The Journal of Asia TEFL

Conditions and Tasks: The Effects of Planning and Task Complexity on L2 Speaking

Nayoung Kim

Korea Polytechnics, South Korea

Motivated by different notion of task criteria in the Limited Attention Capacity Hypothesis and the Cognition Hypothesis, the purpose of this study is to investigate how planning and task complexity affect second language (L2) learners' spoken performance. The study's motivation is as follows: 1) the effects of online planning (as task condition) and task complexity (as task characteristics) and their interactions were examined; and 2) instead of repeated and redundant measures for complexity, accuracy, and fluency (CAF), multi-layered CAF measure indices were used to capture a careful performance. The study's subjects were 77 undergraduates divided into four groups: no planning in simple task, no planning in complex task, online planning in simple task, and online planning in complex task. Online planning was modified from Yuan and Ellis's (2003) and Wang's (2009) studies, while task complexity was manipulated by [±elements] and [±reasoning]. The study employed 11 general and specific measure indices for CAF and found that no planning condition significantly enhanced overall accuracy and verb-related specific accuracy or speed and repair fluency regardless of task complexity. Meanwhile, task complexity significantly improved phrasal syntactic complexity regardless of planning. Based on these findings, the study explored two competing theories on L2 speaking.

Keywords: task condition, task characteristics, planning, task complexity, limited attention capacity hypothesis, cognition hypothesis, CAF

Introduction

In psycholinguistic task-based research, both the Limited Attention Capacity Hypothesis (LACH)[1] and the Cognition Hypothesis (CH) have received significant attention, particularly as task sequencing criteria in task-based language teaching (TBLT). Both hypotheses share the same assumption that tasks are a central unit of analysis in language teaching and that tasks should be sequenced in terms of certain sequencing criteria. However, the LACH and CH differ in how they define task characteristics and/or task conditions, allowing language teachers to decide on task sequences and how they affect second language (L2) learners' task performance in terms of complexity, accuracy, and fluency (CAF). In other words, the LACH claims that both task characteristics and task conditions can be considered in order to grade task difficulty, and those difficult tasks lead to an improvement of either complexity or accuracy at the expense of the other due to L2 learners' limited attentional resources (Skehan, 2016). However, the CH claims that only task characteristics (i.e., task complexity) is a sole basis for grading task complexity, and that complex tasks lead to an increase in both complexity and accuracy simultaneously (Robinson, 2015).

Norris (2010) argued that, to test the LACH and CH, researchers often manipulate a task into a simple and complex task and then measure typical CAF under different task characteristics. If the predicted results are found between the two tasks, they are used as evidence in favor of the LACH or CH. Révész

(2014) points out:

> Researchers need to supply specific and separate evidence for all proposed factors, from the independent variables including cognitive task complexity, through the explanatory processes of speech production and attentional allocation, to the dependent variables of CAF measures. (p. 88)

Influenced by the two competing theories, LACH and CH, to view task criteria differently on L2 spoken performance, the present study was motivated by the following issues. First, new independent variable should be tested. There is none of studies (in my knowledge) to examine the relations between online planning and task complexity on L2 speaking (except in Kim's (2016) study[2] on L2 writing). Following the CH, the relations of pre-task planning and task complexity were popularly examined (Gilabert, 2007; Sasayama & Izumi, 2012). However, online planning with task complexity on L2 speaking was not explored due to vague task criteria of online planning in the CH although online planning occurs frequently when L2 learners speak and thus the LACH-oriented studies (Wang, 2009; Yuan & Ellis, 2003) have treated this importantly. Second, redundant usage of dependent variables (CAF) should be avoided. For instance, many measures of syntactic complexity heavily rely on subordinating complexity (e.g., dependent clauses per clause, relative clauses per T-unit, and S-nodes per T-unit). Therefore, it is necessary to use more various and different measure indices to earn the reliable results of task-based research. In terms of these concerns, the present study attempts to investigate the effects of planning (no planning vs. online planning) and task complexity (simple vs. complex tasks) on intermediate L2 learners' spoken performance, syntactic complexity, accuracy, and fluency. By manipulating two independent variables in a controlled laboratory setting, the study examines the influence of task conditions (i.e., planning) and/or task characteristics (i.e., task complexity) on the multilayered measures of CAF.

Literature Review

Task Conditions vs. Task Characteristics

The LACH and CH have presented specific rationales to justify task difficulty/task complexity. The central pedagogical purpose of the LACH and CH in task design is to enable the sequencing of tasks from simple to difficult/complex tasks for L2 learning and teaching. Given this purpose, it is important to uncover how task conditions and task characteristics produce expected outcomes in L2 performance. However, these two hypotheses are built on contrasting theoretical foundations and models; thus, their effects on second language (L2) performance in terms of (syntactic and lexical) complexity, accuracy, and fluency (CAF) are expected to be different (see Figure 1).

The LACH, based on L2 learners' limited attentional capacity, proposes a three-way distinction for task analysis: *cognitive complexity, code complexity*, and *communicative stress* (Skehan, 1998, 2003). Recently, Skehan (2009, 2014, 2015, 2016) explained the LACH in two ways: 1) *task characteristics*, which determine task difficulty based on tasks' own features; and 2) *task conditions*, which determine task difficulty based on task performance. Skehan (2015) emphasized "the need to consider conceptualization and formulation/articulation as distinct from one another in the way they impact task difficulty" (pp. 127-128). In other words, the factors that make conceptualization or formulation more demanding are different. For instance, task familiarity, cognitively complex manipulation of information, task types, and pre-task planning for generating ideas place more demands on conceptualization, whereas tasks' lexical and syntactic demands, general time pressure conditions, and online planning tend to place more demands on formulation. Another key issue between task characteristics and conditions is that they both contribute to task difficulty and L2 performance (CAF) separately and/or together. Thus, Skehan (2014) claims that:

> Any task is likely to subsume a bundle of (task) features, and… tasks, given their nature, are likely to be strongly influenced by context, and so what is difficult in one (pedagogic) context or for one particular learner may not be so difficult in another, through, for example, cultural knowledge or age differences. (p. 6)

