ABSTRACT

Title of Dissertation:	THE INTERACTIVE EFFECTS OF TASK COMPLEXITY, TASK CONDITION, AND COGNITIVE INDIVIDUAL DIFFERENCES ON L2 WRITING
	Jiyong Lee, Doctor of Philosophy, 2018
Dissertation directed by:	Dr. Michael H. Long, Second Language Acquisition

Second language (L2) researchers, L2 teachers, and textbook designers have shown great interest in the relationship between task characteristics and interlanguage development. Although the literature is inundated with research on the effects of task complexity on speech, less attention has been paid to its effects on writing. To this end, the present study investigated how increasing task complexity led to changes in cognitive load, and in turn, changes in L2 written performance. It also explored whether limiting the number of acceptable solutions to a task, i.e., task closure, had an effect on writing. Finally, the roles of working memory capacity (WMC) and aptitudes for implicit and explicit learning in task performance were investigated as well.

Eighty-three Korean learners of English and seven L2 teachers deemed as experts were recruited for the study. The L2 learners were randomly assigned to one of two conditions: the Open condition, in which participants carried out open task

versions, and the Closed condition, in which they carried out closed versions. Participants carried out two tasks, each with a simple and complex version. Learner self-ratings, expert judgments, and time-on-task were used to obtain independent evidence that increasing task complexity led to changes in cognitive load. An Ospan task and the LLAMA D and F were used to measure WMC and aptitudes for implicit and explicit learning, respectively.

A series of mixed effects models revealed that tasks intended to be more complex were perceived as such by both learners and experts. While significant task complexity effects were found on lexical diversity and one measure of accuracy, its effects on syntactic complexity measures were not significant. Task closure effects were only found for lexical diversity, such that open versions elicited more diverse vocabulary than complex versions. Cognitive individual differences also played a role, such that higher WMC was related to greater lexical diversity, and higher aptitude for implicit learning was associated with greater accuracy in writing.

THE INTERACTIVE EFFECTS OF TASK COMPLEXITY, TASK CONDITON, AND COGNITIVE INDIVIDUAL DIFFERENCES ON L2 WRITING

by

Jiyong Lee

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2018

Advisory Committee:
Professor Michael Long, Chair
Professor Robert DeKeyser
Professor Steven Ross
Dr. Jared Linck
Professor Jeff MacSwan

Acknowledgements

My Ph.D. journey has definitely had its ups and downs, but one thing is for sure: I will always cherish the memories of my time in Maryland. I am truly blessed to have met such wonderful people during these years, and I would like to express my gratitude to each and every one of them.

First, I would like to thank my advisor, Dr. Mike Long, for his guidance, insightful suggestions, and dry sense of humor that is a gift you need to be born with. I have learned so much from him in many ways, more than I ever expected when I first came to UMD. I would also like to extend my gratitude to my committee members. I still remember the first day I met Dr. Robert DeKeyser when I was an MA student in Korea, and I would like to thank him for his intellectual inspiration throughout the years. I would also like to thank Dr. Steve Ross and Dr. Jared Linck for their valuable comments and advice on statistical analyses, and Dr. Jeff MacSwan for his feedback and support.

I would like to give a huge shout-out to my colleagues and friends in the SLA program, especially Hyojin Jeong, Yoonjee Hong, and Fatima Montero. I am very fortunate to have made such good friends, and I truly thank you for your friendship. I would also like to thank the people who came to my dissertation defense to show their support. I am also very grateful to my ARHU family, who have been there to witness the birth of my child and the birth of my dissertation. I never would have been able to go to Korea for data collection if it weren't for their generosity.

My deepest gratitude goes to my family. I would like to thank my parents and parents-in-law for their unwavering support and encouragement. I am especially indebted to my father, who helped me with participant recruitment for my dissertation and never hesitates to give me advice whenever I am stuck in a rut. Special thanks to my mother and mother-in-law, who took turns coming to the States to take care of Riley while Jae and I are at work. I would also like to express my sincere appreciation to my father-in-law, brother-in-law, sister-in-law, and my brother for their patience and belief in my abilities. And last but certainly not least, to Jae and Riley. I am the luckiest woman in the world to have Jae as a husband and Riley as my daughter. I really could not have survived this journey without you. Thank you for being my rock. I love you.

Table of Contents

List of Tables

List of Figures

Chapter 1. Introduction

Ever since Long's (1985) proposal of task as a meaningful and viable unit of analysis in identifying learners' needs, defining syllabus content, organizing language acquisition opportunities, and measuring student achievement, tasks have drawn the attention of teachers, textbook writers, and language researchers. As the basis of Task-Based Language Teaching (TBLT), tasks have been subject to intense scrutiny for their facilitative role in second language (L2) development. As the goal of task use is for learners to develop their L2 while performing tasks successfully, research has focused on the criteria with which tasks could be classified and sequenced. Two major frameworks—Robinson's (2001a, 2003, 2005b, 2011) Cognition Hypothesis and Skehan's (1996, 1998) Trade-Off Hypothesis—provide different approaches as to how tasks should be sequenced in order to obtain desired changes in linguistic complexity, accuracy, and fluency (CAF).

In recent years, task-based research witnessed an increase in the use of various methods to see whether task complexity manipulations along the Here-and-Now dimension, reasoning demands, and number of elements actually lead to the expected changes in cognitive load. More attention has also been paid to how certain individual differences interact with task complexity effects, mostly focusing on working memory capacity. However, few studies have examined how language aptitude, apart from working memory, plays a role in task performance. Furthermore, there has been very little research on whether closing a task, i.e., restricting the number of acceptable solutions, has a differential effect on the task's cognitive load and on task performance. To this end, this dissertation attempts to examine the interactive effects of task complexity, task closure, and individual differences in working memory capacity and aptitudes for implicit and explicit learning on cognitive load, and in turn, on L2 writing.

Chapter 2. Review of the Literature

2.1 Models of Task Complexity

2.1.1 Robinson's Cognition Hypothesis

Consistent with the proposals by Long (1985, 2015) and Long and Crookes (1992) that pedagogic tasks should be designed and sequenced according to increased task complexity, so that they eventually resemble the full demands of real-world tasks, Robinson (2001a, 2003, 2005b, 2011) proposed the Triadic Componential Framework (TCF; see Table 1) and the Cognition Hypothesis (CH). This framework is the most detailed attempt to date to distinguish the componential dimensions of task complexity and predict their effects on language performance. It specifies three superordinate categories: Task complexity (cognitive factors), Task condition (interactive factors), and Task difficulty (learner factors).

Task *complexity* refers to the intrinsic cognitive demands of a task, which are relatively fixed and inherent qualities, independent of learner characteristics. They are expected to account for within-learner variance in task performance. Determined by participation and participant factors, task *condition* concerns the situational settings and conditions under which tasks are performed. Task *difficulty* is associated with the learner's perceptions of task demands, which contribute to between-learner variation in task performance. These three components do not operate in a mutually exclusive manner, and complex interactions among the dimensions and factors are expected to lead to multiple effects on L2 learning and performance (Robinson & Gilabert, 2007).

In fact, Robinson (2001b) states that there may be a stable relationship between task complexity and task difficulty.

Table 1. The Triadic Componential Framework for task classification (from Robinson 2007)

1. *Task Complexity* (Cognitive factors)	2. *Task Condition* (Interactive factors)	3. *Task Difficulty* (Learner factors)
(a) *Resource-directing variables* making cognitive/conceptual demands	(a) *Participant variables* making interactional demands	(a) *Ability variables* and task-relevant resource differentials
+/- here and now +/- few elements -/+ spatial reasoning -/+ causal reasoning -/+ intentional reasoning -/+ perspective-taking	+/- open solution +/- one-way flow +/- convergent solution +/- few participants +/- few contributions needed +/- negotiation not needed	h/l working memory h/l reasoning h/l task-switching h/l aptitude h/l field independence 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 +/- single task +/- task structure +/- few steps +/- independency of steps +/- prior knowledge	+/- same proficiency +/- same gender +/- familiar +/- shared content knowledge +/- equal status and role +/- shared cultural knowledge	h/l openness to experience h/l control of emotion h/l task motivation h/l processing anxiety h/l willingness to communicate h/l self-efficacy

Task complexity has two components—*resource-directing* variables and *resource-dispersing/depleting* variables—that can be manipulated for the purposes of syllabus design and task sequencing. These dimensions vary in the way they influence the allocation of attentional resources during task performance. Increasing task complexity along resource-directing dimensions is claimed to place greater conceptual/functional demands on the learner. This, in turn, is expected potentially to channel learner's attentional and memory resources to certain features of the linguistic system necessary for completing the task. For instance, deictic expressions (e.g., *this, that, here, there*) can be used to express temporality of reference, and logical subordinators (e.g., *so, because, therefore*) to justify one's beliefs or to give reasons. In a task where such concepts are required for task performance, the learner's attention will be drawn to the ways that the L1 and L2 differ in grammaticizing concepts. This allocation of attentional resources will result in the facilitation of interlanguage development, i.e., new form-function and conceptual mapping in the L2. Because accuracy and complexity are both driven by the nature of the functional linguistic demands of a task itself, there will be a correlation between the two. However, fluency is argued to contrast with accuracy and complexity. Consequently, greater accuracy and linguistic complexity, but less fluency of the output is anticipated. On the other hand, increasing complexity along resource-dispersing dimensions places greater demands on the ability to *access* the currently established and developing L2 knowledge repertoire. Practice is assumed to facilitate this accessibility, thus having a positive effect on accuracy, linguistic complexity, *and* fluency.

Motivated by the multidimensional structure of the TCF, Robinson's CH claims that increases in task complexity should be the logical basis for task sequencing and syllabus design, as opposed to task difficulty, task condition, or any linguistic text feature. When tasks require certain concepts to be expressed and understood, their conceptualizations are demanding of cognitive resources. As these cognitive demands of the task are increased, learners are primed to direct their attentional and memory resources to aspects of the L2 system needed to understand the concepts. Accordingly, selective attention is facilitated and 'noticing the gap' occurs, thereby stimulating conceptual L2 grammaticization. To this effect, three major predictions are made when task complexity is increased along resource-directing dimensions: (1) learners will be pushed to greater accuracy and linguistic complexity but less fluency, (2) interaction and negotiation of meaning will be promoted, resulting in heightened attention to, and incorporation of, task input and modification of output, and (3) individual differences in ability and affective variables contributing to perceptions of task difficulty will differentiate task performance and language learning.

Based on the predictions of the CH, Robinson (2010) and Baralt, Gilabert, and Robinson (2014) propose two instructional-design principles for sequencing tasks. The first principle states that "only the cognitive demands of tasks contributing to their intrinsic conceptual and cognitive processing complexity are sequenced" (2010, p. 247). The second principle concerns the steps involved to sequence tasks, such that one should "increase resource-dispersing dimensions of complexity first (e.g., from + to – planning time), and then increase resource-directing dimensions (e.g., from – to + intentional reasoning)" (2010, p. 247). This will ensure that the access and proceduralization of current interlanguage resources take place under optimal conditions (Robinson, 2003, 2007; see Figure 1). To be more specific, resource-directing dimensions are kept simple while complexity along resource-dispersing dimensions is increased to target-task levels in the first stage. Practice along the resource-dispersing dimensions will facilitate access to the learner’s L2 knowledge base, promoting automatic access to and control over the interlanguage system and resulting in improved task performance. In the second stage, resource-directing dimensions are increased to target-task levels.

		Stage 1		Stage 2
Low Performative and Low Developmental Complexity	→	High Performative and Low Developmental Complexity	→	High Performative and High Developmental Complexity
(a) Resource-directing + few elements + no reasoning + Here-and-Now		(a) Resource-directing + few elements + no reasoning + Here-and-Now		(a) Resource-directing - few elements - no reasoning - Here-and-Now
(b) Resource-dispersing + planning + prior knowledge + single task		(b) Resource-dispersing - planning - prior knowledge - single task		(b) Resource-dispersing - planning - prior knowledge - single task

Figure 1. Sequencing tasks along dimensions of complexity (Adapted from Robinson, 2003)

The two stages for task sequencing constitute a model, the SSARC model, that Robinson (2010) proposes for increasing L2 pedagogic task complexity, represented in the following (Baralt, Gilabert, & Robinson, 2014):

Step 1. SS (stabilize, simplify) = i × e [(‘s’rdisp) + (‘s’rdir)]^{n}

Step 2. A (automatize) = i × e [(‘c’rdisp) + (‘s’rdir)]^{n}

Step 3. RC (restructure, complexify) = i × e [(‘c’rdisp) + (‘c’rdir)]^{n}, where

i = current interlanguage state,

e = mental effort,

‘s’ = simple task demands,

‘c’ = complex task demands,

rdisp = resource dispersing dimensions of tasks,

rdir = resource directing dimensions of tasks,

^{n} = potential number of practice opportunities on tasks

2.1.2 Skehan’s Trade-Off Hypothesis

Skehan (1996, 1998) and Skehan and Foster (1997) propose an alternative model to task sequencing, the Trade-Off Hypothesis (TOH) or the Limited Attentional Capacity Hypothesis (LACH). The biggest difference between the TOH and the CH lies in their basic assumption: while the CH assumes multiple, non-competing pools of attentional resources that learners may draw on, the TOH assumes a single pool of resources accessible to learners. Because this capacity is limited, mapping of form-meaning relationships is restricted. This reflects VanPatten’s (1990) view that meaning is given priority when meaning and form are competing for attention—form can be attended to only if the recovery of meaning is necessary, or when there is cognitive capacity to spare. According to Skehan (2014), L2 learners face problems when performing demanding tasks due to these processing limitations, leading to a trade-off effect, such that increased fluency may occur with either greater accuracy or complexity (at best), but not with both simultaneously (Skehan & Foster, 1997).

Another major difference between the two major frameworks is that the CH takes a deductive approach and the TOH takes an inductive approach to research (Skehan, 2016). In other words, the former takes a “top-down” approach, such that the central concept of task complexity can be narrowed down to resource-directing and resource-dispersing variables, each of which lead to specific predictions. On the other hand, Skehan’s approach claims to be more inductive in nature, as it is generally based on “case-by-case” observations, identification of task features from general theory, previous research, or classroom experience, and then empirical studies on whether certain task features have a systematic relationship with performance. To this effect, Skehan (2016) makes a distinction between a focus on tasks and task design, and a focus on task conditions that concern choices made about how tasks are implemented in task-based research.

Task conditions that are claimed to affect performance include familiarity of information, interactivity of tasks (monologic vs. dialogic), degree of structure (e.g., clear and sequential macrostructure), complex outcomes (straightforward vs. multifaceted), and transformation of information. For instance, Skehan (2003, 2014) suggests that familiar tasks, those based on concrete information, structured tasks with a clear time line or macro-structure, and those containing a post-task phase may lead to greater accuracy and fluency. Interactive tasks, those requiring transformation or manipulation of materials, or those containing a pre-task planning phase may result in greater linguistic complexity. In fact, Skehan (2016) argues that conditions such as pre-task planning, task repetition, and post-task activities have a greater and more consistent effect on L2 performance, compared with variables in Robinson's CH, such as ± Here-and-now and ± few elements.

Table 2. Task sequencing features (Adapted from Foster & Skehan, 1996; Skehan, 2001)

Code complexity	Cognitive complexity	Communicative stress
Syntactic complexity	(a) Familiarity of information	Time pressure
Lexical complexity	Familiarity of material in task	Modality
Redundancy & variety	Familiarity of task-type & discourse genre	Scale
		Stakes
	(b) On-line processing	
	Reasoning operations required	
	Nature of input used in task	
	Degree of organization of input material	

Claiming that limited attentional resources cause trade-off effects, the main goal of the TOH is for learners to develop balanced language proficiency in all areas of accuracy, complexity, and fluency through effective task choice and the effective use of task conditions. Furthermore, tasks should be sequenced so that certain task conditions direct attention selectively to achieve the desired outcome (Skehan, 1998). To this end, Skehan (1996) proposes a number of principled criteria by which tasks can be analyzed, compared, and sequenced (see Table 2). *Code complexity* refers to formal language factors concerning traditional areas of syntactic complexity and lexical diversity required for a task. *Cognitive complexity* concerns the content of production and is related to the conceptualization stage of Levelt's (1989) model of first language speaking. A further distinction can be made: familiarity and processing. Pre-existing knowledge that is held in memory is directly accessed and drawn upon when a learner is familiar with certain aspects of a task. On the other hand, processing concerns active intellectual engagement with the task—the extent to which they must actively think through task content, because understanding the material or access to existing knowledge and using it untransformed is not sufficient for arriving at a solution. Lastly, *communicative stress* concerns factors that affect the pressure of communication. Time pressure involves how quickly a task needs to be done, and whether there is any urgency in the manner in which it has to be done. Modality involves the contrast between receptive modalities (listening/reading) and productive modalities (speaking/writing). Scale refers to factors associated with task-based approaches to teaching, such as number of participants and number of relationships involved. Stakes concerns how important it is to do the task, and control to the degree of influence participants have on a task and on how it is done.

Bearing these features in mind, Foster and Skehan (1996) propose a framework (see Table 3) that identifies three stages in task implementation. In each stage, specific pedagogic goals and associated techniques are suggested for the development of language performance in terms of complexity and accuracy. The purpose of pre-task activities is to teach, mobilize, or make language relevant to task performance salient. This can be attempted in two ways. One would be to set up the relevant language for a task in the form of explicit or implicit pre-teaching, or give learners a pre-task to do and then provide them with the language they need. Another approach would be to ease the cognitive load during task performance, resulting in more attention allocated to the actual language used, which in turn could lead to greater linguistic complexity or accuracy. Reducing cognitive processing load can be achieved in a range of activities. Pre-task activation sessions, where learners are prompted to recall pre-existing schematic knowledge, can have an effect on the cognitive familiarity of the task. They can also observe similar tasks or be given related pre-tasks to do so that they have activated schemas when performing the real task. More effectively, they can engage in pre-task planning of the language they need to use or the meaning they want to express, resulting in more accurate, complex, and fluent language production. During the mid-task stage, the appropriate level of task difficulty should be chosen, so that learners do not devote excessive attention to conveying any sort of meaning when a task is too difficult, and they do not get bored and fail to meet all task requirements when a task is too easy. Finally, if learners are aware that there is a subsequent post-task activity, a focus on form can filter through the prior (main) task and reduce the likelihood of learners allocating attention exclusively to meaning. Without such awareness, communication goals may be so dominant that lexicalized communication strategies would take priority, and learners would pay less attention to accuracy.

