contextual support). Skehan (1996, 1998), on the other hand, characterizes these cognitive factors of task complexity as, in Robinson's terms, resource-dispersing variables. Because of the greater linguistic complexity seen under +reasoning conditions, however, his Trade-Off Hypothesis was not supported.

The discussion so far assumes that the input, that is, the content contained in the pictures used in this study, was identical in terms of their intrinsic complexity. In other words, the other variable of interest, that is reasoning demands, was only operationalized in terms of whether or not pictures were sequenced, and not whether the stories told in those picture sets were equal to each other in terms of the content of those series or in terms of the reasoning demands imposed by the sets. However, a comparison of the picture sets, using finer-grained components of reasoning demands, that is, intentional, spatial, and causal reasoning, revealed inequalities which potentially undermine the original design and intent of the study. By operationalizing +/- reasoning demands as the task of putting picture sequences in order, the study ignored the conflating problem of whether the picture sequences were equal in all other respects. In other words, if reasoning demands are to be operationalized as they were in the study, each of the sets of pictures should demand the same amount of intentional, spatial, and causal reasoning.

Consequently, the relationship between task complexity and the linguistic complexity of L2 production seen in this study could be interpreted as an artifact of the picture sequences themselves, and not the results of the conditions employed by the researcher. The results from the second comparison group lend additional support to that idea. Even without sequencing or narrating the picture sets, the second comparison group was able to discern the apparent difficulty of doing so by using common-sense intuitions about the demands imposed by the pictures themselves, without consideration of experimental conditions.

LIMITATIONS AND FUTURE DIRECTIONS

One limitation of the study is related to the possibility that differential reasoning demands were imposed by the picture sets themselves, and not by the experimental conditions. The tasks which were to be made more complex by having participants sequence the pictures, may have been inherently more complex in terms of the reasoning demanded by the picture sets. Therefore, many sources of reasoning demands provided in the input could have been conflated with each other. In other words, the lack of control over content equivalence across the picture prompts may have mitigated the design of the study.

Another limitation of the present study that needs to be discussed is that it provides only an external perspective on reasoning demands and contextual support. An internal perspective would be very valuable, as participants could describe more precisely the types of mental activity they engage in as they narrate under different experimental conditions. These qualitative accounts could then be used to corroborate or qualify some of the findings obtained through the external measures.

Concerning the measures used in this study, the indices of content complexity (i.e., i-units) and syntactic complexity (i.e., words per t-units) were in many respects similar. The operationalization of idea units as single clauses, including infinitival, gerundive, nominal verb phrases and conjucts, all syntactic constructs, resulted in a measure of content complexity that was heavily syntactic in nature. Moreover, as discussed before, the measure of syntactic complexity itself may not have been sensitive enough to reflect the complexity of L2 performance vis-à-vis the experimental conditions.

For future studies, steps need to be taken to link L2 production with L2 acquisition. Even if tasks which have greater reasoning demands elicit "reasoning language," it remains to be seen whether such tasks, over time, and over increased reasoning demands, lead to second language development. Similarly, tasks that are complex along other resource-directing dimensions should be examined for what linguistic features they elicit.

ACKNOWLEDGMENTS

I would like to thank my study participants for making this investigation possible. I would also like to express my gratitude to Prof. ZhaoHong Han for her guidance and insight. I am also indebted to Dr. Monika Ekiert for her invaluable assistance at every stage of this project. Finally, I would like to express my appreciation for my wife, Tao-Lin Carol Choong, and the rest of my family for their continuing support of my studies.

REFERENCES

Bardovi-Harlig, K. (1994). Reverse-order reports and the acquisition of tense: Beyond the principle of chronological order. *Language Learning, 30*, 433-447.

Becker, A., & Carroll, M. (Eds.) (1997). *The acquisition of spatial relations in a second language*. Amsterdam: John Benjamins.

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

Brown, R., & Bellugi, U. (1964). Three processes in the child's acquisition of syntax. *Harvard Educational Review, 34*, 133-151.

Bygate, M. (2001). Effects of task repetition on the structure and control of oral language. In M. Bygate, P. Skehan & M. Swain (Eds.), *Researching pedagogic tasks: Second language learning, teaching, and testing* (pp. 23-48). Harlow: Longman.

Carrell, P. (1985). Facilitating ESL reading by teaching text structure. *TESOL Quarterly, 19*(4), 727-752.

Cornell, E., Heth, D., & Alberts, D. (1994). Place recognition and way finding by children and adults. *Memory and Cognition, 22*, 633-643.

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

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

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

Housen, A., & Kuiken, F. (2009). Complexity, accuracy, and fluency in second language acquisition. *Applied Linguistics, 30*(4), 461-473.

Iwashita, N., Elder, C., & McNamara, T. (2001). Can we predict task difficulty in an oral proficiency test? Exploring the potential of an information-processing approach to task design. *Language Learning*, *51*, 401-436.

Lee, E., & Rescorla, L. (2002). The use of psychological state terms by late talkers at age 3. *Applied Psycholinguistics*, *23*, 623-641.

Long, M. H. (2000). Focus on form in task-based language teaching. In R. Lambert & E. Shohamy (Eds.), *Language policy and pedagogy: Essays in honor of A. Ronald Walton* (pp. 179-192). Philadelphia: John Benjamins.