Therefore, a difficult task is affected by both task characteristics and conditions, and it can be interpreted differently by different learners and/or in different contexts based on the LACH.
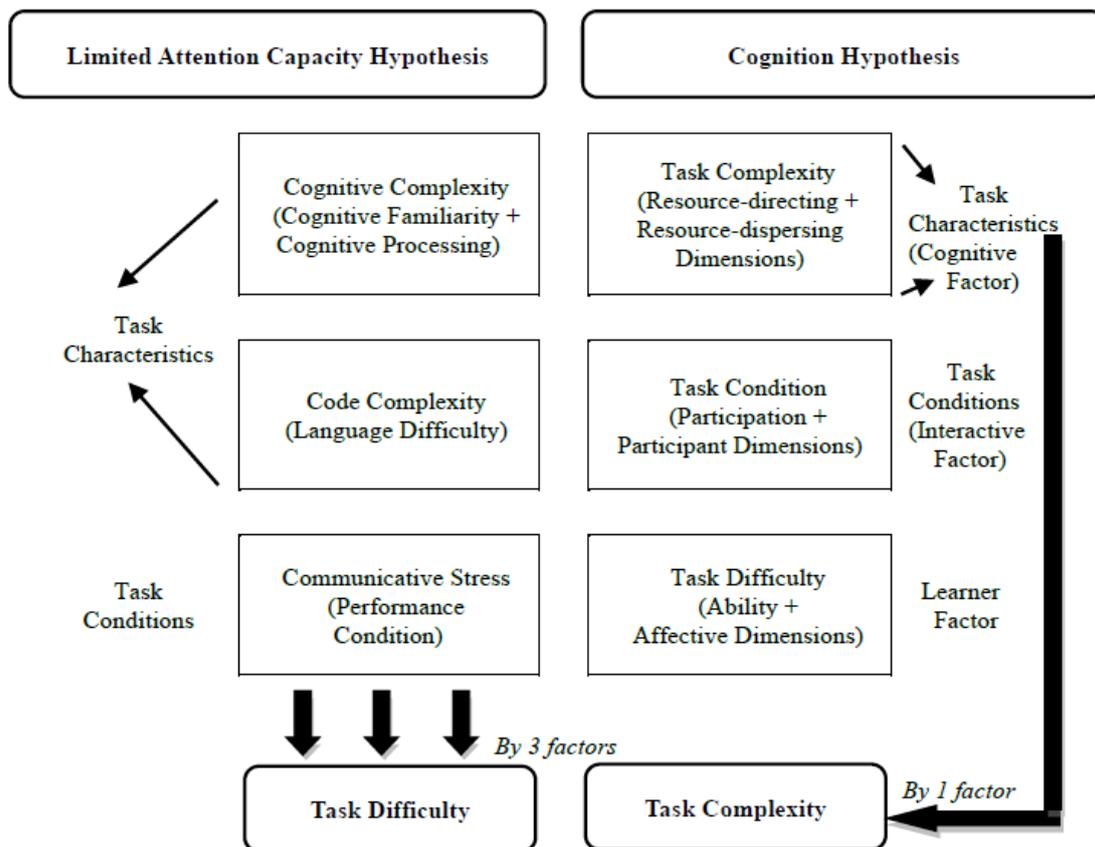

Figure 1. Two different models: LACH and CH.

On the other hand, the CH provides a triadic componential framework for task sequencing consisting of *task complexity, task condition*, and *task difficulty* (Robinson, 2001, 2005, 2007, 2011, 2015). First, *task complexity* refers to cognitive factors that affect tasks' intrinsic cognitive demands. Robinson (2011) claims that:

> Since cognitive complexity is the sole basis of sequencing decisions and that therefore given a sufficient detailed taxonomy of task characteristics affecting their cognitive complexity, then this should also be a pedagogically feasible basis for decision-making by task. (p. 10)

Task complexity consists of the *resource-directing dimension* and *resource-dispersing dimension*. The resource-direction dimension refers to task characteristics that make cognitive and conceptual demands on cognition. Robinson (2015) stated, "languages grammaticize, lexicalize, and syntacticize these conceptual domains in similar and also differing ways" (p. 98). Thus, when the resource-directing dimension increases from simple to complex, it leads to interlanguage development. In contrast, the resource-dispersing dimension refers to task characteristics that make performative and procedural

demands on cognition; however, increasing the resource-dispersing dimension does not direct "learner attention and efforts at conceptualization to any particular aspects of language code" (Robinson, 2015, p. 98). The resource-direction dimension includes [±elements], [±reasoning], and [±perspective-taking] while the resource-dispersing dimension includes [±planning], [±task structure], and [±prior knowledge]. Second, *task condition* refers to the interactive factors affecting the nature and amount of interaction between participants and tasks. Task condition consists of the *interactional dimension*—including [±few participants], [±one-way flow], and [±few contributions needed]—and *interactant dimension*—including [±same proficiency], [±same gender], [±shared cultural knowledge], and [±equal status and role]. Third, *task difficulty* refers to learner factors, consisting of the ability dimension and the affective dimension. The former includes [high/low working memory], [high/low aptitude], and [high/low task-switching] while the latter includes [high/low openness to experience], [high/low task motivation], [high/low willingness to communicate], and [high/low control of emotion].

Although Robinson mentioned task complexity as a task characteristic that affects tasks' cognitive complexity directly, he did not clearly distinguish between task characteristics and conditions in the CH. Therefore, the present study[3] uses these two terms following Skehan's (2016) justification that task characteristics concern tasks' inherent features (task types, task structure, task complexity) while task conditions concern "choices that can be made about how tasks are implemented" (p. 35), including planning, post-task, task repetition, and monologic/interactive participation. Thus, this study employs two independent variables—task conditions (planning) and task characteristics (task complexity)—to examine the LACH and CH models.

Task Condition: Planning

Ellis (2005, 2009) claimed that *pre-task planning*—which occurs before learners perform a task and includes the selection of content and linguistic forms—is classified into *rehearsal* and *strategic planning*. Meanwhile, *online planning*—which occurs while learners perform a task—is classified into *pressured online planning* and *unpressured online planning*. In the present study, *online planning* refers to *unpressured online planning*, and *no planning* refers to a condition where the least possible planning chance is given before or during the task (this is the same condition as *pressured online planning condition* or *pressured speaking condition* in Ellis's definition).

Planning as a task condition has been explained theoretically through Levelt's model of language speaking (Ellis, 2005; Skehan, 1998, 2009, 2015). Levelt (1989, 1999) distinguished three components of speaking processing: 1) *conceptualization*, where the speaker establishes the goal and generates content ideas; 2) *formulation*, where the speaker encodes messages into lexical, synthetic, and phonological items; and 3) *articulation*, where the speaker produces actual speech. Skehan (1998) proposed that L2 learners have a limited attention capacity; thus, they focus on one performance aspect at the expense of others, which is also the LACH's theoretical foundation. In terms of Levelt's (1989, 1999) speaking process, pre-task planning supports conceptualization leading to fluency while online planning stimulates formulation leading to accuracy. (Due to the present study's focus on online planning and no planning in L2 speaking, the relevant findings on oral tasks are presented here.)

In Ellis and Yuan's (2005) study, 42 Chinese, college-level learners of English were divided into three groups: 1) pressured speaking and then online planning in writing (*n* = 14); 2) online planning in speaking (*n* = 14); and 3) pressured writing (*n* = 14). Participants had 30 seconds to view pictures and were then asked to either speak for a limited time (5 minutes) under pressured speaking conditions or speak for unlimited time under unpressured online planning conditions. The findings with statistical significance in speaking revealed that unpressured online planning led to improvements in syntactic complexity (clauses/T-units), syntactic variety (number of different verb forms), and accuracy (proportion of error-free clauses and correct verb forms)—but not fluency—compared to pressured online planning. Similarly, Yuan and Ellis (2003) compared three groups: 1) no planning (0.5 minutes to view pictures plus 5 minutes to tell a story); 2) pre-task planning (10 minutes to view pictures and plan the story plus 5

minutes to tell a story); and 3) online planning (0.5 minutes to view pictures plus unlimited time to tell a story). The findings revealed that online planning increased syntactic complexity and accuracy compared to no planning. Yuan and Ellis (2003) concluded that L2 learners' online planning allows them to access formulation due to increased opportunity to plan, leading to the concurrent increase in complexity and accuracy.