Table 3. A framework for task implementation (Foster and Skehan, 1996)

Stage	Goal	Typical Technique
Pre-task		
Linguistic	Introduce new forms to interlanguage repertoire	Explicit and implicit teaching Consciousness-raising
Cognitive	Reduce cognitive load	Plan linguistically & cognitively Observe similar tasks
	Push learners to express	Plan

	more complex ideas	Observe
Mid-task		
Task choice	Balance difficulty of task	Use analytic scheme
Task calibration	Increase or reduce difficulty	Introduce surprise
		Provide (visual) support
Post-task		
	Raise consciousness for a focus on form	Use public performance and post-task activities

Recognizing that findings motivated by theory are more ideal than inductively generated findings, Skehan (2016) posits that Levelt's (1989) model of first language (L1) speech production can be used as a more useful framework to account for task condition effects. In Levelt's model, there are three major phases to speech production: 1) conceptualization, in which a preverbal message is generated and monitored, 2) formulation, in which the lemmas relevant to the preverbal message are first accessed and grammatical and phonological encodings subsequently occur, and 3) articulation, in which phonetic and articulatory plans are produced for overt speech. Skehan attempts to connect task condition effects to this model of speech production in terms of pre-task planning, online planning, post-task activities, interactivity, and existence or lack thereof support materials. As illustrated in the schematic representation in Figure 2, pre-task planning is connected to conceptualizer operations, as it concerns planning of ideas to be expressed. Greater linguistic complexity and greater fluency is expected to take place when there are pre-task planning activities. Online planning has a connection with formulator operations, such that time pressure felt during speech selectively and positively affects accuracy. Post-task activities, interactivity, and provision of support materials are claimed to have a connection with both conceptualizer and formulator operations. Post-task activities seem to have a beneficial effect on accuracy, due to formulator operations and monitoring, where selective attention is paid to avoid making errors. In the case of interactivity, the conceptualizer operation is involved when learners interact to bring ideas together for task completion. The formulator operation becomes involved when time pressure is eased and there is opportunity to regroup while the interlocutor is speaking. When support materials are provided, memory demands are reduced, resulting in freed up attentional resources for the conceptualizer and formulizer operations. A drawback to this would be that they can also constrain the scope and extent to which learners can take different approaches to task performance because the conceptualizer may place extra load on the formulator. On the other hand, the absence of support materials may place greater memory demands on learners, as they are given more flexibility as to how to carry out the task. The conceptualizer is less constrained, and eases its demands on the formulator. Finally, repetition is connected with the formulator operation, in that traces of the lemma partially retrieved during the first performance persist and facilitate the second performance, resulting in more effective syntax building and message creation. Based on the connection between L2 speech performance and Levelt's model, Skehan (2016) makes two predictions: 1) separation between performance measures should diminish as the L2 mental lexicon grows with greater proficiency, such that complexity and accuracy should correlate more at higher proficiency levels, and 2) the conceptualizer-derived variables (e.g., information organization) and formulator-derived variables (e.g., structure) can be used to induce interaction or synergetic effects.

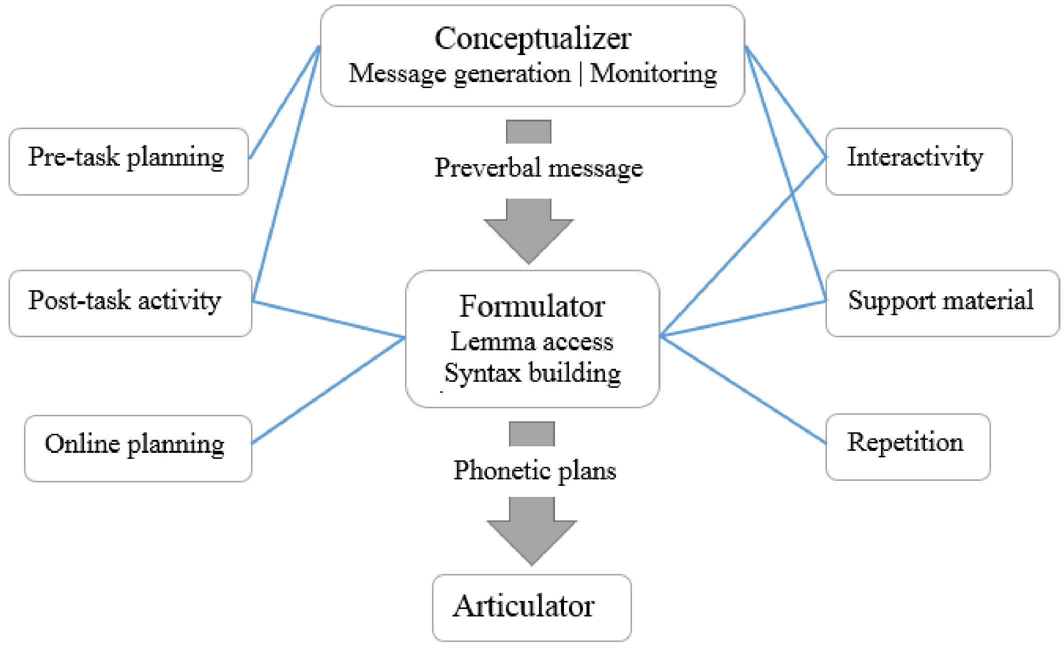


Figure 2. Task-based application of Levelt's (1989) model of L1 speech production

2.2 Task Complexity Effects on L2 Performance

2.2.1 Empirical Studies of Task Complexity

SLA research has witnessed a proliferation of empirical studies testing either the CH or the TOH. Most studies testing the CH manipulate resource-directing variables, such as ± Here-and-Now, ± few elements, and ± reasoning demands. On the other hand, ± planning, ± familiarity, ± post-task activities, and ± task repetition have been the focus of studies testing the TOH. Although they claim to provide evidence supporting one model or the other, many have in fact obtained mixed findings.

Foster and Skehan (1996) investigated the effects of task complexity and ± planning on language production. Thirty-two learners of English carried out three types of interactive oral tasks. Based on the required level of learner attention, familiarity, and predictability of information, a personal information-exchange task was hypothesized to be the least complex, a narrative task mid-complex, and a decision-making task the most complex. Participants were divided into three groups that performed tasks in differing conditions: no planning, undetailed planning, and detailed planning. It was found that planned conditions led to greater fluency and greater syntactic complexity. However, the undetailed planning condition resulted in the greatest accuracy. Task-type was found to interact significantly with planning condition, such that planning effects on accuracy and linguistic complexity were greater in the more complex narrative and decision-making tasks than in the simple personal-information-exchange task. Because planning had a positive effect on fluency and complexity, but not on accuracy, when considering task-type, a trade-off effect was shown, and the authors concluded that their findings supported the TOH.

Similar results were found in Skehan and Foster's (1997) study. Forty learners of English performed the same task-types as in Foster & Skehan (1996). The effects of ± planning and ± knowledge of a post-task activity were investigated. It was found that students under planned conditions, compared with unplanned conditions, generally showed greater fluency, accuracy, and complexity in their oral output. When planning effects were examined across task-types, stronger effects on accuracy were found in the mid-complex narrative task, but not in the complex decision-making task. On the other hand, planning effects on complexity were found in the decision-making task, but not in the narrative task. Knowledge of a post-task activity was found to have a minimal effect on accuracy.

Robinson (1995) found limited support for the CH in a small-scale study. Twelve learners of English with intermediate L2 English proficiency performed oral narrative tasks with increasing complexity along the ± Here-and-Now dimension. Supporting evidence for the CH was found, in that learners used more varied lexical items in the complex [- Here-and-Now] condition. They also tended to use more syntactically complex structures in the complex version. However, no significant differences were found for syntactic complexity and fluency. Robinson attributed the lack of significant findings to the small sample size, questionable reliability and validity of the outcome measures, relatively low proficiency of the learners, and the openness of monologic tasks.

Other studies that have examined the effects of task complexity along the ± Here-and-Now dimension include Gilabert (2007) and Ishikawa (2007). With 46 learners of lower-intermediate English proficiency, Gilabert investigated how the ± Here-and-Now dimension interacted with ± planning time. Participants performed an oral narrative task with comic strips. Statistical analyses revealed that while there was a positive task complexity effect on accuracy and fluency in planned and unplanned conditions, linguistic complexity remained the same in both conditions. Increasing task complexity was even found to reduce lexical complexity, contrary to the predictions of the CH. The planned condition was also found to be more beneficial for fluency and lexical complexity.

The findings of Ishikawa (2007) are very similar. In Ishikawa's study, 54 Japanese learners of English performed a written narrative task after seeing a cartoon strip. Task complexity was determined by the use of present/past tense and presence/absence of contextual support. Findings were claimed to support the CH regarding accuracy, fluency, and structural complexity, in particular. Task complexity effects were not found for lexical diversity, although results seemed to point in that direction on two measures. Ishikawa concluded that the lack of a trade-off effect was counterevidence to the TOH's assumption of a single memory source.

Analyzing data from the same participants, Kuiken and Vedder (2007, 2008) investigated the effects of ± few elements and L2 proficiency on written production. Based on their scores on an English cloze test, 84 Dutch learners of Italian and 75 Dutch learners of French were assigned to either a low- or high-proficiency group. They carried out a letter-writing task in which they had to write to a friend about choosing a holiday destination. The number of requirements involved determined the level of complexity.

Kuiken and Vedder (2008) found that both groups were significantly more accurate on the complex task, particularly with respect to fewer first and second degree errors (i.e., minor to slightly more serious deviations in spelling, meaning, grammatical form, or word order) on the complex task. For the learners of Italian, there were no significant differences in syntactic complexity and lexical variation between the simple and complex task. Learners of French showed similar results concerning syntactic complexity. However, lexical variation, measured by type-token ratio, was significantly higher on the complex task. No interaction was found between language proficiency and task complexity for both groups. Because the authors found that the complex task generated greater syntactic complexity and accuracy in writing, they concluded that their findings provided partial support for the CH, and no support for the TOH.

Using more specific measures, Kuiken and Vedder (2007) examined learners' written output more closely. Accuracy was measured by errors in grammar, lexicon, orthography, and appropriateness, and a distinction was made between high- and low-frequency words. They obtained results showing that high-proficiency learners of Italian outperformed low-proficiency learners in terms of grammatical, orthographical, and other errors. Furthermore, the complex task was found to result in fewer lexical errors and more high-frequency words. For the learners of French, high-proficiency learners showed greater accuracy in terms of grammatical, lexical, and other errors. They also made fewer orthographical, appropriateness, and other errors in the complex task. Unlike the students of Italian, more low-frequency words were used in the complex task.

Table 4 provides a summary of the studies mentioned above. Operationalizations of task complexity have varied across studies, and different kinds of outcome measures have been employed. Findings have been mixed, failing to provide unambiguous support for either model.

Table 4. Studies of task complexity effects

Author	**Participants**	**Task**	**Modality & Interactivity**	**Task complexity factor**	**Results**
Foster & Skehan (1996)	32 learners of English	Information-exchange, narrative, decision-making	Speech Dialogic	± planning	+ planning: Fluency ↑ Syntactic complexity ↑ - planning: Accuracy ↑

					Task-type x planning condition
Skehan & Foster (1997)	40 learners of English	Information-exchange, narrative, decision-making	Speech Dialogic	± planning, ± knowledge of post-task activity	Accuracy ↑ Complexity ↑ Fluency ↑ Task-type x planning condition
Robinson (1995)	L2 learners of English	Narrative	Speech Monologic	± Here-and-Now	Accuracy ↑ Syntactic complexity ≈ Lexical diversity ↑ Fluency ≈
Gilabert (2007)	46 learners of English	Narrative	Speech Monologic	± Here-and-Now, ± planning	Accuracy ↑ Syntactic complexity ≈ Lexical diversity ↓ Fluency ↑
Ishikawa (2007)	54 Japanese learners of English	Narrative	Writing Monologic	Present/past tense, ± contextual support	Accuracy ↑ Syntactic complexity ↑ Lexical diversity ≈ Fluency ↑
Kuiken & Vedder (2008)	84 Dutch learners of Italian, 75 Dutch learners of French	Letter-writing task	Writing Monologic	± few elements	Accuracy ↑ Syntactic complexity ≈ Lexical diversity ↑

2.2.2 Limitations of Task Complexity Research

Although the two models of task complexity have generated considerable interest among educators and language researchers, they are not without limitations. Lee (2018) claims that there are three major problematic areas, any of which may contribute to the lack of unambiguous support for either model: 1) a lack of consistent operationalization of complexity dimensions, 2) a lack of consistency in the choice and operationalization of outcome measures, and 3) a failure to include native speaker baseline data.

Under the TOH, it has been claimed that such task characteristics as familiarity of task material, reasoning operations required, and degree of structure may have an effect on performance. However, there is no clear explanation as to how each feature could be operationalized, or how they may interact with one another. Likewise, the CH does not provide clear guidelines for operationalizing task complexity variables, and does not suggest how task condition and task difficulty variables may interact with one another. In other words, tasks are holistic and involve a combination of factors (R. Ellis, 2017). Skehan (2016) even states that the distinction between resource-directing and resource-dispersing variables is unclear. For instance, Kuiken and Vedder (2007) and Skehan (2016) point out that there is no clear-cut distinction between ± few elements and ± reasoning (spatial, causal, intentional), as the presence of one almost automatically implies the the other. Moreover, the hypothesis does not specify the numbers of elements that constitute as [+ few] or [- few elements]. Due to such unspecified operationalizations, the topics of previous research have been refined to task complexity manipulations along ± Here-and-Now, ± few elements, and ± intentional reasoning. However, D. Ellis (2011) and Long (2015) point out that even these oft-manipulated dimensions are difficult to operationalize uncontroversially. It is not clear how to count the number of elements manipulated in a task, and whether each element are equally salient to the learner, as distinct from what the designer or researcher may have intended. If, say, five of ten elements are unnoticed, is the task considered simple or complex?

Regarding performance measures, D. Ellis (2011) criticized early work (e.g., Foster & Skehan, 1996 and Robinson, 1995) for their operationalizations of complexity, accuracy, and fluency. For instance, as measure of fluency, Foster and Skehan (1996) included the number of reformulations, replacements, false starts, and repetitions, and Robinson (1995) counted the frequency of two-second pauses. According to Jackson and Suethanapornkul's (2013) report, a total of 84 different CAF measures were employed in merely 17 studies. Norris and Ortega (2009) found that 13 measures were used to measure complexity, alone, in 16 studies. In fact, some studies, such as Ishikawa (2007), use several different measures to assess complexity (e.g., mean length of T-unit, S-nodes per T-unit, clauses per T-unit, dependent clauses per clause, and S-nodes per clause). Having a variety of measures for one dependent variable inevitably leads to multicollinearity problems, and may also contribute to the lack of consistent findings.

The third problem involves a lack of studies that include baseline native speaker data. To be fair, Skehan's and Robinson's approaches concern task performance and language development in the second language. However, due to unclear operationalizations and explanations of possible interactional processes, it would be beneficial, or even necessary, to test predictions with L1 speakers. While a range of individual differences, such as L1 background, L2 proficiency and age of onset vary drastically among L2 learners, native speakers are a more homogenous, comparison group and are assumed have complete control over the L1 (Long, 2015). By testing native speakers, as well as non-native speakers, any changes in performance can be attributed to task complexity manipulations alone, unfiltered through incomplete L2 competence (see D. Ellis, 2011; and Long, 2015). Similarly, Michel (2011) claims that native speaker data are needed in order to "fully understand the measures we use for the evaluation of non-native performances" (p. 149). Foster and Tavakoli (2009) also support this view, stating that including such data allows for a clear distinction between performance features due to L2 processing and those due to task performance.

Other problematic areas in task-based research involve implementation variables and participant interpretation. R. Ellis (2017) claims that the complexity of a task can never be considered separately from how it is implemented, which is not considered by many studies. Examples of implementation variables include planning time, pre-task activities, and time-limit. Even if the same task is used, the cognitive load of the task may be different depending on how it is implemented. However, it should be noted that task complexity is an inherent quality of a task, which should be differentiated from task difficulty. While implementation variables may make a task more or less difficult, it will not make it more or less complex. Another general problem in this field concerns the possibility of differences in interpretation between participants and task designer (Skehan, 2016). A task designer will most likely have a goal in mind when choosing/designing a task. However, participants' reactions to the task may be different from what was intended or expected. They may not notice the greater complexity of the "complex" task, or they may intentionally avoid the added element(s), opting for the path of least resistance. The task designer may believe that L2 learners will be familiar with the task material or task-type, but familiarity depends on personal experience, which may vary among those who complete the task. Such problems may also contribute to unexpected or mixed results.

2.3 Validation of Task Complexity Manipulations

2.3.1 Measures of Cognitive Load

Numerous studies have attempted to support or refute the CH or the TOH, but with mixed or null findings. In order to account for this inconsistency, Norris (2010), Révész (2014), and R. Ellis (2017) argue that an important step has traditionally been assumed, not empirically tested. Consequently, a growing number of researchers are first investigating whether task complexity manipulations actually result in the intended changes in *cognitive load*, and whether these changes, in turn, have an effect on the accuracy, complexity and fluency of learner production. Cognitive load can be defined as the burden placed on a learner's capacity for cognitive processing, or the processing capacity of working memory. When a learner performs a complex task as opposed to a simple one, it is assumed that the task will place a greater burden on working memory. It is insufficient to observe output changes and assume that they are the result of increased task complexity; task complexity needs to be established first, and independently. Various methods have been suggested to address this issue: 1) subjective self-ratings, 2) subjective time estimations, 3) dual task methodology, 4) time-on-task, 5) psychophysiological techniques (e.g., eye-tracking), and 6) stimulated recall protocols. The following provides details regarding the first four measures.

Subjective self-rating of perceived difficulty is one of the measures employed in earlier studies of the CH. Robinson (2001b) found that when 44 Japanese learners of English performed an oral interactive task, increased task complexity in terms of the amount of information and availability of prior knowledge had a significant effect on their ratings of overall difficulty and stress. However, ratings of interest and motivation were found to be unrelated to task complexity manipulations. Using the same questionnaire with 60 learners of English, Gilabert, Barón, and Llanes (2009) found that complex tasks led to high ratings of perceived difficulty and stress, but low confidence ratings. However, there were no significant differences in participants' interest and motivation ratings for the simple and complex versions of the task. In Ishikawa's (2011) study, 46 Japanese learners of English performed oral tasks whose complexity was increased along the ± intentional reasoning dimension. In general, complex tasks were found to be perceived as more difficult. This suggests that self-ratings of difficulty, stress, confidence, etc., are a valid measure of cognitive load during task performance.