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

Olson, D., Torrance, N., & Hildyard, A. (1985). *Literacy, language, and learning: the nature and consequences of reading and writing*. Cambridge: Cambridge University Press.

Revesz, A. J. (2009). Task complexity, focus on form, and second language development. *Studies in Second Language Acquisition*, *31*, 437-470.

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, UK: 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 grading and sequencing pedagogic tasks. In M. Garcia Mayo (Ed.), *Investigating tasks in formal language learning* (pp. 7-27). Clevedon, UK: Multilingual Matters.

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

Robinson, P., Cardierno, T., & Shirai, Y. (2009). Time and motion: Measuring the effects of the conceptual demands of tasks on second language speech production. *Applied Linguistics, 30*(4), 533-554.

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

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

Skehan, P. (1998). Task-based instruction. *Annual Review of Applied Linguistics, 18*, 268-286.

Skehan, P., & Foster, P. (1999). The influence of task structure and processing conditions on narrative retellings *Language Learning*, *49*, 93-120.

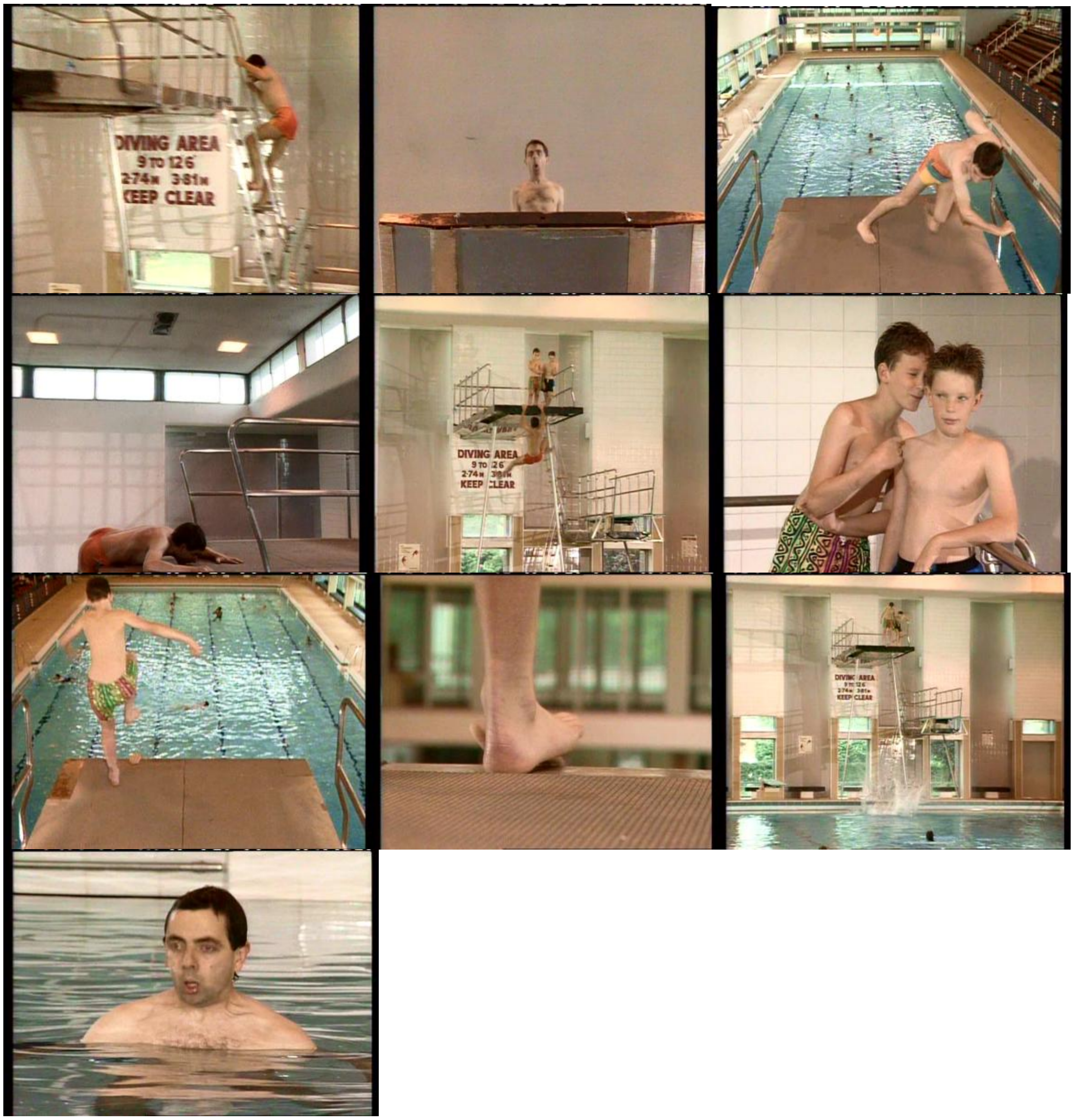
Skehan, P., & Foster, P. (2001). Cognition and tasks. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 183-205). New York: Cambridge University Press.

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

APPENDIX

Picture Sequence 1

Picture Sequence 2



Picture Sequence 3

Picture Sequence 4