However, other scholars (Kim, 2018; Skehan, Xiaoyue, Qian, & Wang, 2012; Wang, 2009, 2014) have argued that the online planning condition is problematic because actual planning while performing tasks is difficult to control. Moreover, an "unlimited amount of time" may not differ from a "limited amount of time" particularly in speaking because speaking itself is a more pressured condition (than writing). In a carefully controlled online planning studies on L2 spoken tasks (Kim, 2018; Wang, 2009, 2014), online planning did not increase accuracy without the help of pre-task planning. For example, Wang (2009) slowed down a *Mr. Bean* video by 60% in oral narrative tasks and set six planning conditions: 1) control condition (watching the video and narrating the story simultaneously); 2) watched condition (watching the video and then watching and narrating simultaneously); 3) watched and strategic condition (watching the video, then planning for 10 minutes, and then watching and narrating simultaneously); 4) (pure) online planning condition (watching a slower version of the video and narrating the story simultaneously); 5) watched online planning condition (watching the video and then watching a slower version of the video and narrating simultaneously); and 6) task repetition condition (watching and narrating simultaneously and then repeating the task). The findings revealed that pure online planning did not increase CAF, but the watched online planning condition improved both complexity and accuracy. Wang (2009) suggested that the "watched" condition facilitates conceptualization while online planning facilitates formulation. Therefore, the effects of carefully controlled online planning (in terms of speaking) on CAF are positive when "something more is needed to trigger its effectiveness" (p. 176).

Task Characteristics: Task Complexity

Task complexity manipulated by resource-directing and resource-dispersing dimensions is task characteristics (Robinson, 2011). One of the key features of the cognition hypothesis is that complex tasks improve both complexity and accuracy simultaneously by demanding greater conceptualization efforts while preparing content, leading to a "paring down of conceptual information into a linguistically relevant representation for subsequent encoding at the microplanning stage with positive consequences" (Robinson, 2011, p. 15). In the present study, task complexity is controlled by [±elements] and [±reasoning]; thus, empirical studies related to these factors are reviewed below.

Ishikawa (2008) manipulated task complexity by [±elements] and [±reasoning] in three control, simple, and complex task groups. Participants were asked to leave an answering machine message explaining human relationships. The simple task group explained two people's relationship and its changes using simple reasons presented in the picture, whereas the complex task group explained four people's relationship and its changes using complex reasons. The control group explained four people's relationships without reasoning. The findings revealed that the simple task produced greater lexical and syntactic complexity and accuracy than the control (no reasoning) task. The control task led to greater fluency than the other tasks. Interestingly, increases in lexical and syntactic complexity and accuracy were found in the simple tasks—not in complex tasks—compared to the control tasks.

Contrastingly, in Michel, Kuiken, and Vedder (2012), participants were also asked to leave a phone message for a friend in need of advice, and task complexity was controlled by [±elements]. The study found that the complex task improved lexical complexity and accuracy, but not syntactic complexity and fluency. Moreover, Sasayama and Izumi (2012) manipulated task complexity by [±elements], producing a simple task with fewer characters and a complex task with more characters in picture-based stories. The researchers found that complex tasks positively affect the specific measure of syntactic complexity (noun modifiers) but negatively affect global accuracy and fluency. Thus, task complexity manipulated by [±elements] and/or [±reasoning] does not seem to improve both syntactic complexity and accuracy

simultaneously, as Robinson (2001, 2005, 2007, 2011) has claimed.

Skehan (2016) argued that the task characteristic of number of elements/reasoning has little overall effect on both structural complexity and accuracy. Skehan (2015, 2016) concluded that the CH provides well-organized task characteristics through task complexity theoretically, yet task complexity has not been supported fully by empirical studies (e.g., Ishikawa, 2008; Michel, Kuiken, & Vedder, 2012; Révész, 2011; Robinson, Cadierno, & Shirai, 2009; Sasayama & Izumi, 2012). Therefore, the effects of task complexity in speaking on CAF should be carefully investigated.

Research Questions

In order to fill the gap in previous task-based research on the LACH and CH, the present study investigates the effects of planning (as task conditions) and task complexity (as task characteristics) in spoken performance on multilayered CAF. In light of the previous review, this study addresses the following research questions:

(1) To what extent do planning and task complexity influence syntactic complexity in college L2 learners' spoken performance?
(2) To what extent do planning and task complexity influence accuracy in college L2 learners' spoken performance?
(3) To what extent do planning and task complexity influence fluency in college L2 learners' spoken performance?

Methods

Participants

This study employed a 2x2 research design in a laboratory setting at a private college in Seoul, South Korea. The participants were 77 Korean female undergraduates. The average participant age was 20.09 years old, and academic years included freshman ($n = 35$), sophomore ($n = 20$), junior ($n = 9$), and senior ($n = 13$). Participants were majoring in different subjects. Based on an earlier study (Kim, 2016), this study assumed that the independent variable of planning would work best with intermediate proficiency L2 learners rather than novice or advanced learners. Thus, intermediate proficiency learners who had achieved Level 2 (intermediate level) within six months according to the General English Language Test (GELT)[4] were voluntarily recruited. Participants' English proficiency was determined again through a proficiency assigning task,[5] and their intermediate proficiency level was rated and confirmed by the researcher and a native English speaker using the speaking proficiency guidelines from the American Council on the Teaching of Foreign Languages (Swender, Conrad, & Vicars, 2012). Each participant received a numeric score from the two raters: 1 (Novice High), 2 (Intermediate Low), 3 (Intermediate Mid), 4 (Intermediate High), or 5 (Advanced Low). The 77 participants had a finalized proficiency level of intermediate ranging from Intermediate Low ($n = 11$) to Intermediate Mid ($n = 54$) to Intermediate High ($n = 12$) (Cronbach's alpha = .862). If the assigned numeric score differed by more than two levels between the raters, they came to an agreement through a one-on-one conference. Thus, participants were divided into four groups for the main task: no planning in simple task (NPS, $n = 18$), no planning in complex task (NPC, $n = 19$), online planning in simple task (OPS, $n = 20$), and online planning in complex task (OPC, $n = 20$).

Main Task Treatment

The main task for the study's four groups was a narrative spoken task requiring participants to tell a

story based on a series of pictures. The task was manipulated by two independent variables: planning (no planning and online planning) and task complexity (simple task and complex task). Following Yuan and Ellis (2003), no planning—which is a fast, high-pressure speaking condition—was the condition restricting both pre-task planning and online planning. Modifying three prior studies (Kim, 2018; Wang, 2009; Yuan & Ellis, 2003), online planning in this study was the unpressured online planning condition in which pre-task planning was restricted; however, thinking while performing the spoken task was allowed. Moreover, all participants were asked to use the total extended amount of time for speaking. In terms of the pilot study, the pressured no planning condition was set for two minutes, while the online planning condition was set for four minutes. During the pilot study following Yuan and Ellis's (2003) study, the participants were asked to record their narration as fast as possible while seeing the pictures, and then their different time spent on speaking was detected. Two minutes were set because it was the fastest time among the participants' fast and pressured time. It was assumed that it would be pressured. Online planning time was modified from Wang's (2009) study in which her online planning condition was 60% slower video for speaking. Four minutes were set based on 50% slower and longer than the normal speaking time (2m 40s) for narrating the pictures during the pilot study. In addition, task complexity was manipulated by [±elements] and [±reasoning] in the resource-direction dimension. The simple task adopted from Heaton (1975) featured a smaller number of characters (two main characters) in the story, and the reason leading to the event's problems was relatively simple. The complex task adopted from Yule (1997) featured a greater number of characters (three main characters and two minor characters) and the reason for the event's problems was relatively complex.