The rationale for self-estimations of duration is that time seems to pass quicker when a person is performing a difficult or attention-demanding task, as opposed to one that is easy or less attention-demanding. There are two paradigms in the estimation of a target duration: the prospective paradigm (*experienced duration*) and retrospective duration (*remembered duration*) (Block & Zakay, 2008). In the former, a person is aware while performing a task that s/he must estimate its duration. In the latter, s/he is aware of making an estimation only after the time period has ended. Attentional models in psychology claim that time estimations are determined by the amount of attention allocated to the processing of temporal information. In the prospective paradigm, attention is assumed to be shared by a non-temporal information processor and a temporal information processor (Block, 1992), with the former focusing on external stimuli regardless of time, and the latter focusing on time-related attributes of external stimuli. However, previous studies have provided evidence that the two processors require some of the same attentional resources. As a result, fewer attentional resources may be allocated to temporal information when the load of non-temporal processing is increased (Block, 1992; 2003). People are expected to be less accurate in estimating the duration of a task, the greater the cognitive load placed on them, and a negative linear relationship between prospective judgment length and load of non-temporal information processing is assumed. In other words, more complex (and/or more difficult) tasks may reduce the available cognitive resources allocated to temporal processing, resulting in shorter duration judgments (see Block, 1992; Block, Hancock, & Zakay, 2010; and Zakay, 1992). For this method, the *ratio* of subjective duration estimation to objective duration should be used to make comparisons across task-types and within the same task-type.

Using time judgments and perceptions of task difficulty, several studies have attempted to validate task complexity manipulations (Malicka & Levkina, 2012; Baralt, 2013; Révész, Michel, & Gilabert, 2015; Sasayama, 2016; and Lee, 2018). Malicka and Levkina (2012) divided 37 EFL learners into high- and low-proficiency groups. Each performed two oral instruction-giving tasks. Complexity was manipulated along the dimensions of ± few elements and ± spatial reasoning. Affective variables were measured in terms of task difficulty, learner confidence, interest, and motivation. Learners were also asked to identify the task that took longer to perform, and estimate the time it took to complete the tasks. The researchers found that learners tended to perceive the complex task as more difficult, irrespective of proficiency level. On the other hand, compared with the high-proficiency group, more participants in the low-proficiency group considered the complex task to take longer to perform. In addition, 60 percent of the high-proficiency group and 65 percent of the low-proficiency group accurately estimated the time it took to complete the tasks.

Baralt (2013) investigated how task complexity mediated the effects of recasts in two different modes, face-to-face (FTF) and computer-mediated communication (CMC). Task complexity was increased along a resource-directing dimension, ± intentional reasoning. After performing two interactive dialogic story-retell tasks at two levels of complexity, 84 learners of Spanish estimated the time it took them to carry out the task and completed a questionnaire that asked about how hard the task was, whether they felt it was challenging, and what the overall task difficulty was. Contrary to Baralt's assumptions, learners perceived FTF tasks to be more difficult than CMC tasks. However, complex tasks resulted in greater time estimations than the actual time required, while modality did not have a significant effect.

According to Brünken, Plass, and Leutner (2003), dual-task performance and brain activity measures are objective direct methods of measuring cognitive load. The former assumes that simultaneous performance of two tasks will affect distribution of attentional resources. The underlying principle is that performance on the secondary task reflects the level of cognitive load generated by the primary task (Révész, 2014). Processing of a secondary task will vary when different versions of a primary task require varying degrees of attentional resources, resulting from differences in cognitive load induced by the primary task. Choice reaction tasks are mostly employed as the secondary task, in which learners make simple decisions. For example, participants may be required to indicate the correctness of simple math equations or respond to screen/letter color changes. The dependent variables for dual task methods are reaction time and error rate on the secondary task.

Measured by the time from task onset to task completion, time-on-task has been introduced to task-based research in very recent years. If learners complete a task successfully, time-on-task reflects the time taken for them to familiarize themselves with the task, process the input and materials to solve the task, come up with a solution, and provide their response in either spoken or written form. It is mainly used in the fields of education and educational psychology, in either one of two approaches: 1) as an indicator of a (latent) construct, e.g., reasoning speed (Goldhammer & Klein Entink, 2011), or 2) a predictor to account for differences in task success (Goldhammer, Naumann, Stelter, Toth, Tóth, Rölke, & Klieme, 2014). In task-based research, Vasylets (2016) is the only empirical study thus far that used time-on-task as a dependent variable to investigate the differential effects of task complexity in different modes. It was used as a measure of fluency, because speech is faster than writing by default, and using a more traditional measure, such as words per minute, would be pointless. Participants were found to spend significantly more time on a complex writing task. Based on the results of Goldhammer et al.'s (2014) large-scale study showing that time-on-task was moderated by item difficulty and individual ability, where item difficulty was determined by the degree to which a task required cognitive effort, it is reasonable to believe that time-on-task could be used as a valid measure of cognitive load. In other words, time spent on a task could be indirect evidence of the cognitive effort needed to carry out a task, and it was assumed that time-on-task would increase with increases in task complexity.

2.3.2 Validation Studies of Task Complexity

Four studies of particular interest attempted to validate task complexity manipulations using a combination of various methods (Révész, Sachs, & Hama, 2014; Révész, Michel, & Gilabert, 2015; Sasayama, 2016; and Lee, 2018). Using expert judgments, dual task methodology, and eye-tracking, Révész et al. (2014) measured the cognitive load of oral narrative tasks whose complexity was increased along the ± causal reasoning dimension. Two experts in applied linguistics judged all of the 32 experimental items, and versions intended to be greater in complexity were found to be rated as such. The dual task method employed involved responses to color changes during task performance. Sixteen learners of English and 16 native English speakers had to respond to these changes as quickly and accurately as possible. Although reaction times on the secondary task did not significantly differ between simple and complex task versions, accuracy rates were found to be a sufficiently sensitive measure of cognitive load. It was found that ESL learners obtained higher accuracy rates when performing the simple versions, and native speakers obtained higher accuracy rates than the ESL learners. Eye-tracking also provided support for the validity of task complexity manipulations in terms of fixation counts and fixation duration. Compared with native speakers, ESL learners also showed greater fixation durations, but not higher counts.

Révész et al. (2015) attempted to validate task complexity using the dual task method, participant self-ratings, and expert judgments. Comparisons were made of the performance of 48 native English speakers and 48 ESL speakers. Three oral task-types—a picture narrative, a map task, and a decision-making task—were employed, each in a simple and complex version. Adopting the dual task method in Révész et al. (2014), the secondary task required participants to respond to screen color changes. They also completed a self-rating questionnaire regarding the mental effort required by the task and overall perceived task difficulty. Sixty-one ESL teachers also provided their expert judgments of the tasks. Results showed that the dual task method was a good measure of cognitive load, in that participants' accuracy on the secondary task was higher on the simple version than the complex version. However, task complexity effects were not found for reaction time. Both ESL learners' and teachers' self-rated perceptions of mental effort and perceived task difficulty provided further support for the validity of task complexity manipulations, as the ratings were found to be higher for complex versions of tasks. In short, it can be concluded that complex task versions placed a greater cognitive load on participants than simple versions.

Sasayama (2016) employed the dual task method, time estimations, and self-ratings of task difficulty and mental effort. Fifty-three adult Japanese learners of English, divided into three groups by L2 proficiency, performed four narrative tasks that differed in the number of elements involved. Reaction times on the secondary task were slower while performing more complex tasks than simpler tasks, although differences were not statistically significant. Results of the duration estimations also indicated that the most complex and simplest tasks placed the intended degree of cognitive load on the learners. However, duration estimations of the two tasks in the middle range of complexity ran counter to predictions. With regard to self-ratings, learners also judged the most complex task as significantly more difficult and requiring more mental effort than the other three. An interaction effect was also found between L2 proficiency, task complexity, and measure of cognitive load. In order to account for the unexpected results of the tasks in the mid-complexity range, Sasayama suggested two possibilities: storyline and picture quality, and code complexity.

In an effort to elucidate task complexity effects unfiltered by non-native competence, Lee (2018) examined the cognitive load and speech of 42 native English speakers during task performance. Participants performed three types of tasks—a direction-giving map task, a seating arrangement task, and a car accident description task—whose complexity was progressively increased at three levels each. Task complexity was manipulated in terms of the number of elements involved. Three methods were used to measure the cognitive load of the three different versions of each task: 1) self-ratings, 2) prospective duration estimation, and 3) dual task methodology. Increasing task complexity was found to lead to greater cognitive load, as proven by self-ratings of perceived difficulty, mental effort, stress, and interest, shorter prospective duration estimations, and slower reaction times on the secondary task of the dual-task methodology. Regarding linguistic outcomes, the mid-complex task versions elicited the most syntactically complex structures, while the most complex versions generated the greatest lexical diversity. In order to account for the reason why the most complex task versions did not generate the greatest syntactic complexity, it was claimed that the added complexity in those versions was either ignored or went unnoticed by some participants, resulting in simplification of the task by the participant.

Table 5. Studies validating task complexity manipulations

Author	Participants	Task	Task complexity factor	Validation method	Results
Robinson (2001b)	44 Japanese learners of English	Direction-giving task	± Here-and-Now	Self-ratings of difficulty	Difficulty ↑ Stress ↑
Gilabert, Baron, & Llanes (2009)	60 learners of English	Narrative reconstruction, map task, decision-making task	± Here-and-Now, ± few elements, ± intentional reasoning	Self-ratings of difficulty	Difficulty ↑ Stress ↑ Confidence ↓
Malicka & Levkina (2012)	37 EFL learners	Instruction-giving task	± few elements, ± spatial reasoning	Self-ratings of difficulty, time estimations	Difficulty ↑ Generally accurate time estimations
Baralt (2013)	84 learners of Spanish	Story-retelling task	± intentional reasoning	Self-ratings of difficulty, time estimations	Difficulty: F2F > CMC Time estimations ↑
Révész, Sachs, & Hama (2014)	16 learners of English, 16 native English speakers	Reasoning about causes of events	± causal reasoning	Dual task method, eye-tracking, expert judgments	Dual method: Accuracy ↓ Eye-tracking: Fixation duration: ↓ Difficulty ↓
Révész, Michel, & Gilabert (2015)	48 ESL learners, 48 native English speakers, 61 ESL teachers	Picture narrative, map task, decision-making	± familiarity, ± reasoning demands	Dual task method, self-ratings of difficulty, expert judgments	Dual method: Accuracy ↓ Mental effort ↓ Difficulty ↓
Sasayama (2016)	53 Japanese learners of English	Story-telling task	± few elements	Self-ratings of difficulty, dual task method, time estimations	Difficulty ↑ Mental effort ↑ Mixed time estimations by task-type Dual task: reaction time ≈
Lee (2018)	42 native English speakers	Map task, seating arrangement task, car accident description task	± few elements	Self-ratings of difficulty, dual task method, time estimations	Difficulty ↓ Mental effort ↓ Stress ↓ Dual method: Reaction time ↑ Time estimations ↓

2.4 Task Closure Effects on L2 Performance

A major line of task-based research investigates effects of task condition—how certain choices can be made as to how tasks are implemented, such as whether there is pre-task planning, whether there is a post-task activity, or whether the task is monologic or dialogic. A manipulation of task condition that has received little attention is the distinction between open and closed tasks (or ± open solution under Robinson's framework). According to Long (1989), an open task does not have a predetermined solution, but has a wide range of acceptable solutions. It can vary in its degree of 'openness,' in that a task that stipulates the topic-information is less open than one that allows learners to choose topics to discuss freely. A debate, ranking favorite hobbies, or choosing the ten most popular athletes, are examples of open tasks. A closed task, on the other hand, requires that participants attempt to reach a single solution or a one of a small, finite set of correct solutions. An example of a closed task is a seating arrangement task in which the participant is required to arrange the best seating plan for a number of guests with seating preferences that must be met (Lee, 2018). Long emphasizes that in order for this task closure feature to work, it is crucial that learners are aware that the tasks they are to carry out are open or closed.

Although neither the CH nor the TOH make any predictions regarding this feature, Long (1990) and Loschky and Bley-Vroman (1993) provide a rationale for using closed tasks over open tasks. The more 'open' an open task is, the greater the tendency is for learners to treat topics briefly, to drop them altogether when the task is deemed too difficult or when there is a major communication breakdown, to provide and incorporate feedback less because there is no need to, and to recycle linguistic material less. On the other hand, because closed tasks are designed to force learners to find a single solution, which they know exists, they are less likely to give up when trouble arises. There will be greater quality and quantity of negotiation for meaning, more topic and language recycling, and more provision and incorporation of feedback, all of which are conducive to interlanguage development and L2 acquisition. In short, closed tasks are claimed to be superior to open tasks. Nunan (1991) cautions that language proficiency may play a moderating role, such that closed tasks may be more beneficial than open tasks to lower-intermediate to upper-intermediate learners, whereas they may be too difficult for beginners. However, it could be argued that closed tasks such as picture identification, picture matching, and Spot the difference can easily be accomplished by beginner learners. Taking an opposite view, in favor of open tasks, Turner and Paris (1995) claim that open tasks are better because they allow learners to make personal choices, provide challenges, allow learners to take control over their learning, foster interaction through collaboration, foster constructive comprehension, and promote feelings of competence and efficacy. In fact, they argue that when an open task is moderately challenging, learners adapt their strategies rather than give up. They will even take personal responsibility for them, because they will see the task as controllable. However, closed tasks are more likely invoke a sense of frustration or discouragement because they feel they failed in working out the solution.

Only a handful of studies have directly investigated the effects of task closure. In order to examine whether specific contexts and task closure have an effect on learners' responses to math problems, Sullivan, Warren, and White (2000) found that task closure effects differed depending on the task that learners performed. Three tasks were employed (Differing Units, Perimeter and Area, and Embedded Rectangles), each with two versions: one consisting of parallel closed and open contextual tasks, and the other consisting of parallel closed and open no-context tasks. Results showed that students received lower scores on the closed version of the Differing Units task, but received higher scores on the closed version of the two other tasks. In order to account for the mixed results, the researchers analyzed the responses by breaking down task elements, and suggested that the open versions that were more difficult than the closed tasks involved linking two concepts and using the links to conjecture and generalize. On the other hand, the closed version that was more difficult involved the need to convert from one unit to another, e.g., meters to centimeters.

In a study of task closure in language learning, Rahimpour (2009) used narrative tasks to investigate whether closed versions would elicit greater complexity, accuracy, and fluency than open versions. Task closure and the ± Here-and-Now variable were manipulated. It was found that closed tasks were significantly better than open tasks at eliciting greater fluency, operationalized as the number of words between pauses. Although the difference was not statistically significant, closed tasks were better at eliciting more error-free t-units than open tasks. Rahimpour concluded that the closed task condition was more demanding and motivating to the learner, lending support for the idea that closed tasks are superior. However, a major flaw of this study was in how task closure was operationalized. Instead of manipulating tasks so that the one intended to be closed would have one solution or a small, finite set of solutions, the existence of an interlocutor who was required to select and order relevant pictures based on participants' descriptions was the factor that determined whether a task was closed or not.

In a research synthesis of 15 studies, Cobb (2010) attempted to investigate the effectiveness of task-based interaction in the classroom. One of the moderator variables examined included task closure. Among the 15 studies included in the analysis, the number of closed tasks was twice that of open tasks (nine versus four), and Cobb reported that superior effect sizes were found for closed tasks. However, this finding seems problematic in two ways. First, Cobb mentioned that very few authors explicitly indicated whether they used open or closed tasks, so two coders had to rely on their judgments based on information in the original reports. Therefore, the coding of open/closed tasks was potentially subject to error. Second, the effect sizes reported are not an indication of the differential effects of open vs. closed tasks. What can be claimed, based on the results, is that treatments used with closed tasks in some studies had a greater effect size than those used with open tasks in different studies. This was not the result of direct comparisons between open vs. closed tasks within the same studies.

The only empirical study that has compared the effects of task closure in the context of task complexity was conducted by Montero (2018). 62 beginner learners of Spanish were divided into two groups: one that performed a closed version of a task and the other an open version of the same task. Participants were required to rearrange and describe a number of colored geometric shapes in front of them. In the closed condition, the arrangement of the shapes was predetermined and shapes were either a triangle, a square, a star or a circle. In the open condition, participants could rearrange the shapes however they wanted, and the shapes were of an odd form. The complexity of the tasks were also manipulated in terms of number of shapes. Increases in task complexity generally led to greater accuracy, syntactic complexity, and lexical diversity, lending support for the CH. However, counter to the predictions of the study, open task versions were either comparable to or more effective than the closed versions. Montero concluded that due to the task design of the study, the open versions allowed participants to be creative, resulting in complex descriptions.

Due to the dearth of research on task closure, there is little empirical evidence to support the claim that closed tasks are more effective than open tasks. However, reflecting post hoc interpretations of previous studies on L2 interaction, Long (1989) and Loschky and Bley-Vroman (1993) propose that closed tasks are better at promoting negotiation work, resulting in greater precision and more recycling of the language, because restrictions on the number of acceptable solutions force learners to persevere until they find the solution. Based on this rationale, the present study predicted that closed tasks would elicit more complex and accurate language, as shown in Table 6, and examined the effects of task closure on L2 written performance separately and in combination with task complexity.

Table 6. Predicted effects of open and closed tasks

	Open task		Closed task	
Measure	Simple	Complex	Simple	Complex
Complexity	-	+	-	++
Accuracy	-	+	-	++

2.5 Working Memory in L2 Studies

One line of investigation that is receiving increasing attention in task-based research concerns individual differences among learners and how they are relevant for task performance. Because the present study aimed to see if tasks intended to be more complex actually led to an increase in cognitive load, it also aimed to find out whether a cognitive factor in L2 learning, working memory capacity (WMC), could account for differences, if any, in the desired changes in task performance. Working memory is "a limited capacity system, which temporarily maintains and stores information, supports human thought processes by providing an interface between perception, long-term memory, and action" (Baddeley, 2003, p. 829). According to Conway, Kane, Bunting, Hambrick, Wilheim, and Engle (2005), it is a multi-componential system that actively maintains information in the face of ongoing processing and/or distraction, which is the result of domain-specific storage and rehearsal processes and domain-general executive attention. Baddeley and Logie (1992) claim that WM allows for comprehension and mental representation of the immediate environment, preservation of information about immediate past experience, acquisition of new knowledge, and formulation and achievement of current goals.

The most well-known model of WM was proposed by Baddeley and Hitch (1974), with an additional element (the episodic buffer) included in a later version (Baddeley, 2000). Emphasizing its role as a system for complex cognitive tasks, such as language comprehension, arithmetic, syllogistic reasoning, and complex dynamic tasks, this multiple-componential model of WM consists of four parts: the central executive, phonological loop, visuospatial sketchpad, and episodic buffer. The central executive is a supervisory system that controls the flow of information of the 'slave' subsystems and other cognitive processes. The subsystems—the phonological loop, visuo-spatial sketchpad, and episodic buffer—take on passive roles as short-term information repositories and are controlled by the central executive. These short-term memory (STM) storage spaces are determined by the quantity of information maintained and how long that information is available. Working memory capacity (WMC) is determined by storage and processing components, with each component measurable separately or in combination. For instance, simple span tasks such as the forward digit span task, word span task, and non-word span task, primarily tap the ability to store and rehearse information. Complex span tasks, conversely, tap the ability to store information while faced with additional processing demands (Linck, Osthus, Koeth, and Bunting, 2014; Juffs & Harrington, 2011). Certain methods can also be used to measure the components of WM separately with, for instance: 1) a listening span task, counting span task, backward digit span task, and day/night Stroop test for executive processes, 2) the digit span task, recall of words test, and non-word repetition test for phonological short-term memory (PSTM, i.e., verbal WM), and 3) pattern recall and Corsi blocks for visuospatial WM (Gathercole, 1999).

Focusing on writing, Kellogg (1996, 1999) attempted to account for how working memory functions in writing tasks by incorporating Baddeley's working memory model. This proposal attempts to explain how specific writing processes, borrowing the Hayes and Flower (1980) model of text formulation, rely on the main components of WM. Three writing processes—formulating, executing, and monitoring—each involve two-level basic processes. The formulation process involves the planning of ideas and translating them into linguistic expressions later to be handwritten, typed, or dictated. The execution process consists of motor unit programming and the execution of muscle movements. The monitoring process consists of reading an already produced text and editing the mental and textual representation output, and this is claimed to oversee the formulation and execution systems. The order of these processes are not fixed, and the model supports simultaneous activation of the processes when processing demands on working memory do not exceed capacity limitations.

With regard to working memory, it is assumed that the central executive is related to the planning process when writers generate ideas, organize schemas, or use supporting visual graphics or orthography. Along with the reading and editing processes, planning entails a high level of reflective thought and self-regulation. Translating an idea into a sentence involves a grammatical component and then a phonological component, after which the phonological representations of the sentence constituents are temporarily processed and stored in verbal working memory. In addition to the verbal component, translating may also place demands on the central executive, especially when a writer is struggling for words and sentence structures. However, its demands would be small during conversational speech, when sentence generation is largely automatic. Programming demands on the central executive may be small, if not negligible, when motor skills are highly practiced (such as would be the case for a skilled typewriter). During the editing process, error detection is expected to consume substantial capacity of the central executive, as errors can occur at various levels of a text structure, and detecting an error would involve high levels of metacognitive awareness. Editing can occur before, or any time after, the execution process, but if it is sufficiently delayed after sentence formulation, the writer must engage in a reading process before editing. The reading process itself is claimed to be very complex and is expected to place great demands on both the central executive and verbal components of working memory. Table 7 provides a summary of the types of working memory components used by writing processes.

Table 7. WM components used by writing processes (Adapted from Kellogg, 1999)

		Working memory component		
	Basic Process	Spatial	Central executive	Verbal
Formulation	Planning	✓	✓	
	Translating		✓	✓
Execution	Programming		✓	
	Executing			
Monitor	Reading		✓	✓
	Editing		✓	

Seeking to address the reason why some learners have difficulty in acquiring an L2, SLA researchers have been interested in how IDs in WMC might account for variation in L2 learning and use. The relationship between WMC and attentional demands of L2 tasks has been of particular interest, and most studies have investigated the role WM plays in the noticing of feedback and interaction-driven L2 learning. To date, there is a paucity of work investigating how certain components of WM moderate the relationship between task complexity and L2 production.

Two studies have directly examined the moderating effect of WMC on task performance (Kormos & Trebits, 2011; Kim, Payant, & Pearson, 2015). In Kormos & Trebits' (2011) study, 44 Hungarian learners of English performed two narrative tasks differing in complexity, which was operationalized by the absence/existence of a storyline. A backward digit span task was used to measure WMC. Their findings do not provide support for the CH, as no effects were found for task complexity in terms of accuracy, global syntactic complexity, and fluency. WMC was found to play a limited role—WMC effects were found for clause length, but not for subordination.

Employing a pretest-posttest-delayed posttest design, Kim et al. (2015) examined the relationship between task complexity, WM, and English question development. Eighty-one learners of English performed three types of oral interactive tasks in dyads with a native English speaker. Reasoning demands were manipulated so that each task-type had two versions of varying complexity. Participants were divided into two groups, one that carried out the simple versions of the tasks and the other that performed the complex versions. During task performance, they received recasts from their interlocutor following various linguistic errors. Their WM was measured by an aural running span task. It was found that only WM was significantly related to the amount of noticing, and it was also the only significant predictor of question development. The researchers concluded that with complex tasks in particular, higher WM is more beneficial for interaction-driven learning.

The present study attempted to contribute to the literature on the relationship between task complexity effects and WMC. It was predicted that those with higher WMC would have more attentional resources to meet the cognitive/conceptual demands enhanced by increases in task complexity. In terms of Kellogg's (1996) model, greater WMC would have a beneficial impact on the formulation and monitoring processes, such that writers would be able to plan ideas better, generate more accurate and complex sentences, and detect and modify their errors better. Therefore, WMC was expected to have a positive effect on written performance, especially when writers need to exert more mental effort while carrying out complex tasks.

2.6 Aptitudes for Language Learning

As part of the effort to investigate the nature of variation in L2 learning and use, a focus of interest in SLA research in recent years has been cognitive individual differences (IDs). Foreign language aptitude is one such factor considered to be an important determinant of success in language learning. L2 aptitude is characterized as individual cognitive strengths during L2 learning and performance in various contexts and at various stages (Robinson, 2005a). Defined as "the variable or variables that determine the amount of time a student needs to learn a given task, unit of instruction, or curriculum to an acceptable criterion of mastery under optimal conditions of instruction and student motivation" (Carroll, 1989, p. 26), four components of aptitude were proposed: 1) phonetic coding ability, 2) grammatical sensitivity, 3) memory abilities, and 4) inductive language learning ability (Carroll, 1990). Table 8 describes these components in detail. This traditional Carrollian concept of aptitude was developed to differentiate between successful and unsuccessful learners in instructed learning settings. Aptitude batteries that have stemmed from this line of research include the Modern Language Aptitude Test (MLAT) (Carroll and Sapon, 1959), Pimsleur's Language Aptitude Battery (PLAB) (Pimsleur, 1966), the Defense Language Aptitude Battery (DLAB) (Petersen and Al-Haik, 1976), and the LLAMA Language Aptitude Tests (Meara, 2005).

Table 8. Carroll's four-factor aptitude model (from Dörnyei & Skehan, 2003)

Majors	Definitions
Phonemic coding ability	Capacity to code unfamiliar sound so that it can be retained
Grammatical sensitivity	Capacity to identify the functions that words fulfill in sentences
Inductive language learning ability	Capacity to extrapolate from a given corpus to create new sentences
Associative memory	Capacity to form associative links in memory

A longstanding issue with the traditional concept of aptitude is that it lacks a clear definition. Because the construct was developed on the basis of selecting successful learners, Dörnyei (2005) points out that language aptitude is defined by what language aptitude tests measure. In response to such criticisms, Robinson (2001c, 2005a) proposed a theoretically-motivated model of an interaction of aptitudes, also known as 'complexes.' In this model, aptitudes interact with each other, such that the cognitive resources and abilities have a combined facilitative effect on language processing and learning (also see DeKeyser & Koeth, 2011). As a consequence, certain instructional treatments or tasks—such as implicit negative feedback and other Focus on Form techniques—are assumed to be more beneficial for certain L2 learners with a particular aptitude complex. According to this model, aptitude is a complex construct with multiple cognitive characteristics. Primary abilities include pattern recognition, speed of processing in phonological WM, and grammatical sensitivity. The notion that WM may be a key component of language aptitude has been supported by several other researchers, as well (Miyake & Friedman, 1998; DeKeyser & Koeth, 2011; Kormos, 2013, Skehan, 2016).

Contrary to these views, findings that WM and language aptitude are two distinct constructs have also been reported. In a validation study of the LLAMA aptitude test, Grañena (2013) performed a series of exploratory principal component analyses on LLAMA subtests and tests measuring WM, STM, processing speed, and attention control. It was found that language aptitude was different from processing speed and WM, which were measured by an operation span task and a visual letter span task, respectively. Attempting to clarify the relationship between WM, aptitude, and L2 learning, Yalçın, Çeçen, and Erçeti (2016) found that WMC, assessed by a reading span task and an operation span task, correlated only with the LLAMA F scores that tapped grammatical inferencing abilities. This result suggests that the ability to store and manipulate verbal information in WM is crucial in memorizing individual items and processing relationships among them for grammatical inferencing. However, researchers state that this is also counterevidence to the argument that WM is an indicator of aptitude, and suggest a diminished role of WM as extra component of aptitude. Using the operation span task and a grammatical inferencing subtest of the LLAMA (LLAMA F), Yilmaz (2013) compared the role of WM and language aptitude on the effects of explicit correction and recasts. Forty-eight native speakers of English performed an information-gap task after learning 59 Turkish words. Depending on the group they were assigned to, either recasts or explicit corrections of the plural and locative case morphemes were provided during task performance. It was found that WMC and language aptitude moderated the effect of feedback type on plural and locative scores, such that those with high WMC and language aptitude benefited more from explicit correction than from recasts. However, those with low WMC and language aptitude showed no differences in the effectiveness of the two feedback types. Comparing high and low WMC and language aptitude learners within each feedback type, an advantage for high WMC was found for the explicit correction group only, and no differences were found for the recast group.

Based on the results of her validation study, Grañena (2013) proposed that aptitude has two dimensions that differ in their relevance for language learning: explicit language aptitude (ELA) and implicit language aptitude (ILA). ELA is relevant for explicit language learning and processing through reasoning and deliberate hypothesis testing, and ILA is relevant for acquiring patterns in input without awareness of the rules (implicit induction). Aptitude was measured via the LLAMA test battery, whose subtests measure the ability to learn unknown vocabulary (LLAMA B), recognize unfamiliar sounds (LLAMA D), associate sounds to corresponding orthographic characters (LLAMA E), and learn the grammar of an unfamiliar language (LLAMA F). In her validation study, Grañena (2013) found that

LLAMA D loaded separately from the other subtests, and LLAMA B, E, and F loaded on a common factor. A closer look into the nature of the subtests accounts for these results: LLAMA D does not include a study phase, nor allows time to rehearse, minimizing problem-solving and strategy use and involving online processing. On the other hand, LLAMA B, D, and F include a study phase, allow time to rehearse materials, and provide opportunities for problem-solving and strategy use, thus involving explicit associative learning.

Drawing on Grañena's findings, Yilmaz and Grañena (2016) investigated the relationship between ELA and three types of feedback: explicit feedback in the form of immediate rejection followed by correction, implicit feedback in the form of recasts, and no feedback. Employing an oral pretest, immediate posttest, delayed posttest design, the researchers elicited the English indefinite article from 48 L2 learners of English. A significant interaction between ELA and task condition was found, indicating that explicit feedback is more beneficial to learners with high aptitude for explicit learning (i.e., high ELA). In an attempt to account for the non-significant effects for recasts, the researchers postulated that compared with more explicit types of negative feedback, recasts lack the directness of rejecting errors and presenting the correct forms, which is expected to enable learners to focus their attention on language forms.

To date, no study has examined the extent to which ILA and ELA are related to task complexity effects on L2 performance. Would these broader types of aptitude that capture explicit and implicit cognitive abilities predict the quality and quantity of L2 written performance when task complexity is increased? Do closed tasks work better for learners with high ILA or ELA? These were questions that the present study sought to answer. Because writing involves high levels of conscious effort and cognitive processing, it could be predicted that ELA would be more predictive of task performance than ILA. However, due to the lack of empirical findings, this prediction was tentative at best. In fact, ILA could have a greater association with task performance because writers may rely more on implicit use of L2 knowledge (i.e., without awareness) when they are focusing on the content of their writing while carrying out a complex task.

Chapter 3. The Current Study

In light of the previous research discussed in Chapter 2, the present study sought to address the following research questions. Directional hypotheses were formulated according to the predictions of the CH because a resource-directing variable was manipulated to increase task complexity: ± number of elements. Although not supported by empirical findings, the hypothesis for RQ 5 was driven by rationales from Long (1989) and Loschky and Bley-Vroman (1993).

RQ1. Does increasing the number of task elements lead to systematic increases in cognitive load for L2 learners?

H1-1. Increases in number of elements will lead to higher self-ratings of perceived difficulty.

H1-2. Increases in number of elements will lead to higher self-ratings of mental effort.

H1-3. Increases in number of elements will lead to higher self-ratings of stress.

H1-4. Increases in number of elements will lead to higher self-ratings of interest.

H1-5. Increases in number of elements will lead to higher time pressure felt during the planning stage.

H1-6. Increases in number of elements will lead to higher time pressure felt during the writing stage.

H1-7. Increases in number of elements will lead to lower expectations on task performance.

H1-8. Increases in number of elements will lead to longer time-on-task.

Learner self-ratings, expert judgments, and time-on-task were used to measure cognitive load. The learner self-rating survey included questions regarding the overall perceived difficulty of the task, mental effort required for task performance, stress felt during task performance, the level of interest felt, time pressure felt during the planning and writing stages, and expectations on task performance. All ratings except those for performance expectations were expected to be higher for complex tasks. As an objective measure of cognitive load, the time spent on tasks were also used: time-on-planning, time-on-writing, and time-on-whole task. These times were also expected to increase with increased task complexity.

RQ2. Does increasing the number of task elements lead to systematic changes in L2 written performance?

H2-1. Increases in task complexity will lead to greater syntactic complexity.

H2-2. Increases in task complexity will lead to greater lexical diversity.

H2-3. Increases in task complexity will lead to greater accuracy.

Because task complexity was manipulated along one of Robinson's CH resource-directing variables, ± number of elements, hypotheses were formulated based on the predictions of the CH. Therefore, increases in task complexity were expected to elicit greater syntactic complexity, lexical diversity, and accuracy in L2 writing.

RQ3. Does task closure bring changes in cognitive load for L2 learners?

H3-1. Compared with open tasks, closed tasks will show higher self-ratings of perceived difficulty.

H3-2. Compared with open tasks, closed tasks will show higher self-ratings of mental effort.

H3-3. Compared with open tasks, closed tasks will show higher self-ratings of stress.

H3-4. Compared with open tasks, closed tasks will show higher self-ratings of interest.

H3-5. Compared with open tasks, closed tasks will show higher time pressure felt during the planning stage.

H1-6. Compared with open tasks, closed tasks will show higher time pressure felt during the writing stage.

H1-7. Compared with open tasks, closed tasks will show lower expectations on task performance.

H1-8. Compared with open tasks, closed tasks will show longer time-on-task.

To date, there is a paucity of studies that investigated the differential effects of task closure on L2 development and task performance. However, several researchers have claimed that closed tasks are more beneficial than open tasks because the restriction on the number of acceptable solutions forces learners to find the correct solution. Because learners are expected to pay more attention when carrying out closed tasks and exert more effort on finding the correct solution, it was hypothesized that closed task versions would place a greater cognitive load on participants than open versions.

RQ4. Does task closure have a differential effect on L2 written performance?

H4-1. Compared with open tasks, closed tasks will lead to greater syntactic complexity.

H4-2. Compared with open tasks, closed tasks will lead to greater lexical diversity.

H4-3. Compared with open tasks, closed tasks will lead to greater accuracy.

For the same reason behind the direction of the hypotheses regarding the cognitive load differences between open and closed tasks, closed tasks were expected to elicit greater syntactic complexity, lexical diversity, and accuracy.

RQ5. Is working memory capacity (WMC) predictive of changes in L2 written performance due to increases in task complexity?

H5-1. Participants with high WMC will show greater syntactic complexity in their writing.

H5-2. Participants with high WMC will show greater lexical diversity in their writing.

H5-3. Participants with high WMC will show greater linguistic accuracy in their writing.

WMC is characterized as the capacity for temporary storage and processing of information, and Kellogg's model claims that writing processes rely on WM components. According to this model, people rely on the central executive function when formulating ideas and translating them into linguistic expressions. In this regard, those with high WMC were expected to retain more ideas and expressions in memory, which would later be reflected in their writing performance in terms of syntactic complexity, lexical diversity, and accuracy.

RQ6. Is implicit language aptitude (ILA) predictive of changes in L2 written performance due to increases in task complexity?

H6-1. Participants with high ILA will show greater syntactic complexity in their writing.

H6-2. Participants with high ILA will show greater lexical diversity in their writing.

H6-3. Participants with high ILA will show greater accuracy in their writing.

RQ7. Is explicit learning aptitude (ELA) predictive of changes in L2 written performance due to increases in task complexity?

H7-1. Participants with high ELA will show greater syntactic complexity in their writing.

H7-2. Participants with high ELA will show greater lexical diversity in their writing.

H7-3. Participants with high ELA will show greater accuracy in their writing.

Due to its relative novelty in SLA research, few empirical studies have investigated the relationship between aptitudes for language learning—ILA and ELA—and task performance. When compared with speech, writing involves high levels of conscious effort and cognitive processing relevant for the planning and organization of ideas, writing, and revising. Hence, it was predicted that those with high ELA, i.e., those who have more controlled use of language and are better at monitoring their L2 and writing, will show greater linguistic complexity and precision in their writing. However, because participants may focus more on the goal of the tasks (finding a solution) and/or content, especially when carrying out complex tasks, it may also be the case that ILA would have a positive relationship with task performance, as such tasks would require automatic use of L2 knowledge. With these possibilities taken into consideration, hypotheses for RQs 5 and 6 were formulated predicting positive relationships between aptitudes for language learning and task performance.

Chapter 4. Methodology

4.1 Participants

Eighty-three Korean learners of English as a foreign language (EFL) were recruited from Inha University in South Korea (36 males, 47 females). Their ages ranged from 18 to 27 years old (M = 23.13, SD = 1.88). They were undergraduate students, the majority of whom were majoring in English Education (see Table 9 for a complete list of majors). All were living in Korea at the time of the study, and with the exception of one student who had lived in the United States for three years for family reasons, none had lived in an English-speaking country for more than one year. They were randomly assigned to one of two conditions: the Open condition in which they performed open tasks, and the Closed condition in which they performed closed tasks. Participants met with the researcher for one session of an average of 90 minutes. They were compensated with 20,000 won (equivalent to approximately 20 dollars) for their participation at the end of the session.