Experimental Procedure

The experiment was carried out in a laboratory setting featuring the proficiency-assigning task and the main task. The L2 learners who scored a Level 2 (intermediate level) on the GELT within six months participated in the study voluntarily. Participants entered the laboratory individually, and each time the researcher explained the instructions in their first language (Korean) while also providing written directions. For the proficiency-assigning task, four topics about the past personal story were given; each participant chose one topic to speak about for two minutes. While telling and recording their story, participants were asked to check the time on a timer located in front of them (see Figure 2).

Subsequently, for the main task, each participant was assigned randomly to one of the four groups (NPS, NPC, OPS, or OPC). All participants were asked to view a series of pictures (simple version or complex version) for one minute which was set from the pilot study. Note-taking was not allowed while viewing the pictures. In the NPS and NPC conditions, participants were required to tell the story within two minutes, which were set following the pilot study as a fast and pressured speaking condition. As they did previously, participants also had to check the time on the timer. The difference between the NPS and NPC conditions was a series of pictures manipulated differently by [±elements] and [±reasoning] of task complexity. In the OPS and OPC conditions, participants were required to tell the story within four minutes, which were set following the pilot study as an extended and unpressured speaking condition. Participants were asked to use up the whole four minutes while looking at the time on the timer. The difference between the OPS and OPC conditions was a series of pictures controlled differently by [±elements] and [±reasoning]. After finishing the main task, participants were required to fill out a background questionnaire.
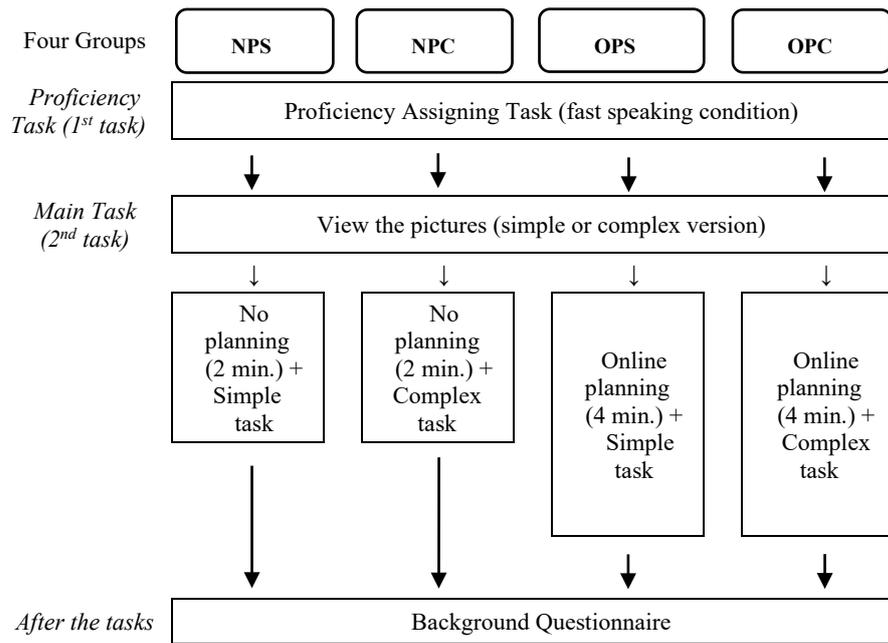
Figure 2. Experimental procedure.

Measurement and Data Analysis

Participants' oral performance was measured through syntactic complexity, accuracy, and fluency (CAF) (see Table 1). First, based on Norris and Ortega (2009), the following five indices were selected to measure syntactic complexity: 1) the mean length of T-unit (MLT); 2) the Coordination Index (CI); 3) clauses per T-unit (C/T); 4) dependent clauses per clause (DC/C); and 5) the mean length of clauses (MLC). The rationale behind adopting five measuring indices is that a variety of metrics to detect different aspects of syntactic complexity—entailing overall complexity (e.g., MLT), coordinate complexity (e.g., CI), subordinate complexity (e.g., C/T and DC/C), and phrasal complexity (e.g., MLC)—provide a more in-depth understanding of complexity (Bulte & Housen, 2012; Norris & Ortega, 2009). Since the present study's participants were intermediate learners, T-units were selected rather than AS-units, unlike in my previous study (Kim, 2019).

Second, accuracy was measured in terms of global and specific measures modified from Ellis and Yuan (2005) with the following four indices: 1) the number of errors per T-unit (E/T); 2) the number of error-free T-units per T-unit (EFT/T); 3) verb-related errors per T-unit (VE/T); and 4) errors other than verb-related ones per T-unit (OE/T). The rationale for adopting four measuring indices is that general accuracy (e.g., E/T and EFT/T) was chosen considering participants' English level proficiency with T-unit (not AS-unit), whereas context-specific local accuracy (e.g., VE/T and OE/T) was chosen due to task type. Participants would be expected to use the past tense of verbs when narrating the story; thus, appropriate verb usage would be crucial in this context. Verb-related errors include errors in tense, verb forms, subject-verb agreement, missing verbs, unnecessary verbs, and verb inflection. By distinguishing two specific error types (verb errors and non-verb errors), a more insightful pattern of context-specific errors could be examined.

TABLE 1
Eleven Measures of CAF

Syntactic Complexity	General Syntactic Complexity			
	(Overall)	(Coordinate)	(Subordinate)	(Phrasal)
MLT	●			
CI		●		
C/T			●	
DC/C			●	
MLC				●
Accuracy	General		Specific Accuracy	
	(Overall)		(Verb-related)	(Others)
E/T	●			
EFT/T	●			
VE/T			●	
OE/T				●
Fluency	Speed Fluency	Repair Fluency (Restart, Repetition, Self-correction)		Breakdown Fluency
W/M	●			N/A
DysF		●		

Finally, following Tavakoli and Skehan (2005), the following two indices were selected to measure fluency: 1) the number of words per minute (W/M); and 2) dysfluency ratio per 100 seconds (DysF). Fluency is classified into speed fluency, repair fluency, and breakdown fluency (Foster, Tonkyn, & Wigglesworth, 2000; Tavakoli & Skehan, 2005). The rationale for adopting two measuring indices is that excessive amounts of repair—including false starts, repetition, and self-correction—were expected to result in dysfluent speaking even if many words were spoken in terms of speed fluency. Thus, the present study included speed fluency (e.g., W/M) and repair fluency (e.g., DysF). Since pausing behavior for breakdown fluency did not occur frequently, it was not included.

For data analysis, the researcher and two raters (Korean- and English-speaking raters) analyzed the spoken data. For syntactic complexity and fluency, the researcher and Korean rater separately counted the number of clauses, dependent clauses, T-units, coordinate clauses, sentences, words, and dysfluent words. The resulting numbers were compared and, if a difference existed, the final number was recounted through a one-on-one conference. For accuracy, the errors in the present study included morphosyntactic and lexical errors. The researcher and English-speaking rater separately counted the number of errors, error-free T-units, verb-related errors, and non-verb errors for up to 10% of the data. The two compared the resulting data and reached an agreement through a one-on-one conference. Subsequently, the rest of the data was analyzed independently following the conference agreement. The inter-rater reliability for the coding errors was high (Cronbach's alpha = .989). Disagreements were appropriately discussed.

To identify statistical significance, a number of statistical analyses were conducted using the Statistical Package for Social Sciences (SPSS). Since the assumptions of multivariate normality and homogeneity of the covariance were violated and the sample size was not big enough, a multivariate analysis of variance (MANOVA) was not carried out in this study. Instead, univariate analyses of variance (ANOVA) were conducted. First, a series of two-way ANOVA was conducted to calculate the main effects of the study's independent variables (planning and task complexity) and their interaction. In terms of Levene's test (for homogeneity), if the variance was not equal, it led to an increase in Type I error rates. Second, the Breusch-Pagan test (for heteroscedasticity) was conducted to adjust variance, and if this test revealed the presence of conditional heteroscedasticity, the weighted least square method was employed. Among 11 indices, six indices (MLT, C/T, MLC, EFT/T, VE/T, W/M) were detected by two-way ANOVA, three indices (CI, DC/C, and E/T) by the Breusch-Pagan test, and two indices (OE/T and DysF) by the weighted least square method.

Results

Effects of Planning and Task Complexity on Syntactic Complexity

The study's first research question focuses on the effects of planning and task complexity on syntactic complexity in L2 spoken performance. Table 2 outlines descriptive statistics for syntactic complexity in the four groups (NPS, NPC, OPS, and OPC), while Table 3 shows the effects of planning and task complexity and their interaction on syntactic complexity.

TABLE 2
Descriptive Statistics for the Four Experimental Groups on Syntactic Complexity

Measures	Planning	TC	*N*	*M*	*SD*
MLT	No PL	Simple	18	11.88	2.78
		Complex	19	13.16	2.38
	Online PL	Simple	20	11.35	2.53
		Complex	20	12.03	2.12
CI	No PL	Simple	18	43.33	21.91
		Complex	19	39.15	11.95
	Online PL	Simple	20	41.36	15.61
		Complex	20	43.70	12.65
C/T	No PL	Simple	18	1.54	.32
		Complex	19	1.61	.24
	Online PL	Simple	20	1.59	.27
		Complex	20	1.51	.21
DC/C	No PL	Simple	18	.32	.14
		Complex	19	.37	.09
	Online PL	Simple	20	.35	.12
		Complex	20	.32	.10
MLC	No PL	Simple	18	7.85	1.81
		Complex	19	8.21	1.14
	Online PL	Simple	20	7.12	.87
		Complex	20	8.00	1.16

Notes. TC = task complexity; PL = planning.

TABLE 3
Effects of Planning and Task Complexity on Syntactic Complexity

Measures	Between-subjects Effects		
	Groups	*F.*	*Sig.*
MLT	Planning	2.216	.141
	TC	3.074	.084
	PL * TC	.285	.595
CI	Planning	.126	.723
	TC	.064	.800
	PL * TC	.810	.371
C/T	Planning	.240	.625
	TC	.005	.943
	PL * TC	1.501	.225
DC/C	Planning	.097	.756
	TC	.118	.732
	PL * TC	1.779	.186
MLC	Planning	2.568	.113
	TC	4.477*	.038
	PL * TC	.807	.372

Notes. TC = task complexity; PL = planning; $p^* < .05$.

Regarding syntactic complexity, a univariate ANOVA revealed that task complexity had significant effects on MLC ($F = 4.477$, $p = .038$, partial $\eta^2 = .058$), while planning did not have significant effects on MLC ($F = 2.568$, $p = .113$, partial $\eta^2 = .034$). Thus, task complexity improved phrasal complexity and the resource-directing dimension of task complexity increased across different planning conditions. However, neither planning nor task complexity produced significant differences in other indices of syntactic complexity (MLT, CI, C/T, and DC/C), and there was no interaction between planning and task complexity in syntactic complexity. Figure 3 shows how, compared to simple tasks, complex tasks increased phrasal syntactic complexity significantly regardless of planning.
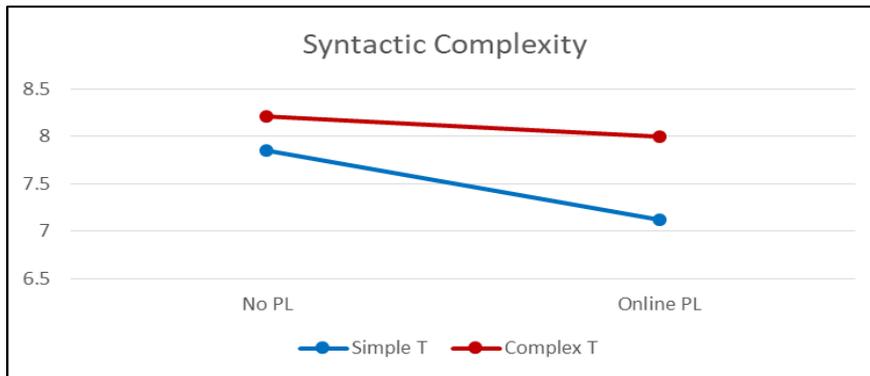

Figure 3. Significant effects of task complexity on phrasal syntactic complexity.

Therefore, task complexity improved phrasal complexity but not other types of syntactic complexity, while planning conditions did not produce an increase in syntactic complexity.

Effects of Planning and Task Complexity on Accuracy

The study's second research question focuses on the effects of planning and task complexity on accuracy in L2 spoken performance. Table 4 outlines descriptive statistics for accuracy in the four groups, and Table 5 shows the effects of planning and task complexity and their interaction on accuracy.
TABLE 4
Descriptive Statistics for the Four Experimental Groups on Accuracy

Measures	Planning	TC	*N*	*M*	*SD*
E/T	No PL	Simple	18	.78	.19
		Complex	19	.76	.33
	Online PL	Simple	20	1.07	.42
		Complex	20	1.01	.34
EFT/T	No PL	Simple	18	.43	.10
		Complex	19	.50	.19
	Online PL	Simple	20	.30	.16
		Complex	20	.36	.14
VE/T	No PL	Simple	18	.62	.21
		Complex	19	.55	.25
	Online PL	Simple	20	.86	.39
		Complex	20	.74	.27
OE/T	No PL	Simple	18	.16	.10
		Complex	19	.22	.20
	Online PL	Simple	20	.21	.14
		Complex	20	.27	.18

Notes. TC = task complexity; PL = planning.

TABLE 5
Effects of Planning and Task Complexity on Accuracy

Measures	Between-subjects Effects		
	Groups	*F.*	*Sig.*
E/T	Planning	12.590***	.001
	TC	.193	.662
	PL * TC	.077	.782
EFT/T	Planning	15.802***	.000
	TC	3.351	.071
	PL * TC	.008	.928
VE/T	Planning	10.844**	.002
	TC	1.944	.167
	PL * TC	.113	.737
OE/T	Planning	1.968	.165
	TC	2.415	.125
	PL * TC	.001	.976

Notes. TC = task complexity; PL = planning; $p^{**} < .01$; $p^{***} < .001$.