Table 9. List of majors

Majors	Number of participants (N = 83)
English Language Education	73
Education	2
Business Administration	2
Computer Engineering	1
Economics	1
English Language and Literature	1
Nursing	1
Ocean Sciences	1
Statistics	1

Three Korean university professors in a field related to English Education or English Literature, and four Korean teachers teaching English at the high school level, also participated in the study to provide expert judgments on the tasks. Except for one professor who had eight to nine years of teaching experience, everyone else had taught at their institutions for over 10 years. All of them had high ratings as reflected in their answers to a question about their familiarity with Task-Based Language Teaching (TBLT) and task complexity on a 9-point Likert scale, indicating good familiarity: an average rating of 7.33 for the professors and an average of 6.5 for the high school teachers. The experts received the same tasks the undergraduate participants carried out and were asked to fill out a questionnaire regarding the cognitive load of the tasks.

4.2 Materials

4.2.1 Cloze Test

An English cloze test developed by Brown (1980) was used to measure participants' English proficiency. This test is based on a 399-word passage, *Man and his Progress,* and is adapted from a book designed for intermediate ESL readers, *Man and His World: A Structured Reader* (Kurilecz, 1969). Considered to be a reliable and valid measure of overall English language proficiency, it is used to assess vocabulary, morphosyntactic knowledge, and discourse competence (Chrabaszcz & Jiang, 2014). Fifty words were systematically deleted from the passage. They included nouns, verbs, adjectives, adverbs, prepositions, articles, and more (Kim, 2003). Written instructions on how to take the test were given in Korean for easier comprehension, and participants were asked to provide one word for each blank.

4.2.2 Tasks

In order to maximize generalizability of findings, the study employed two tasks, a Hotel Task and a Venue Task, whose task complexity was manipulated along a resource-directing dimension of the CH, [± few elements], and a Task Condition participant variable, [± open solution]. As shown in Table 9, the two tasks will each have four versions: 1) simple open, [+ few elements, + open solution]; 2) complex open, [- few elements, + open solution]; 3) simple closed, [+ few elements, - open solution]; and 4) complex closed, [-few elements, - open solution].

Table 10. Operationalization of tasks

Open		Closed	
Simple	Complex	Simple	Complex
+ few elements + open solution	- few elements + open solution	+ few elements - open solution	- few elements - open solution

For the tasks, participants were required to choose the best location (hotel/venue) for a certain event and write a letter to the people who wanted to hold that event. They were given specific instructions to give detailed explanations on why they chose a certain location and why they did not choose the others. Participants were expected to be familiar with the format and scheme of this type of task, as they are exposed to similar listening tasks in English every year when they prepare for the College Scholastic Ability Test (a high-stakes annual college entrance exam similar to the American SAT) in high school. All texts were provided in Korean, as the pilot test showed that many students tended to copy a number of constructions when texts were in English. In order to ensure that only task closure was manipulated and all other conditions were consistent, the information about the locations was the same for both the open and closed tasks. Extra care was taken so that (un)favorable characteristics were evenly distributed among the locations to prevent participants choosing a location too easily. The only difference between the open and closed tasks was that the latter included the event holders' requirements, which participants had to consider when making their decision. As indicated in the task instructions, those in the Open condition were aware that the open task did not have a correct answer, and those in the Closed condition were aware that there was only one correct answer. Task complexity was increased in terms of the number of locations to choose from, the amount of information about the locations (e.g., budget, internet, daily rate, etc.), and requirements involved in the task. Table 11 summarizes how the number of elements was manipulated.

Table 11. Number of elements

	Simple		Complex	
Task	Hotel	Venue	Hotel	Venue
Location	3	3	4	4
Location characteristics	3	3	6	6
Closed task requirements	3	3	6	6

4.2.3 Cognitive Load Measures

Three measures were used to assess the cognitive load of the tasks: self-ratings, expert judgments, and time-on-task.

4.2.3.1 Self-ratings

After participants performed each task version, they completed a questionnaire that required them to indicate their answers on a nine-point Likert scale regarding (1) the overall perceived difficulty of the task, (2) the level of mental effort they thought was required to perform the task, (3) the level of stress they felt during task performance, (4) the level of interest they felt during task performance, (5) the amount of time pressure they felt during the planning stage of the task, (6) the amount of time pressure they felt during the writing stage, and (7) how well they believed they performed the task.

4.2.3.2 Expert Judgments

Expert judgments were also used to evaluate the cognitive load of the tasks. Three professors and four high school teachers considered to be experts were provided with all of the tasks given to the undergraduate participants, and were asked to answer questions on a nine-point Likert scale about how difficult participants would perceive the task and how much mental effort they would exert in order to perform the task. Also included in this questionnaire were open-ended questions that asked their opinions as to what caused a difference, if any, between levels of overall difficulty or mental effort between certain tasks.

4.2.3.3 Time-on-task

To date, the present study is the only one in task-based research that used time-on-task to measure cognitive load. Participants carried out tasks on Enterprise Learning Management System (ELMS), an online system for students and instructors at the University of Maryland. Users can upload/take exams and quizzes, among many other features. The tasks employed in the present study were uploaded as quizzes, and participants' activities were logged during task performance with timestamps. Figure 3 is an example of how a participant's activities were logged. For each task version, the participants saw two questions. In Question 1, task instructions were provided and they were asked to plan what they were going to write. When they had finished planning, they were required to click on a button that read 'I am ready to write' and then click 'next' to proceed to the writing stage. This would lead them to Question 2, where they would type in their letter in a blank box. As in Question 1, they were shown the same task instructions. Dividing each task version into two questions enabled the researcher to measure separately how long it took to plan their letter (time-on-planning), to write their letter (time-on-writing), and time on the task as a whole (time-on-whole task).

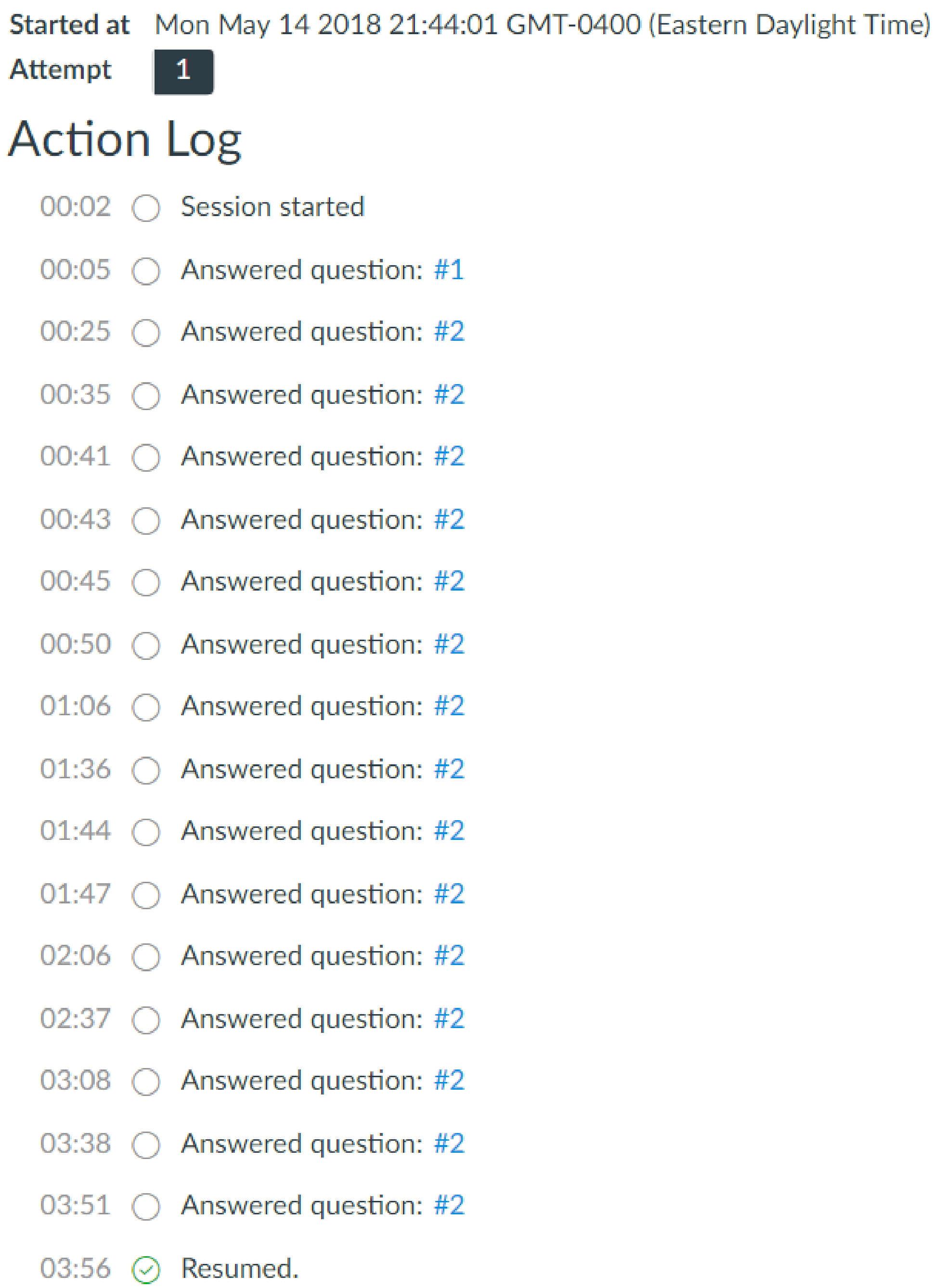


Figure 3. Action log on ELMS website

4.2.4 Working Memory Measure

A modified version of the Operation Span (Ospan) task (Engle, Cantor, & Carullo, 1992; Malone, 2018) was employed as a complex measure of WM to tap the capacity to store and process information. DMDX software (Forster & Forster, 2003) was used to administer the test. Participants saw a series of simple math equations with an uppercase English letter beside each one, read them aloud, and indicated whether the equations were correct by pressing “O” if correct, and “X” if incorrect. At the same time, they had to recall the letters that they saw in blocks of two to five. Participants took ten minutes at most to complete the test. In order to avoid L1 effects, all instructions were given in Korean. Figure 4 shows an example of how the equations and letters are presented.

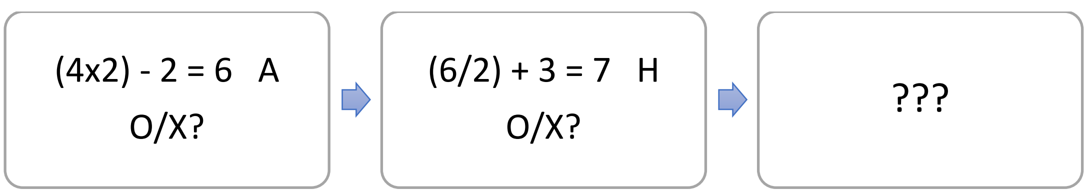


Figure 4. Example of two-block question in Ospan task

4.2.5 Language Aptitude Measures

The LLAMA test (Meara, 2005) was used to measure language aptitude. A computer-based test available for free download, it comprises four subtests, each of which takes ten minutes at most to complete. LLAMA B tests the ability to learn a large amount of vocabulary in a short period of time, LLAMA D to recognize foreign sound patterns, LLAMA E to associate sounds and symbols, and LLAMA F to extract syntactic and morphological patterns of an unfamiliar language. Efforts were made to ensure that there was a minimal L1 effect, and it was found that LLAMA scores were not affected by language background (Grañena, 2013; Rogers, Meara, Barnett-Legh, Curry, & Davie, 2017). LLAMA B uses words taken from a Central American language and arbitrarily assigns the words to target images, LLAMA D uses computer-generated sound sequences based on names of flowers and natural objects in a British Columbian Indian language, LLAMA E uses an unfamiliar alphabet, and LLAMA F's interface does not require any L1 input. For the present study, modified versions of the LLAMA D and LLAMA F were used to measure implicit language aptitude (ILA) and explicit language aptitude (ELA), respectively. Because the original LLAMA test only provides total scores for each subtest, LLAMA D and LLAMA F were modified to be administered on ELMS so that individual responses could be recorded. All instructions were translated into Korean.

4.2.5.1 LLAMA D

LLAMA D was used to measure the ability to recognize sound patterns in an unfamiliar language. In other words, it measures the ability to recognize small morphological variations that signal grammatical features in many languages (Meara, 2005). After listening to a string of 10 computer-generated sound sequences based on a British-Columbian Indian language, participants immediately moved on to the next phase of the test, where they listened to 30 sound sequences one at a time and indicated whether they believed they had heard the sound before or not. Participants were required to play the sounds only once during the recognition test phase.

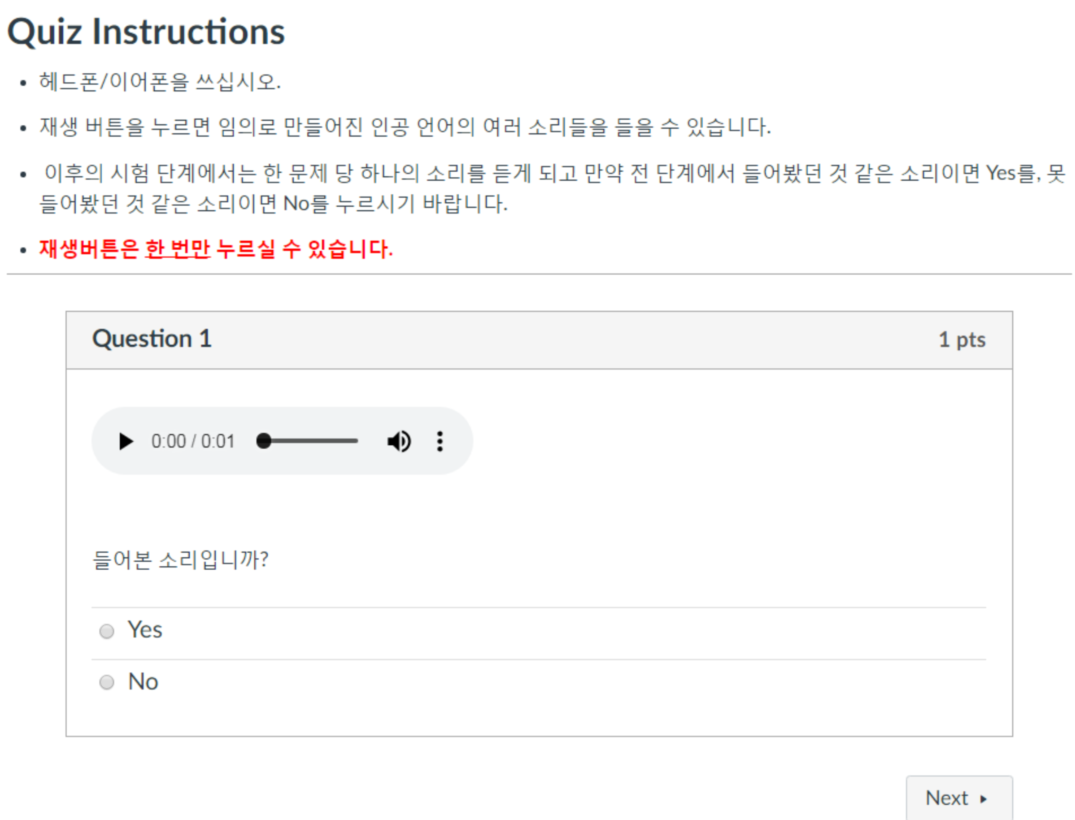


Figure 5. Example of LLAMA D question

4.2.5.2 LLAMA F

LLAMA F was used to tap the ability to infer the grammar rules that govern an unfamiliar language. For up to five minutes, participants saw a set of 20 pictures and sentences describing the pictures. During this timed study phase, they tried to figure out the grammatical rules of the language. They were allowed to take notes, which could be used during the test phase. Then one picture at a time, they saw a total of 30 pictures and two sentences under each, only one of which was grammatically correct. Participants had to choose the sentence they considered correct. The original LLAMA F test contained 20 items in the test phase, but an additional 10 items were added in order to increase the reliability of the test (Suzuki & DeKeyser, 2016).