Regarding accuracy, a univariate ANOVA revealed that planning had significant effects on on E/T ($F = 12.590$, $p = .001$, partical $\eta^2 = .147$), EFT/T ($F = 15.802$, $p = .000$, partical $\eta^2 = .178$), and VE/T ($F = 10.844$, $p = .002$, partical $\eta^2 = .129$). However, task complexity did not significantly affect the three indices of accuracy. Thus, no planning condition increased accuracy despite different levels of task complexity. Neither planning nor task complexity produce significant differences in OE/T, and there was no interaction between planning and task complexity in accuracy. Figure 4 shows how, compared to online planning, no planning significantly increased accuracy regardless of task complexity.
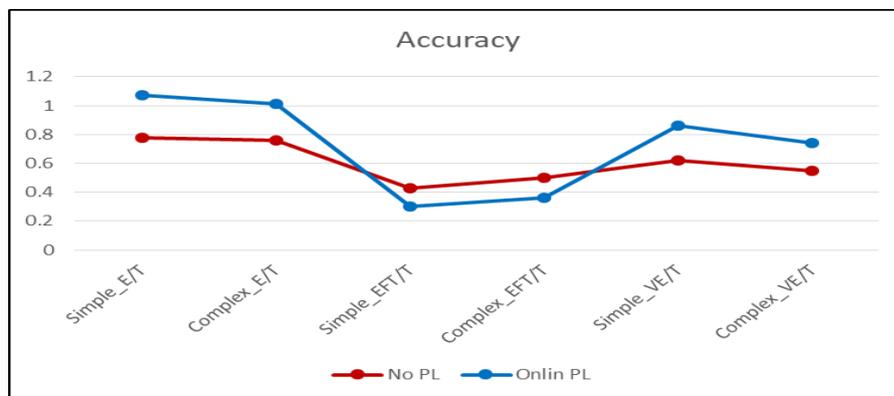

Figure 4. Significant effects of planning on accuracy.

Therefore, the no planning condition improved all accuracy indices (except OE/T) while task complexity did not produce an increase in accuracy.

Effects of Planning and Task Complexity on Fluency

The study's third research question focuses on the effects of planning and task complexity on fluency in L2 spoken performance. Table 6 outlines descriptive statistics for fluency in the four groups, and Table 7 shows the effects of planning and task complexity and their interaction on fluency. Regarding fluency, a univariate ANOVA revealed that planning had significant effects on W/M ($F = 27.817$, $p = .000$, partical $\eta^2 = .276$) and DysF ($F = 19.206$, $p = .000$, partical $\eta^2 = .208$), but that task complexity did not significantly affect W/M ($F = .478$, $p = .492$, partical $\eta^2 = .007$) or DysF ($F = .469$, $p = .496$, partical $\eta^2 = .006$). Thus, the no planning condition increased speed fluency and repair fluency simultaneously,

regardless of how tasks were manipulated by task complexity. No interaction was found between planning and task complexity in fluency.

TABLE 6
Descriptive Statistics for the Four Experimental Groups on Fluency

Measures	Planning	TC	N	M	SD
W/M	No PL	Simple	18	90.12	19.98
		Complex	19	88.50	22.70
	Online PL	Simple	20	68.46	17.12
		Complex	20	64.04	16.50
DysF	No PL	Simple	18	29.31	14.19
		Complex	19	26.84	12.05
	Online PL	Simple	20	17.41	8.65
		Complex	20	16.40	5.74

Notes. TC = task complexity; PL = planning.

TABLE 7
Effects of Planning and Task Complexity on Fluency

Measures	Between-subjects Effects		
	Groups	F.	Sig.
W/M	Planning	27.817***	.000
	TC	.478	.492
	PL * TC	.102	.750
DysF	Planning	19.206***	.000
	TC	.469	.496
	PL * TC	.082	.775

Notes. TC = task complexity; PL = planning; $p^{***} < .001$.

Figure 5 shows how, compared to online planning, no planning increased fluency significantly regardless of task complexity.
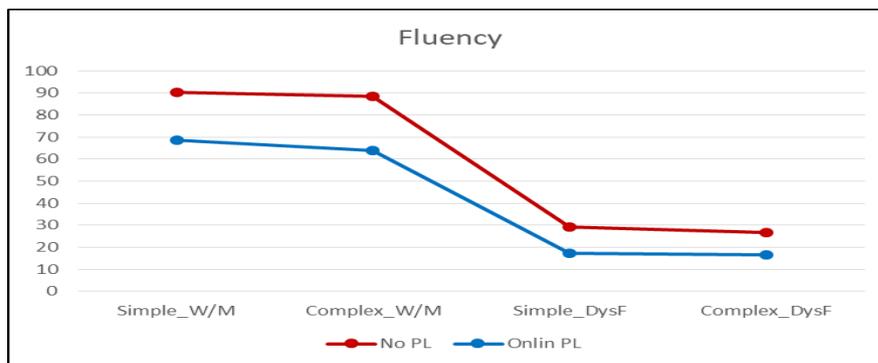

Figure 5. Significant effects of planning on fluency.

Therefore, the no planning condition improved all indices of fluency, while task complexity did not produce an increase in fluency.

Discussion and Conclusion

This study examined how planning (as task conditions) and task complexity (as task characteristics) influence syntactic complexity, accuracy, and fluency in spoken performance. The findings reveal that planning affects accuracy and fluency significantly and that, among syntactic complexity indices, task

complexity affects phrasal complexity significantly. No interaction was found between planning and task complexity, meaning that the effects of planning and task complexity were both additive on CAF in L2 spoken tasks.

Concerning planning as a task condition, no planning improved overall accuracy and verb-related specific accuracy, as well as speed fluency and repair fluency regardless of different levels of task complexity in L2 intermediate learners' spoken performance. In terms of accuracy, non-planners made a smaller percentage of errors (including verb-related errors) and generated a significantly larger percentage of error-free T-units, while online planners did not produce accurate spoken performance despite different levels of task complexity (see Figure 4). The results contradict some prior studies (Ellis & Yuan, 2005; Yuan & Ellis, 2003) but align with others (Kim, 2018; Wang, 2009, 2014). In Yuan and Ellis (2003), non-planners were given 30 seconds to view pictures and then five minutes to tell a story, while online planners were given unlimited time to tell a story, which improved syntactic complexity and accuracy compared to the no planning condition. However, in Wang (2009, 2014), the "online planning condition" consisted of simultaneously watching a 60% slower version of a video and telling a story, while the "watched online planning condition" consisted of first watching the regular speed video and then simultaneously watching a slower version of the video and telling a story. The study found that online planning did not improve any CAF while the "watched" condition facilitated online planning, which increased both complexity and accuracy. Compared to the present study, which asked participants to view pictures for one minute and then speak for either two minutes (non-planners) or four minutes (online planners), a possible reason for the different results may be the availability to compensate for the required conceptualization and formulation process. L2 learners engaging in relatively longer speaking times (for five minutes, with 60% slower video playing, or four minutes) had to manage more ideas and speak about them with limited linguistic resources; thus, they might have struggled with both conceptualization and formulation. With the support of a compensating strategy—such as unlimited speaking time (Yuan & Ellis, 2003) or pre-watching the video before speaking (Wang, 2009)—L2 learners might be less pressured to conceptualize and/or formulate. However, regarding fluency, the online planning condition in the present study may enable L2 learners to generate ideas and find appropriate linguistic expressions through a slower rate of speech. The total number of words spoken in online planning was higher (mean score = 247.2 words) than in no planning (mean score = 157.7 words), but the words per minute (wpm) were significantly fewer in online planning (mean score = 66.25 wpm) than in no planning (mean score = 89.31 wpm). Speaking slowly in the online planning condition, participants might have struggled with both conceptualization and formulation in creating diverse content and expressing it because their speech included detailed reasoning and/or descriptions of events in the pictures. However, this condition seemed not to produce any other compensating strategy; thus, it may not increase syntactic complexity and accuracy as in the above studies.

On the other hand, deprived of planning time before and while speaking in the no planning condition, non-planners seemed to focus on minimum amounts of content from the pictures, which led to a simplified story told through fewer words. However, non-planners spoke faster and repaired the broken speech through substitutive repetition and self-correction significantly more often than online planners (see Figure 5). Substitutive repetitions are often used for self-correction to achieve greater accuracy (Sasayama & Izumi, 2012). Thus, a simplified narration can alleviate the demands on conceptualization, while formulation is better supported by substitutive repetition and self-correction inducing a greater chance of monitoring speech. Mehnert (1998) claimed that when one minute of planning was provided before speaking, participants' first priority was accuracy, but complexity was only achieved by the ten-minute planners. The present study's results regarding the no planning condition's ability to improve accuracy are consistent with the one-minute planners in Mehnert's (1998) study. Thus, diverging from theoretical expectations, the no planning condition leads L2 learners to narrate a simplified story, suggesting that less pressured conceptualization may allow them to pay attention to formulation through a monitoring process with repetition and self-correction. In the present study, this condition improved accuracy and fluency compared to the online planning condition.

Regarding task complexity as task characteristics, complex tasks manipulated by resource-directing dimensions [±elements] and [±reasoning] did not improve linguistic performance (i.e., syntactic complexity and accuracy). Task complexity significantly enhanced phrasal syntactic complexity regardless of planning condition but did not increase the accuracy of L2 intermediate learners' spoken performance. This result is aligned with prior studies (Gilabert, Barón, & Levkina, 2011; Ishikawa, 2008; Michel, Kuiken, & Vedder, 2012; Sasayama & Izumi, 2012). An increase in syntactic complexity (Sasayama & Izumi, 2012) or accuracy (Gilabert, Barón, & Levkina, 2011; Michel, Kuiken, & Vedder, 2012) was acquired along with an increase in task complexity at the expense of syntactic complexity or accuracy in terms of fluency, which differed from initial CH claims. Skehan (2016) suggested that as L2 proficiency level increases, the separation between complexity and accuracy will become narrower, possibly leading to an increase in two linguistic forms simultaneously at higher proficiency levels. In my previous study (Kim, 2016), the CH was supported by advanced learners' performance, while the LACH was supported by intermediate learners' performance due to the role of L2 proficiency as a strong indicator of attentional resource capacity. Since the present study was conducted with L2 intermediate learners, their limited attentional resources may not have allowed them to pay attention to accuracy while generating and organizing their thoughts to express a complex story with greater numbers of characters and complex reasoning. Since most participants in the above studies (Gilabert, Barón, & Levkina, 2011; Ishikawa, 2008) were intermediate learners or had an unclear proficiency level (Sasayama & Izumi, 2012), the possible interaction between task complexity and specific L2 proficiency levels in spoken performance might require further investigation to clarify the effects of task complexity as proficiency level increases.

Regarding the sole increase of phrasal syntactic complexity, task complexity manipulated by [±elements] and/or [±reasoning] rarely increased syntactic complexity compared to the increase in lexical complexity. Only one study (Sasayama & Izumi, 2012) among the above studies demonstrated an improvement in syntactic complexity in the area of noun modifiers. The researchers concluded that the global syntactic complexity measure (i.e., clauses per T-unit) did not detect any positive effects of task complexity, yet the specific syntactic complexity measure (i.e., noun modifiers including relative clauses, prepositional phrases, participle phrases, compound nouns, and adjectives) showed greater use in a complex task. The present study's results are consistent with the sensitive detection of noun modifiers in Sasayama and Izumi (2012). Compared to overall, coordinate, and subordinate complexity, phrasal complexity sensitively captured the increase of syntactic complexity in complex tasks (see Table 1). Norris and Ortega (2009) claimed that "mean length of clause (MLC) is radically different from the other length-based measures… and clause length taps a more narrowly defined source of complexification" (p. 561). This is also consistent with the finding in Biber, Gray, and Poonpon (2011) that complexity develops from finite dependent clauses at earlier stages of development to non-finite dependent clauses at intermediate stages and finally to dependent phrases at the final developmental stage. Thus, to ensure a study design that meticulously captures complexity, fine-grained measure indices reflecting proficiency levels are necessary.

Overall, based on the following two findings, intermediate L2 learners' spoken performance seems to be supported by the LACH rather than the CH. First, planning as a task condition has a strong influence over linguistic performance (i.e., accuracy) and fluency. Both task conditions and characteristics affect either syntactic complexity or accuracy at the expense of either accuracy or learners' choices for an oversimplified storyline. This differs slightly from the CH assumption that removing planning time (–planning time) increases cognitive complexity but also disperses attentional resources and "does not direct learner attention and effort at conceptualization to any particular aspects of language code" (Robinson, 2015, p. 98). However, this claim may not consider learners' conscious or subconscious decisions to adjust their spoken process to achieve their goals. In this study, L2 learners' prior goal seemed to be accuracy, which was achieved by a less pressured conceptualization process and self-correction of what they stated repeatedly. Second, task complexity as a task characteristic did not improve both syntactic complexity and accuracy; it increased phrasal syntactic complexity but produced

the same level of accuracy, which is different with what the CH proposes. This might possibly change in favor of the CH because learners' capacity for conceptualization and formulation may become greater as L2 proficiency increases, as observed in studies (Ishikawa, 2006; Kim, 2016).

Further studies that combine specific L2 proficiency levels with task conditions and task characteristics would be useful to provide a clearer picture of how these two factors influence performance and interact with each other as well as with proficiency changes. Through such studies, language teachers could implement task-based language teaching based on different proficiency levels and different task conditions and characteristics.

Notes

1. Skehan (2015) claimed that he prefers the term *LACH* to the term *Trade-off Hypothesis* since the latter term focuses on unavoidable tradeoffs between performances.
2. Kim's (2016) study followed a 2x2x3 research design with three independent variables—including task complexity (simple vs. complex tasks), L2 learners' proficiency (intermediate vs. advanced level), and planning (no planning vs. pre-task planning vs. online planning)—and dependent variables (twelve measures of CAF) in written tasks. The findings were that intermediate learners' complex tasks improved accuracy, which aligns with the LACH. Meanwhile, advanced learners' complex tasks improved both complexity and accuracy, which is consistent with the CH. In addition, intermediate learners' planning helpfully increased CAF, but advanced learners' planning had limited effects on performance. Lastly, there was an interaction between task complexity and online planning that increased complexity in writing.
3. Beyond task conditions and characteristics, learner factors seem to be vague in both the LACH and CH. Robinson (2015) criticized the LACH for not addressing task features and learner factors differently as well as for having learner-dependent models that can be altered differently by different learners. On the other hand, Skehan (2014) criticized the CH for not having the power to explain how different learners perceive task complexity differently due to separate learner factors. The present study does not focus on learner factors, so this was not addressed in the study.
4. The GELT consists of speaking and writing skill tests. Students are divided into three proficiency groups: novice (Level 1), intermediate (Level 2), and advanced (Level 3). These levels are used to enroll students in a general, college-level English class.
5. The proficiency-assigning task was an oral narrative task where participants narrate a personal event for two minutes. In this task, (1) a more specific intermediate level of participants is selected, and (2) participants must warm up their timed speaking and become familiar with the narrative genre before the main task.

Acknowledgements

The Author

Nayoung Kim currently joined Korea Polytechnics as an assistant professor in March 2020 after completing her PhD in second language acquisition from Sookmyung Women's University. Her main research interests include instructional second language acquisition, task-based language teaching with

particular focuses on the roles of tasks, input, interaction, and teachers in class, classroom discourse, and curriculum development.