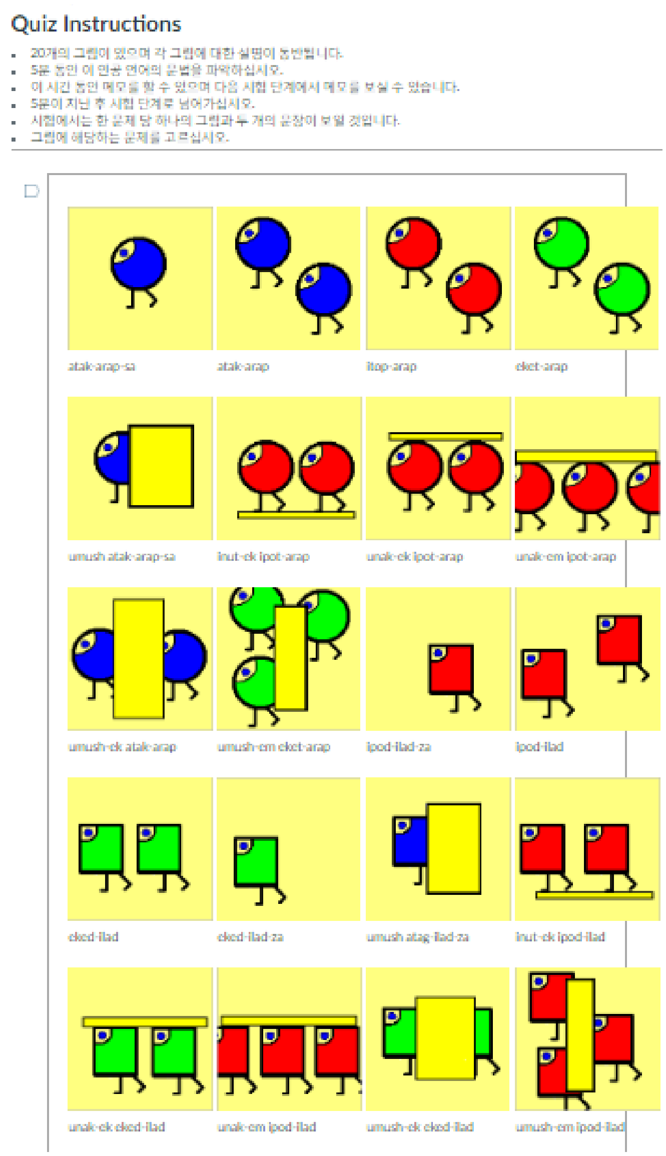


Figure 6. LLAMA F timed study phase

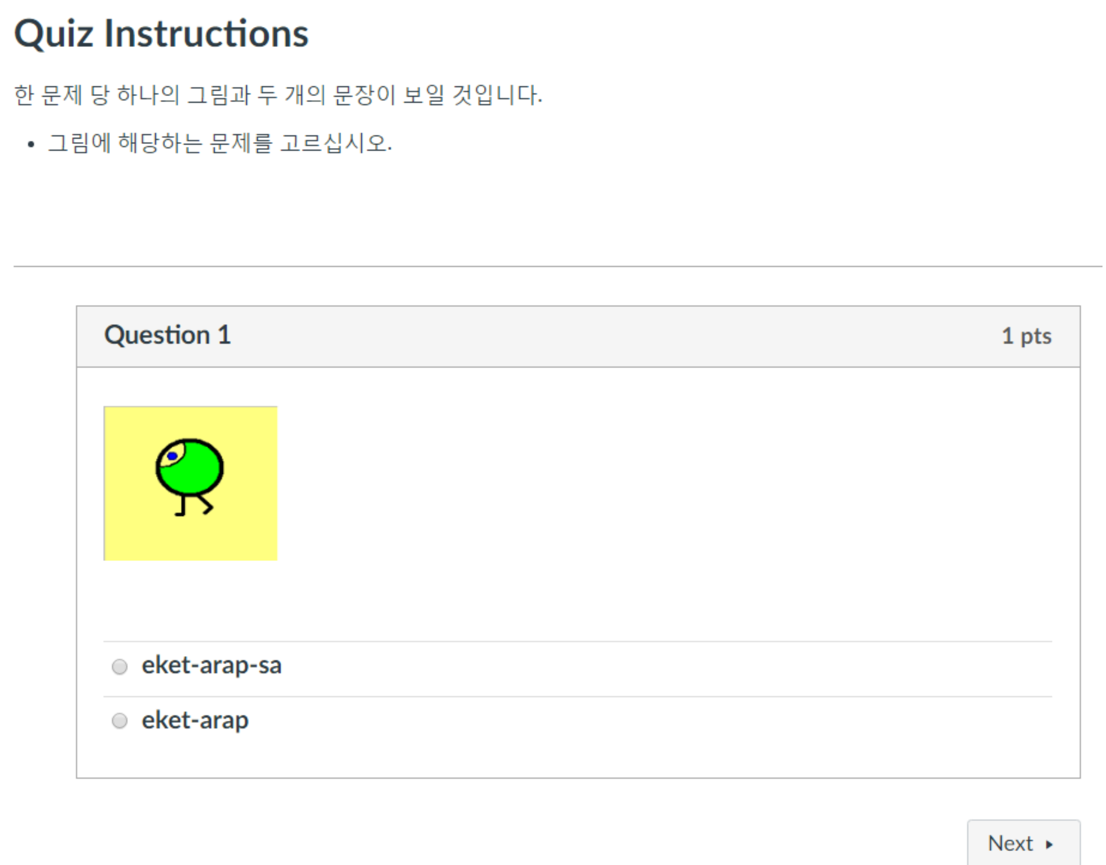


Figure 7. LLAMA F test phase

4.2.6 Language Background Questionnaire

Participants were required to complete a language background questionnaire at the beginning of the session. The questionnaire included questions about their gender, age, major, scores on a standardized English test (e.g., TOEFL, TOEIC, TEPS, and IELTS) if they had taken a test recently, age when they first started to learn English, list of foreign languages they knew, length of residence in an English-speaking country, number of English-related courses at the university level, and amount of English use in terms of reading, listening, speaking, and writing.

4.3 Procedure

Participants met with the researcher individually in an office at Inha University for one session. Upon arrival, they took a cloze best (Brown, 1980) on paper in order to measure their overall English proficiency. The then completed a language background questionnaire and took an Ospan test, administered on a computer. They were randomly assigned to either the Open or Closed condition and performed two tasks, each of which had two levels of complexity—those in the Open condition carried out open tasks, those in the Closed condition closed tasks. In order to prevent sequencing effects, tasks were pseudo-randomly ordered, such that participants would be alternating between the two tasks (see Table 12). After carrying out each task version, participants completed a questionnaire to indicate their ratings regarding perceived difficulty, mental effort, level of stress, time pressure felt during planning and writing, and performance expectations. Upon completion of the tasks, they took the LLAMA D and F tests. The rationale behind placing the Ospan task first, and then the tasks, followed by the LLAMA subtests, was to minimize mental fatigue they would have felt if they had had to take all of the cognitive tests sequentially. On average, each session lasted approximately 90 minutes. Participants could take breaks between tests if needed, but very few actually did. Table 13 summarizes the order and average time on the measures.

Table 12. Task order randomization

	Task sequence			
Participant	1st	2nd	3rd	4th
1	Hotel simple	Venue simple	Hotel complex	Venue complex
2	Hotel simple	Venue complex	Hotel complex	Venue simple
3	Hotel complex	Venue simple	Hotel simple	Venue complex
4	Hotel complex	Venue complex	Hotel simple	Venue simple
5	Venue simple	Hotel simple	Venue complex	Hotel complex
6	Venue simple	Hotel complex	Venue complex	Hotel simple
7	Venue complex	Hotel simple	Venue simple	Hotel complex
8	Venue complex	Hotel complex	Venue simple	Hotel simple

Table 13. Procedure

Order	Task	Average time (in minutes)
1	Cloze test	20
2	Language background questionnaire	5
3	Ospan test	10
4	Tasks	40
5	LLAMA D	5
6	LLAMA F	10
Total		90

4.4 Measurement and Data Analysis

4.4.1 Cloze Test

Brown's (1980) cloze test was used to measure participants' overall English proficiency. The exact scoring method was used, whereby responses were scored as correct only when they were exactly the same as in the original text. Each accurate response received a score of one point, yielding a final score of up to 50 points.

4.4.2 Cognitive Load Measures

In order to answer the research questions regarding cognitive load, a number of measures were used to assess whether increasing task complexity, in terms of ± number of elements, actually required heavier cognitive processing.

4.4.2.1 Self-ratings

Participants' answers to a questionnaire regarding the following were analyzed separately: (1) the overall perceived difficulty of the task, (2) the level of mental effort they thought was required to perform the task, (3) the level of stress they felt during task performance, (4) the amount of time pressure they felt during the planning stage of the task, (5) the amount of time pressure they felt during the writing stage, and (6) how well they believed they performed the task. Because the answers were in digit form, only one rater was needed to compile and analyze the data.

4.4.2.2 Expert Judgments

The answers from experts (three professors and four high school teachers) to questions regarding the overall perceived difficulty and mental effort required to perform each task were evaluated. For these questions, the experts provided nine-point Likert scale ratings as their answers, which were analyzed by one rater. Two raters examined experts' answers to open-ended questions about what they thought caused a difference, if any, in their ratings of overall difficulty and/or mental effort ratings between simple vs complex task versions and between open vs. closed tasks.

Answers to these open-ended questions were divided into two superordinate categories: task-inherent factors and learner-based factors.

4.4.2.3 Time-on-task

In ELMS, where the tasks used in the present study were uploaded for participants to carry out, participants' activities were logged and timestamped. With this feature and the addition of two questions per task version, it was possible for the researcher to measure the time engaged in planning, writing, and the task as a whole. Time in seconds was used for data analyses.

4.4.3 Working Memory Measure

A modified version of the Operation Span (Ospan) task (Engle, Cantor, & Carullo, 1992; Malone, 2018) was used to tap the capacity to store and process information, i.e., complex WM. Participants were required to recall alphabet letters in blocks of two to five letters, while also indicating whether the mathematical equations that they saw were correct or not. They had to recall a total of 11 alphabetical blocks, and indicate the correctness of 40 mathematical equations. One coder examined and analyzed their answers. For the math problems, one point was given for each correct answer, and all participants met the 80% minimum threshold (at least 32 out of 40, $M = 37.89$, $STD = 1.91$), indicating focus on the primary alphabet recalling task. For the primary task, an all-or-nothing load scoring (ANL) method was employed, which reflects the total number of letters recalled in the correct position from each block (see Conway, Kane, Bunting, Hambrick, Wilhelm, & Engle, 2005). ANL was measured by the number of correctly recalled letters divided by the total number of letters.

4.4.4 Language Aptitude Measures

Modified versions of LLAMA D (a sound recognition test) and F (a grammatical inferencing test) were used to tap the abilities relevant for implicit learning and explicit learning, respectively. Each test included a total of 30 questions, and one point was given for each correct answer, yielding final scores of up to 30 points for both tests.

4.4.5 Linguistic Outcome Measures

Participants' written production was assessed in terms of syntactic complexity, lexical diversity, and accuracy. The following provides detailed descriptions of the exact measures used for each.

4.4.5.1 Syntactic complexity

Syntactic complexity was examined in terms of two aspects: length of production and subordination. Because the present study focused on the written production of adult L2 learners, syntactic measures were based on the T-unit, defined as a minimal unit consisting of a main clause and any subordinate clauses embedded or attached to (Hunt, 1964). Subordinate clauses included finite clauses (e.g., subordinate clauses that began with a subordinate conjunction or a relative pronoun) and non-finite subordinate clauses (e.g., a *to*-infinitive clause, -*ing* participle clause, or *–ed* participle clause). Length of production was measured by 1) mean length of T-unit (MLT), a measure of overall or general complexity calculated by dividing the total number of tokens by the total number of T-units, and 2) mean length of clause (MLC), a measure of sub-clausal complexity via phrasal elaboration, calculated by dividing the total number of tokens by the total number of clauses. Complexity via subordination was measured by the number of subordinate clauses per T-unit, calculated by dividing the total number of subordinate clauses by the total number of T-units.

4.4.5.2 Lexical diversity

Lexical diversity was measured by Guiraud's Index of Richness (1954), a mathematical transformation of the type-token ratio. It was calculated by dividing the number of types by the square root of the number of tokens. This measure is claimed to be preferred to the traditional type-token ratio, because it takes text length into consideration. The number of types and tokens were counted by a web version of a computer program that performs lexical text analysis, VocabProfile (Cobb, 2018; Heatley, Nation, & Coxhead, 2002).

4.4.5.3 Accuracy

Participants' accuracy in writing was examined largely in two ways: 1) the proportion of error-free structures, and 2) the proportion of errors. As a global measure of accuracy, the proportions of error-free T-units and error-free clauses were examined, calculated by the total number of error-free T-units divided by the total number of T-units, and by the total number of error-free clauses divided by the total number of clauses, respectively. As a specific measure of accuracy, the proportion of target-like use (TLU) of articles was employed, calculated by dividing the total number of correctly used articles by the total number of noun/noun phrases that required an article or not. In the case of the proportion of errors, the proportion of lexical errors was examined, calculated as the total number of lexical errors divided by the total number of T-units. When looking at error-free structures, errors included *all* errors detected, such as spelling mistakes, morphosyntactic errors, and lexical errors. In terms of lexical errors, wrong choices in words and omission of necessary words were included. Preposition errors were considered to be morphosyntactic errors and not counted as lexical errors. Spelling errors were also not included as lexical errors, because it was difficult to distinguish between typos and errors.

4.4.5.4 Inter-rater reliability

Because syntactic complexity and accuracy measures were subject to personal judgments, the entire data set was rated by two raters: the researcher and a professor with a Ph.D. in English semantics and syntax. The identification of subordinate clauses and lexical errors required the most discussion between the two raters. After a one-hour discussion on how to count such structures, the two raters coded the entirety of the data separately. Discrepancies were later reviewed, reconciled, and recoded. Inter-rater reliability was calculated by dividing the number of items where a discrepancy was found by the total number of items coded. Inter-rater reliability was

.859 for the number of subordinate clauses, and .746 for the number of lexical errors, indicating acceptably good agreement between the two.

4.4.6 Statistical Analysis

In order to answer the research questions stated in Chapter 3, descriptive statistics were computed, followed by a set of linear mixed-effects models (MEMs) with the maximum likelihood estimation. MEMs were preferred over the typical repeated measures analyses of variance (ANOVA) because the data sets for the study involved a series of participants tested on the same series of items. Participants were included in the model as random effects. All statistical analyses were run on STATA. The significance level was set at p = .05. Specific results were reported only when the Wald Chi-Squared Test (a.k.a. Wald Test) yielded significant results, which indicated that the explanatory variables in the model(s) were significant. Because of the vast number of hypotheses and tests, Bonferroni corrections were conducted to avoid an inflated Type I error. Cohen's f^2 was used to measure effect size. Table 14 provides an overview of the measures used in the present study, including the reliability estimates of the measures used to assess cognitive individual differences, estimated using Cronbach's Alpha (α).

Table 14. Overview of measures

Measure	Task	Details	Reliability
Cognitive load	Self-ratings	Overall difficulty Mental effort Stress Performance expectations Time pressure while planning	

		Time pressure while writing	
	Expert judgments	Overall difficulty Mental effort Answers to open-ended questions	
	Time-on-task	Planning Writing Task as a whole	
Linguistic outcome	Syntactic complexity	MLT MLC Number of subordinate clauses per T-unit	
	Lexical diversity	Guiraud's Index	
	Accuracy	Proportion of TLU articles Proportion of error-free T-units Proportion of error-free clauses Proportion of exical errors	
Individual differences	L2 proficiency	Brown's cloze text	
	WMC	Ospan task	.731
	Language Aptitude	LLAMA D (Implicit learning) LLAMA F (Explicit learning)	.454 .758

Chapter 5. Results

5.1 Descriptive Statistics

5.1.1 Individual Differences

Table 15 summarizes participants' scores on the cloze test, the Ospan task, LLAMA D, and LLAMA F. A series of independent samples t-tests was conducted, and results showed that there were no significant differences in scores between the Open and Closed conditions in terms of overall English proficiency, WMC, ILA, and ELA, $t(81) = .641$, $p = .523$; $t(81) = .096$, $p = .924$; $t(81) = 1.533$, $p = .129$; and $t(81) = .202$, $p = .841$, respectively.

Table 15. Descriptive statistics for results on tests measuring individual differences

Measure		Open ($N = 42$)	Closed ($N = 41$)
Cloze test	Mean	15.02	15.63
	Standard Deviation	4.10	4.65
	Minimum	6	8
	Maximum	28	28
Ospan task	Mean	0.71	0.72
	Standard Deviation	0.23	0.27
	Minimum	0	0
	Maximum	1	1
LLAMA D	Mean	19.97	20.83
	Standard Deviation	3.43	2.73
	Minimum	11	14
	Maximum	28	27
LLAMA F	Mean	24.62	24.78
	Standard Deviation	3.62	3.67
	Minimum	14	16
	Maximum	29	30

5.1.2 Cognitive Load Measures

Three measures were used to assess the cognitive load of the simple and complex versions of the Hotel task and the Venue task: 1) self-ratings of overall perceived difficulty, mental effort required, level of stress, level of interest, time pressure felt during the planning and writing stages, and performance expectations, 2) expert ratings of overall perceived difficulty and mental effort required, and 3) time-on-task. L2 learner participants and experts both used a nine-point Likert scale to indicate their ratings. In addition, experts provided answers to open-ended questions on the differences, if any, between simple vs. complex task versions and open vs. closed task versions. Time-on-task was used as an objective measure of cognitive load, and it was further divided into three parts: time-on-planning, time-on-writing, and time-on-whole task.

In order to see whether the learner self-ratings measured one single underlying construct, i.e., cognitive load, a principal component analysis with Verimax ration was conducted. The Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy was .811, suggesting that the partial correlations between the variables were adequate for analysis. Barlett's test of sphericity was significant ($p < .001$), indicating that there were significant relationships among the variables. The analysis yielded two principal components with eigen values greater than 1.0, which accounted for 69.88% of the total variance. The rotated component matrix showed that five of the seven variables loaded on the first component ($\lambda = 3.667$), which accounted for 52.39% of the total variance, and interest and performance expectation ratings loaded on the second component ($\lambda = 1.225$), which accounted for an additional 17.50% of the variance (see Table 16). Based on these findings, it was logical to assume that the level of interest felt during task performance and task performance expectations were not adequate measures of cognitive load, and thus were removed from further analysis.

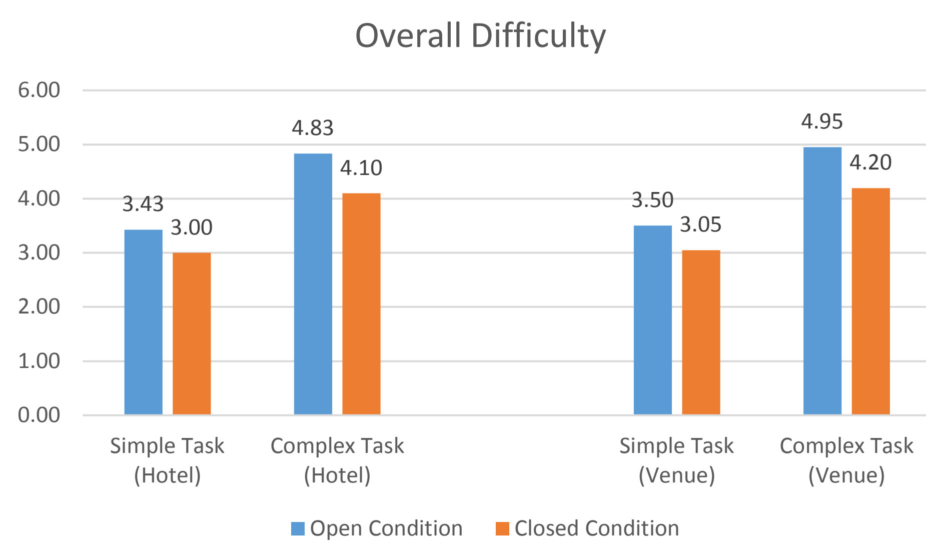
Table 16. Loadings of self-ratings in principal component analysis

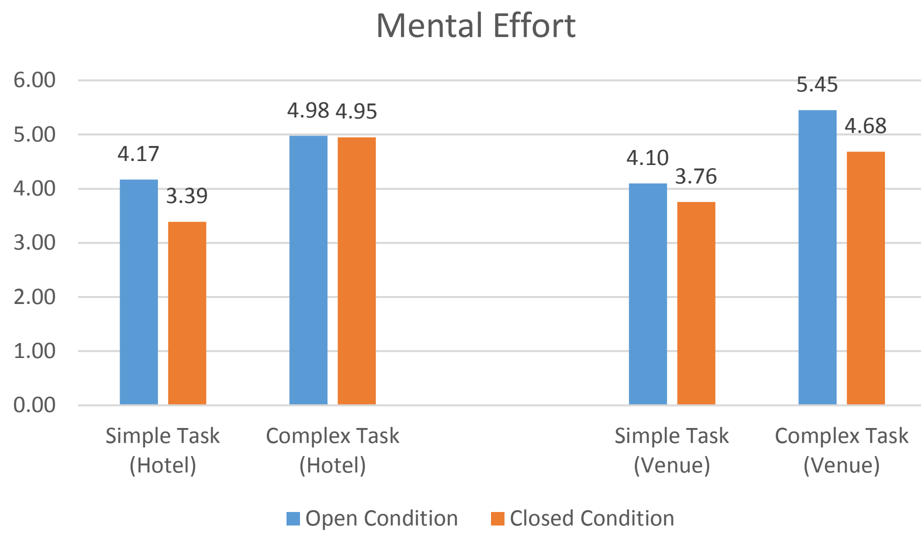
Measure	Component 1	Component 2
Difficulty	.826	
Mental Effort	.843	
Stress	.842	
Time pressure during planning	.775	
Time pressure during writing	.818	
Interest		.728
Performance expectations		.777

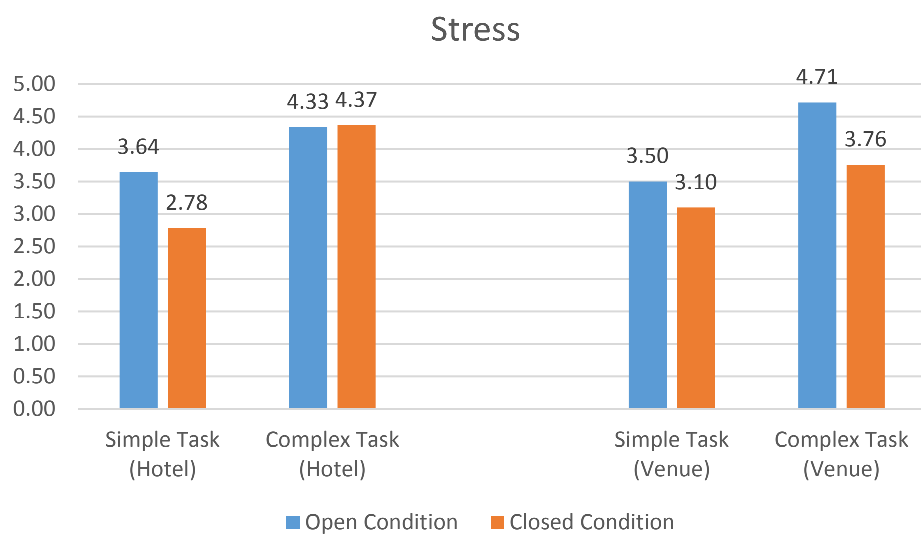
Table 17 and Figure 8 provide descriptive statistics for learner self-ratings. As predicted, the overall ratings for the complex task versions were higher than those of the simple versions. However, considering that the ratings could range from 1 to 9, the ratings were relatively low, which means that the tasks did not require a great degree of cognitive processing. One unexpected finding based on the descriptive statistics was that the average Open condition ratings regarding difficulty, mental effort, stress, and time pressure were higher than the average Closed condition ratings, indicating that open tasks were considered to be more cognitively challenging than closed tasks.

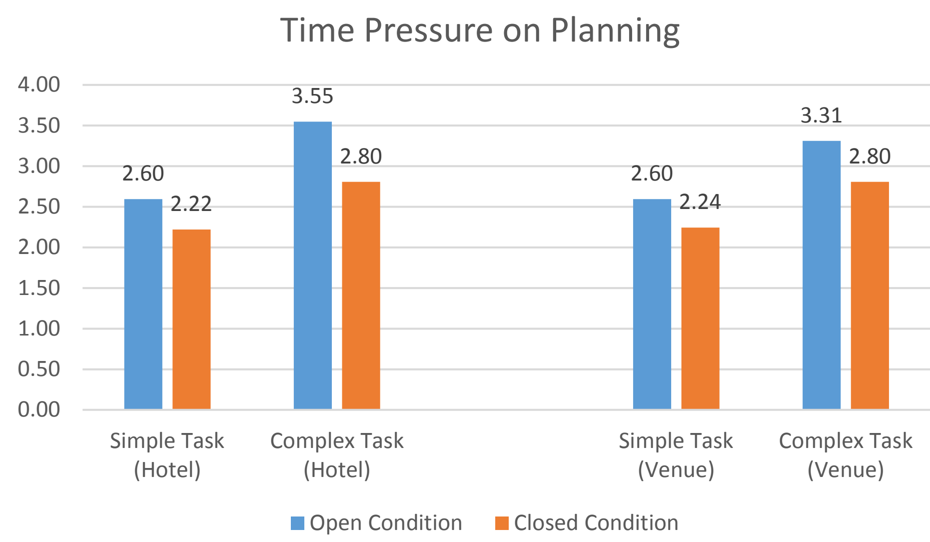
Table 17. Means and standard deviations of L2 learner self-ratings

Task	Hotel				Venue			
Complexity	Simple		Complex		Simple		Complex	
Condition	Open	Closed	Open	Closed	Open	Closed	Open	Closed
Difficulty	3.43 (1.50)	3.00 (1.45)	4.83 (1.74)	4.10 (1.76)	3.50 (1.70)	3.05 (1.50)	4.95 (1.46)	4.20 (1.87)
Mental Effort	4.17 (1.67)	3.39 (1.73)	4.98 (1.72)	4.95 (1.80)	4.10 (1.66)	3.76 (1.84)	5.45 (1.63)	4.68 (1.97)
Stress	3.64 (1.81)	2.78 (1.74)	4.33 (1.84)	4.37 (1.84)	3.50 (1.57)	3.10 (1.56)	4.71 (1.88)	3.76 (1.97)
Planning time pressure	2.60 (1.47)	2.22 (1.52)	3.55 (1.82)	2.80 (1.47)	2.60 (1.52)	2.24 (1.20)	3.31 (1.70)	2.80 (1.78)
Writing time pressure	2.98 (1.66)	2.85 (1.67)	4.26 (2.14)	4.12 (2.08)	3.24 (2.05)	3.00 (1.77)	3.79 (1.87)	3.80 (2.19)




Mental Effort
6.00
5.00
4.00
3.00
2.00
1.00
0.00
4.17
3.39
4.98 4.95
4.10
3.76
5.45
4.68
Simple Task (Hotel)
Complex Task (Hotel)
Simple Task (Venue)
Complex Task (Venue)
Open Condition
Closed Condition


Stress
5.00
4.50
4.00
3.50
3.00
2.50
2.00
1.50
1.00
0.50
0.00
3.64
2.78
4.33 4.37
3.50
3.10
4.71
3.76
Simple Task (Hotel)
Complex Task (Hotel)
Simple Task (Venue)
Complex Task (Venue)
Open Condition
Closed Condition



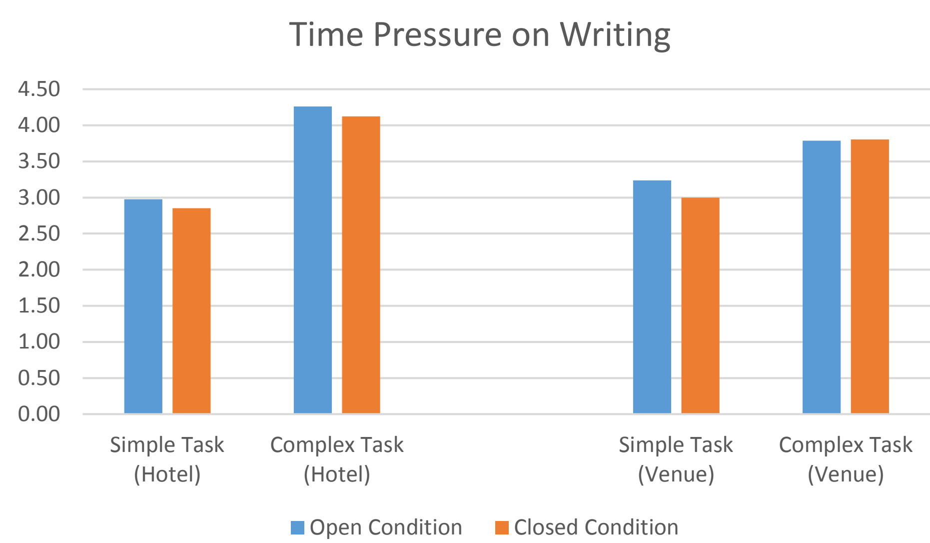


Figure 8. Self-ratings of cognitive load

The patterns of the expert ratings on overall task difficulty and mental effort required were similar to those of L2 learners, in that simple versions were rated lower than complex versions. Furthermore, contrary to predictions, but similar to learner ratings, open tasks were rated as more cognitively challenging than closed tasks. In general, expert ratings were higher than learner ratings, meaning that experts thought that the tasks would place a greater load on the learners than they actually did. Table 18 and Figures 9 illustrate these patterns.

Table 18. Means and standard deviations of expert judgments

Task	Hotel				Venue			
Complexity	Simple		Complex		Simple		Complex	
Condition	Open	Closed	Open	Closed	Open	Closed	Open	Closed
Difficulty	4.57 (1.51)	3.43 (1.72)	6.00 (1.29)	5.14 (1.95)	5.43 (2.07)	4.00 (1.91)	6.71 (1.38)	5.57 (2.07)
Mental Effort	5.00 (1.83)	3.71 (1.80)	6.43 (1.51)	5.29 (1.98)	5.71 (2.29)	4.14 (1.95)	6.86 (1.46)	5.71 (2.06)

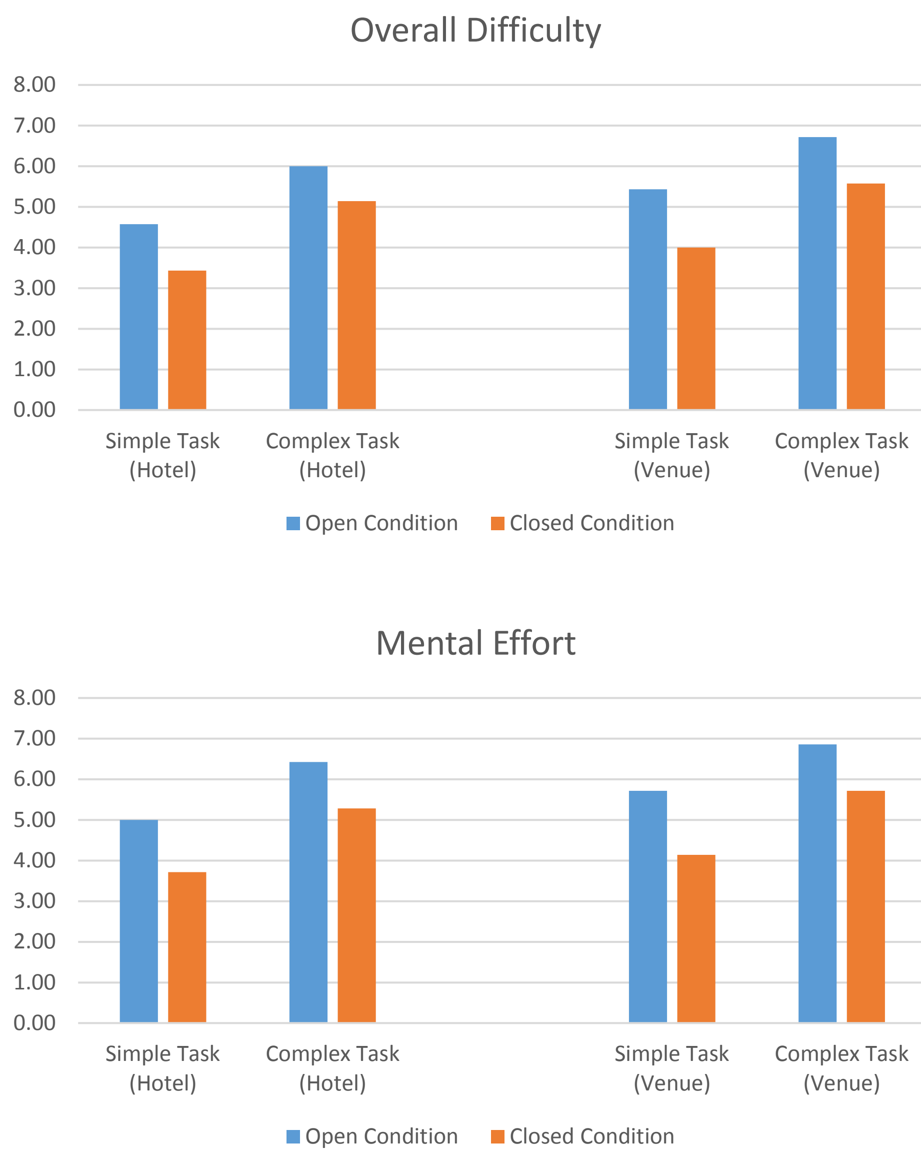


Figure 9. Expert self-ratings

Experts' responses to open-ended questions as to why there were differences in the degree of difficulty or mental effort required by simple vs. complex tasks and open vs. closed tasks were largely divided into two categories: task-inherent factors and learner-based factors (see Table 19). Task-inherent factors indicate characteristics of the task themselves, analogous to a combination of Robinson's Task complexity and Task condition factors. Learner-based factors are similar to Robinson's Task difficulty factor, and involve activities or abilities required for the learner to carry out the task successfully.

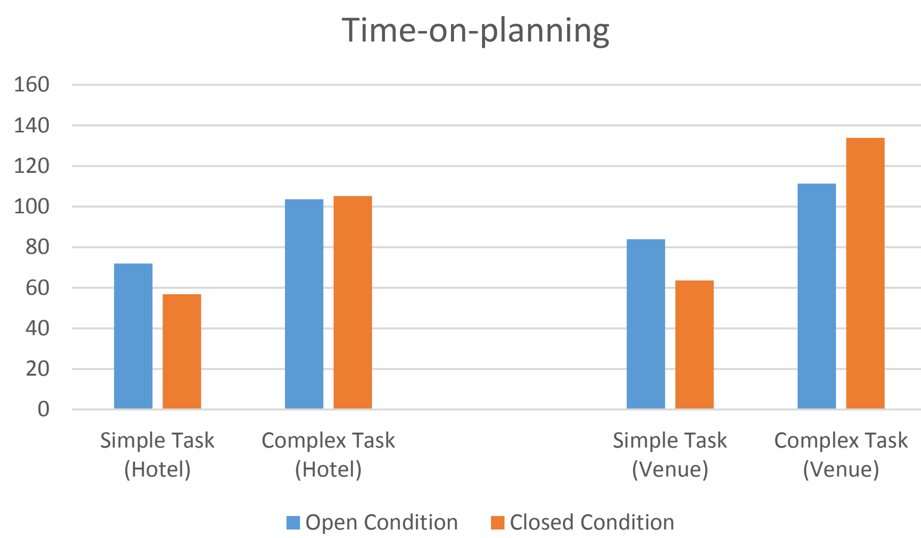
Table 19. Experts' answers to open-ended questions

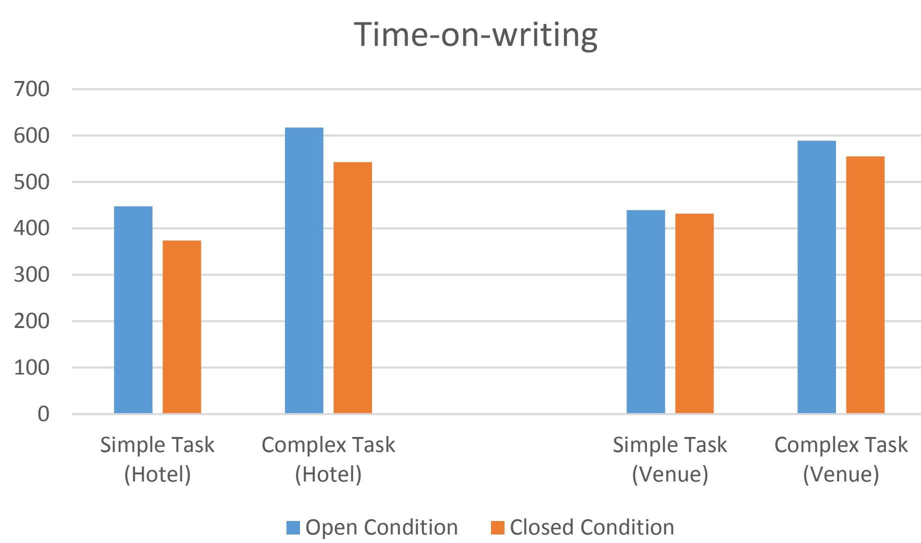
	Task-inherent factors	Learner-based factors
Simple	- Number of elements	- Prioritization of categories
Complex	+ Number of elements	+ Prioritization of categories
Open	- Number of requirements + Planning stage	+ Ability to adapt, synthesize, and infer information - Familiarity - Knowledge of reader preferences In-depth compare & contrast Divergent thinking
Closed	+ Existence of requirements - Planning stage	- Ability to adapt, synthesize, and infer information + Familiarity Repetition of requirements Awareness of one answer Surface compare & contrast Convergent thinking

The time engaged in planning, writing, and task as a whole were also used to assess cognitive load. Time spent on the planning stage was much shorter than anticipated, as it only took roughly one minute for participants to plan their writing for the simple versions, and slightly less than two minutes for the complex versions. However, it should be noted that idea formulation can occur simultaneously while writing. In this regard, time spent on the whole task was deemed a necessary measure to include in the analysis. As expected, time-on-writing was much longer than time-on-planning: slightly over seven minutes for the simple versions, and between nine and 10 minutes for the complex versions. When comparing the open and closed tasks, descriptive statistics for this measure for the learner and expert ratings were consistent: in general, participants in the Open condition spent more time on the tasks than those in the Closed condition, indicating a greater cognitive load in the former condition. The exception to this general finding was found in the case of time-on-planning for the complex tasks, especially the Venue task. In this case, those in the Closed condition took an average of 22 more seconds than those in the Open condition to plan their writing. In other words, the requirements that were included in the closed complex Venue task made participants think more about how to write, most likely because they needed to compare and contrast the information given, in order to find the venue that best met all requirements. Descriptive statistics for time-on-task measures are provided in Table 20 and Figure 10.