General English
Korea Polytechnics
54 Sandan-ro, Heungdeok-gu
Cheongju, Chungcheongbuk-do 28590, South Korea
Tel: +82 43 279 7424
Mobile: +82 10 5380 8776
Email: nkim@kopo.ac.kr

References

Biber, D., Gray, B., & Poonpon, K. (2011). Should we use characteristics of conversation to measure grammatical complexity in L2 writing development? *TESOL Quarterly, 45*(1), 5-35.

Bulte, B., & Housen, A. (2012). Defining and operationalizing L2 complexity. In A. Housen, F. Kuiken, & I. Vedder (Eds.), *Dimensions of L2 performance and proficiency: Complexity, accuracy, and fluency in SLA* (pp. 21-46). Philadelphia: John Benjamins.

Ellis, R. (2005). Planning and task-based performance. In R. Ellis (Ed.), *Planning and task performance in a second language* (pp. 3-34). Amsterdam: John Benjamins.

Ellis, R. (2009). The differential effects of three types of task planning on the fluency, complexity, and accuracy in L2 oral production. *Applied Linguistics, 30*(4), 474-509.

Ellis, R., & Yuan, F. (2005). The effects of careful within-task planning on oral and written task performance. In R. Ellis (Ed.), *Planning and task performance in a second language* (pp. 167-192). Philadelphia, PA: John Benjamins.

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

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

Gilabert, R., Barón, J., & Levkina, M. (2011). Manipulating task complexity across task types and modes. In P. Robinson (Ed.), *Second language task complexity: Researching the cognition hypothesis of language learning and performance* (pp. 105-138). Philadelphia: John Benjamins.

Heaton, J. (1975). *Beginning composition through pictures*. London, UK: Longman.

Ishikawa, T. (2006). The effect of task complexity and language proficiency on task-based language performance. *The Journal of Asia TEFL, 3*(4), 193-225.

Ishikawa, T. (2008). The effect of task demands of intentional reasoning on L2 speech performance. *The Journal of Asia TEFL, 5*(1), 29-63.

Kim, N. (2016). *The effects of task complexity and planning: Reconsidering cognition hypothesis and planning on L2 written performance* (Unpublished doctoral dissertation). Sookmyung Women's University, Seoul, South Korea.

Kim, N. (2018). The effects of online planning on CAF in L2 spoken and written performance. *English Teaching, 73*(3), 3-28.

Kim, N. (2019). L1 versus L2 use in pre-task planning of novice L2 learners' written tasks. *Korean Journal of Applied Linguistics, 35*(3), 23-49.

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

Levelt, W. (1999). Producing spoken language: A blueprint of speaker. In C. Brown & P. Hagoort (Eds.), *Neurocognition of language* (pp. 83-122). Oxford: Oxford University Press.

Mehnert, U. (1998). The effects of different lengths of time for planning on second language performance. *Studies in Second Language Acquisition, 20*(1), 83-108.

Michel, N., Kuiken, F., & Vedder, I. (2012). Task complexity and interaction: Combined effects of task-based performance in Dutch as a second language. In L. Roberts, C. Lindqvist, C. Bardel, & N. Abrahamsson (Eds.), *Eurosla Yearbook 2012* (pp. 164-190). Amsterdam: John Benjamins Publishing Company.

Norris, J. M. (2010). *Understanding instructed SLA: Constructs, contexts, and consequences.* Plenary address delivered at the annual conference of the European Second Language Association (EUROSLA), Reggio Emilia, Italy.

Norris, J., & Ortega, L. (2009). Towards an organic approach to investigating CAF in instructed SLA: The case of complexity. *Applied Linguistics, 30*(4), 555-578.

Révész, A. (2011). Task complexity, focus on L2 constructions, and individual differences: A classroom-based study. *The Modern Language Journal, 95*(Supp.), 162-181.

Révész, A. (2014). Towards a fuller assessment of cognitive models of task-based learning: Investigating task-generated cognitive demands and processes. *Applied Linguistics, 35*(1), 87-92.

Robinson, P. (2001). 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. (2005). Cognitive complexity and task sequencing: A review of studies in a componential framework for second language task design. *International Review of Applied Linguistics, 43*(1), 1-32.

Robinson, P. (2007). Criteria for classifying and sequencing pedagogic tasks. In M. Mayo (Ed.), *Investigating tasks in formal language learning* (pp. 7-26). Clevedon: Multilingual Matters.

Robinson, P. (2011). Second language task complexity, the Cognition Hypothesis, language learning, and performance. In P. Robinson (Ed*.), Second language task complexity: Researching the cognition hypothesis of language learning and performance* (pp. 3-37). Amsterdam: John Benjamins.

Robinson, P. (2015). The cognition hypothesis, second language task demands, and the SSARC model of pedagogic task sequencing. In M. Bygate (Ed.) *Domains and directions in the development of TBLT* (pp. 87-121). Amsterdam/Philadelphia: John Benjamins.

Robinson, P., Cadierno, 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.

Sasayama, S., & Izumi, S. (2012). Effects of task complexity and pre-task planning on EFL learners' oral production. In A. Shehadeh & C. Coombe (Eds.), *Task-based language teaching in foreign language contexts: Research and implementation* (pp. 23–42). Amsterdam, The Netherlands: John Benjamins.

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

Skehan, P. (2003). Task-based instruction. *Language Teaching, 36*(1), 1-14.

Skehan, P. (2009). Modelling second language performance: Integrating complexity, accuracy, fluency, and lexis. *Applied Linguistics, 30*(4), 510-532.

Skehan, P. (2014). The context for researching a processing perspective on task performance. In P. Skehan (Ed.), *Processing perspectives on task performance* (pp. 1-26). Amsterdam/Philadelphia: John Benjamins.

Skehan, P. (2015). Limited attention capacity and cognition: Two hypotheses regarding second language performance on tasks. In M. Bygate (Ed.) *Domains and directions in the development of TBLT* (pp. 123-155). Amsterdam/Philadelphia: John Benjamins.

Skehan, P. (2016). Tasks versus conditions: Two perspectives on task research and their implications for pedagogy. *Annual Review of Applied Linguistics, 36*, 34-49.

Skehan, P., Xiaoyue, B., Qian, L., & Wang, Z. (2012). The task is not enough: Processing approaches to task-based performance. *Language Teaching Research, 16*(2), 170-187.

Swender, E., Conrad, D., & Vicars, R. (2012). *ACTFL proficiency guidelines 2012* (3rd ed.). Alexandria, VA: American Council on the Teaching of Foreign Languages.

Tavakoli, P., & Skehan, P. (2005). Strategic planning, task structure, and performance testing. In R. Ellis (Ed.), *Planning and task performance in a second language* (pp. 239-273). Philadelphia, PA: John Benjamins.

Wang, Z. (2009). *Modelling speech production and performance: Evidence from five types of planning and two task structures* (Unpublished doctoral dissertation). Chinese University of Hong Kong, China.

Wang, Z. (2014). On-line time pressure manipulations: L2 speaking performance under five types of planning and repetition conditions. In P. Skehan (Ed.), *Processing perspectives on task performance* (pp. 27-61). Amsterdam/Philadelphia: John Benjamins.

Yuan, F., & Ellis, R. (2003). The effects of pre-task planning and on-line planning on fluency, complexity, and accuracy in L2 monologic oral production. *Applied Linguistics, 24*(1), 1-27.

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