Table 20. Means and standard deviations of time-on-task (in seconds)

Task	Hotel				Venue			
Complexity	Simple		Complex		Simple		Complex	
Condition	Open	Closed	Open	Closed	Open	Closed	Open	Closed
Time-on-planning	71.88 (69.81)	56.80 (31.47)	103.67 (80.79)	105.20 (56.15)	83.90 (70.38)	63.63 (34.77)	111.38 (82.95)	133.90 (112.90)
Time-on-writing	447.67 (206.10)	374.17 (196.53)	617.26 (340.57)	542.59 (259.62)	439.33 (220.23)	431.68 (246.55)	588.74 (263.17)	555.37 (284.11)
Time-on-whole task	519.55 (214.79)	430.98 (207.32)	720.93 (377.78)	647.78 (259.24)	523.24 (256.59)	495.32 (262.15)	700.12 (287.93)	689.27 (329.80)


Time-on-planning
160
140
120
100
80
60
40
20
0
Simple Task (Hotel)
Complex Task (Hotel)
Simple Task (Venue)
Complex Task (Venue)
Open Condition
Closed Condition


Time-on-writing
700
600
500
400
300
200
100
0
Simple Task (Hotel)
Complex Task (Hotel)
Simple Task (Venue)
Complex Task (Venue)
Open Condition
Closed Condition

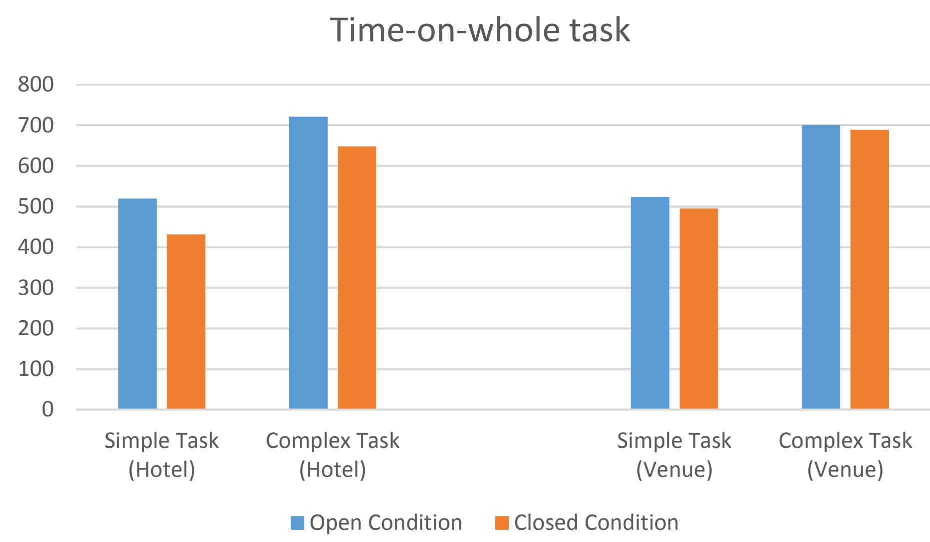


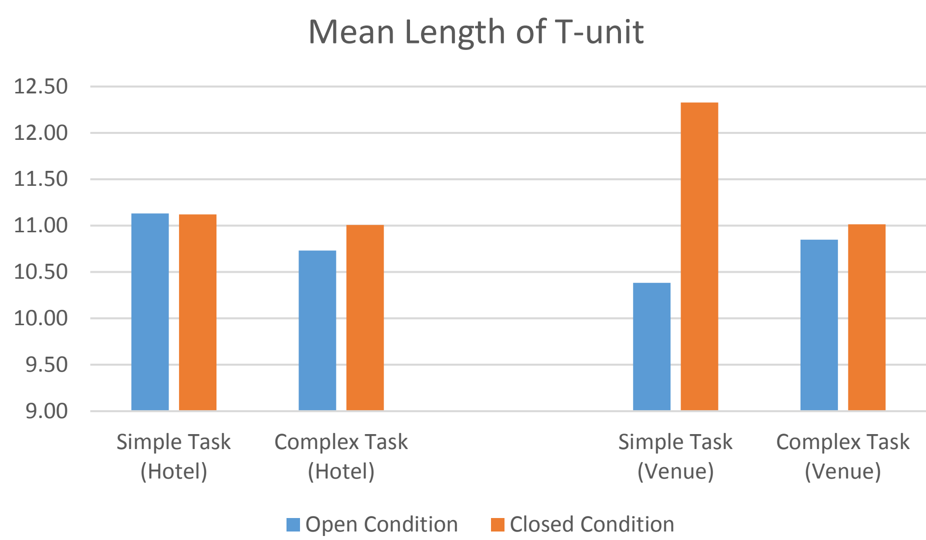
Figure 10. Time-on-task measures

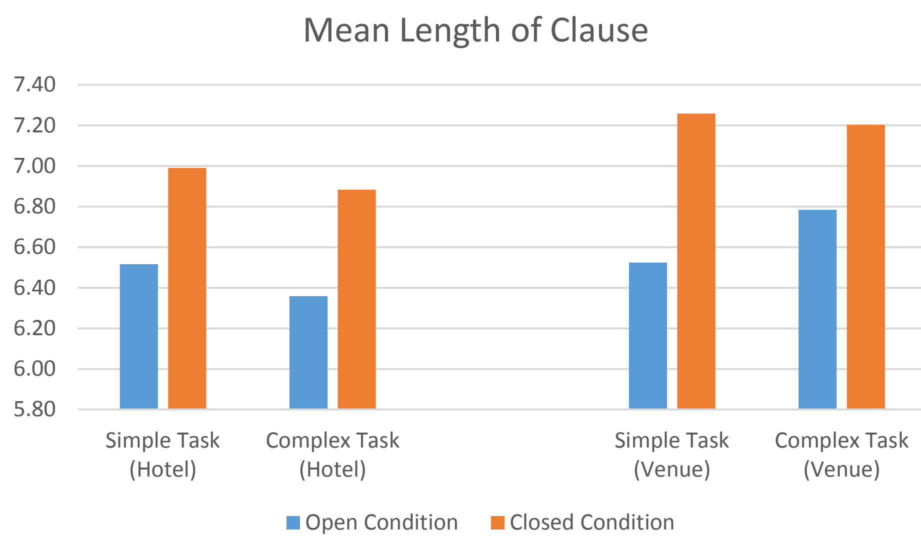
5.1.3 Linguistic Outcome Measures

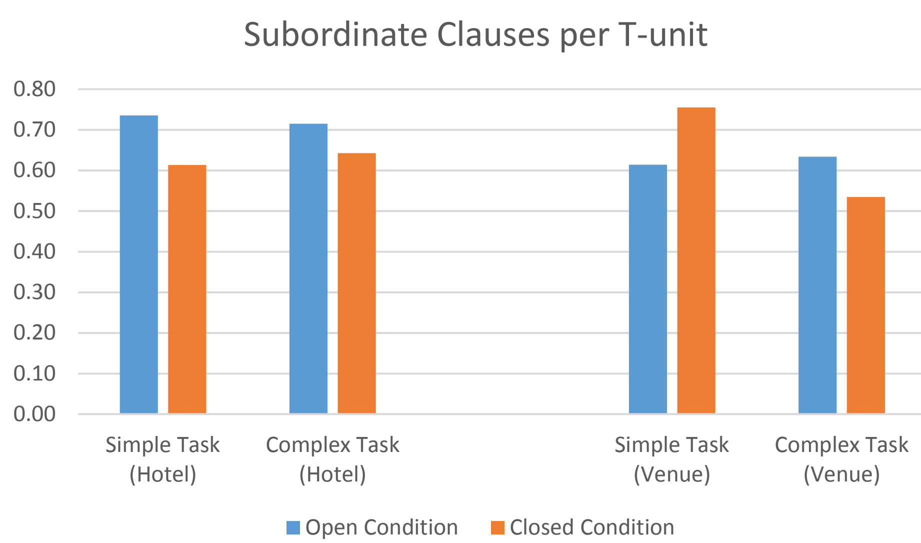
The descriptive statistics for the outcome measures of syntactic complexity and lexical diversity are shown in Table 21 and Figure 11. In terms of the three syntactic complexity measures, MLT, MLC, and subordinate clauses per T-unit, there did not seem to be a consistent pattern between simple and closed versions that matched the predictions of the study. On the other hand, the complex versions appeared to elicit greater lexical diversity for both tasks, indicating that participants used a wider variety of words when carrying out the complex versions. In terms of the comparison between the Open and Closed conditions, the only consistent pattern was found for MLC, i.e., mean length of clauses. Not only was this pattern found in both tasks, but it also was in line with the predictions of the study, in that closed tasks would elicit greater linguistic complexity than open tasks.

Table 21. Means and standard deviations of linguistic complexity measures

Task	Hotel				Venue			
Complexity	Simple		Complex		Simple		Complex	
Condition	Open	Closed	Open	Closed	Open	Closed	Open	Closed
MLT	11.13 (2.70)	11.12 (3.04)	10.73 (2.14)	11.01 (2.77)	10.38 (1.96)	12.33 (5.35)	10.85 (2.44)	11.01 (2.96)
MLC	6.52 (1.18)	6.99 (1.30)	6.36 (0.95)	6.88 (1.24)	6.52 (1.02)	7.26 (1.56)	6.78 (1.40)	7.20 (1.21)
Subclause per T-unit	0.74 (0.43)	0.61 (0.42)	0.72 (0.40)	0.64 (0.53)	0.61 (0.32)	0.75 (0.92)	0.63 (0.36)	0.53 (0.33)
GI	6.16 (0.93)	5.74 (0.75)	6.77 (0.78)	6.14 (0.69)	6.06 (0.82)	5.48 (0.78)	6.91 (1.19)	6.25 (0.70)




Mean Length of Clause
7.40
7.20
7.00
6.80
6.60
6.40
6.20
6.00
5.80
Simple Task (Hotel)
Complex Task (Hotel)
Simple Task (Venue)
Complex Task (Venue)
Open Condition
Closed Condition


Subordinate Clauses per T-unit
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00
Simple Task (Hotel)
Complex Task (Hotel)
Simple Task (Venue)
Complex Task (Venue)
Open Condition
Closed Condition

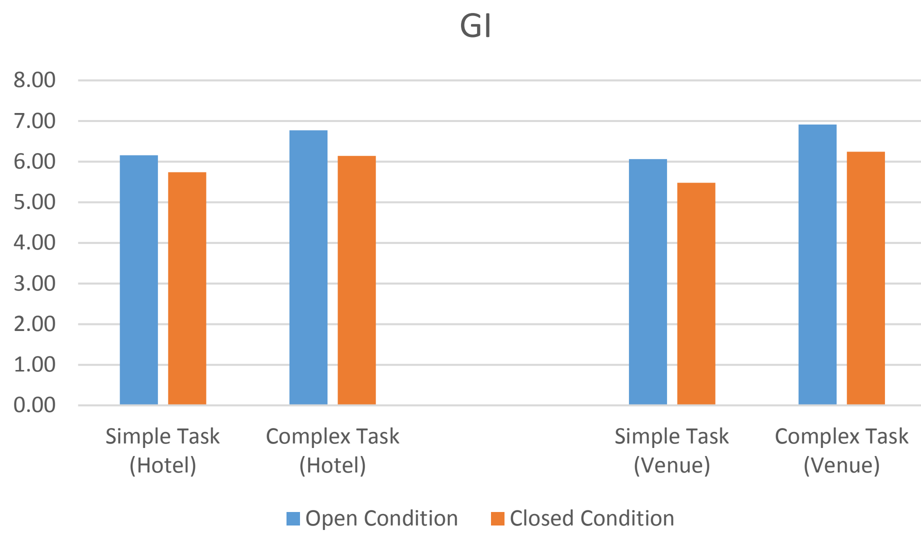
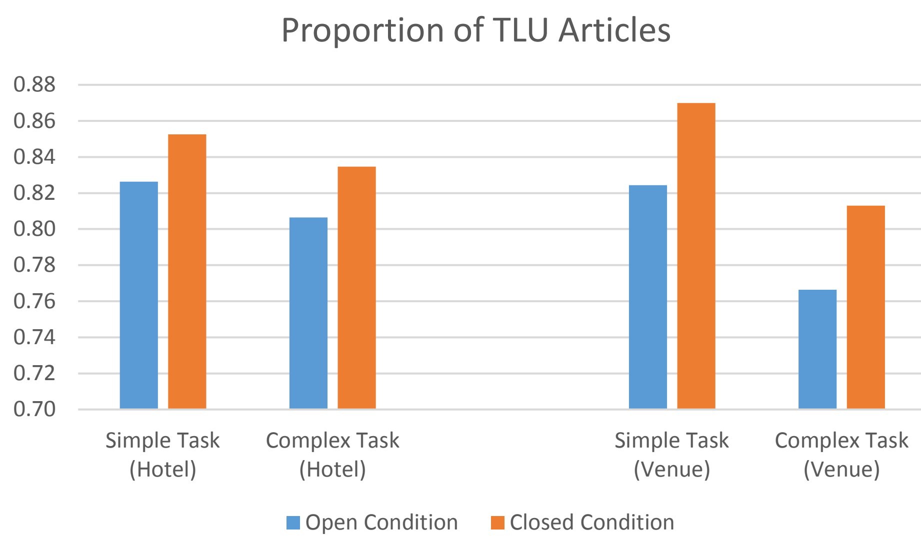


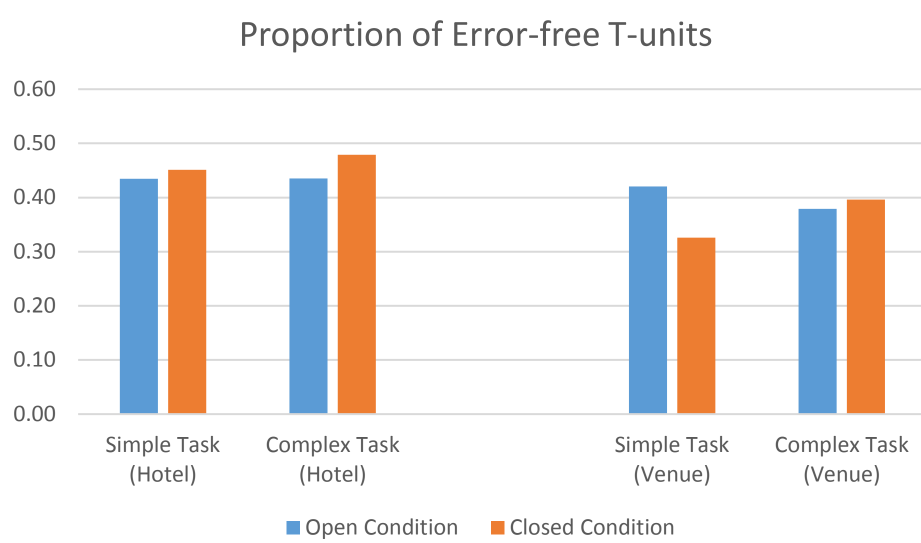
Figure 11. Linguistic complexity measures

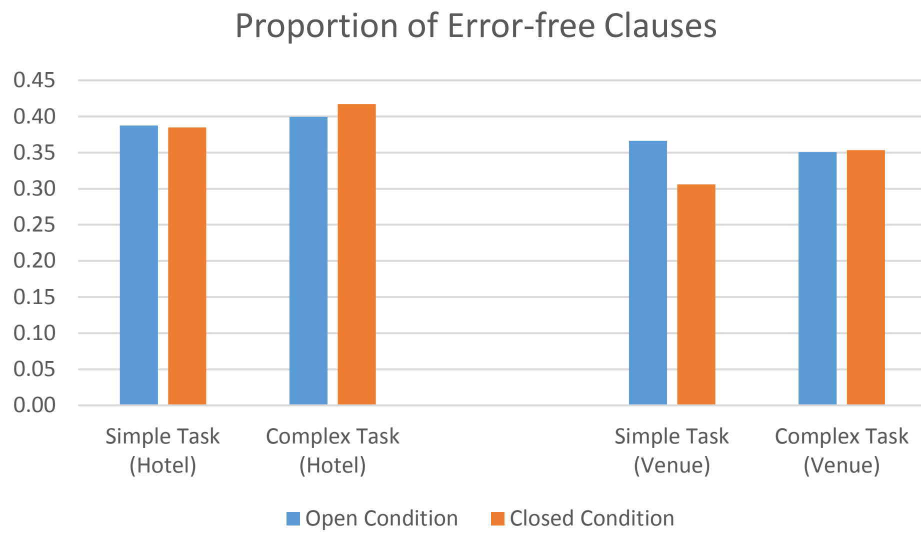
Participants' accuracy was measured in terms of the proportions of TLU articles, error-free T-units, error-free clauses, and lexical errors. Descriptive statistics for these measures are presented in Table 22 and Figure 12. A consistent pattern was found in the proportion of TLU articles between the simple and complex versions of both tasks. Counter to predictions and the CH, participants used fewer target-like articles in the complex versions, lending support for TOH, in that limited attentional resources make it difficult for learners to focus on complexity and accuracy at the same time. On the other hand, there appeared to be a positive effect for task complexity on the proportion of error-free T-units, especially for those in the Closed condition.

Table 22. Means and standard deviations of accuracy measures (proportions)

Task	Hotel				Venue			
Complexity	Simple		Complex		Simple		Complex	
Condition	Open	Closed	Open	Closed	Open	Closed	Open	Closed
TLU articles	0.83 (0.13)	0.85 (0.11)	0.81 (0.10)	0.83 (0.11)	0.82 (0.14)	0.87 (0.08)	0.77 (0.14)	0.81 (0.10)
Error-free T-units	0.43 (0.21)	0.45 (0.27)	0.44 (0.19)	0.48 (0.19)	0.42 (0.21)	0.33 (0.21)	0.38 (0.18)	0.40 (0.18)
Error-free clauses	0.39 (0.22)	0.38 (0.27)	0.40 (0.21)	0.42 (0.21)	0.37 (0.22)	0.31 (0.20)	0.35 (0.19)	0.35 (0.20)
Lexical errors	0.02 (0.02)	0.02 (0.02)	0.02 (0.02)	0.02 (0.02)	0.03 (0.02)	0.04 (0.02)	0.03 (0.02)	0.03 (0.02)


Proportion of TLU Articles
0.88
0.86
0.84
0.82
0.80
0.78
0.76
0.74
0.72
0.70
Simple Task (Hotel)
Complex Task (Hotel)
Simple Task (Venue)
Complex Task (Venue)
Open Condition
Closed Condition


Proportion of Error-free T-units
0.60
0.50
0.40
0.30
0.20
0.10
0.00
Simple Task (Hotel)
Complex Task (Hotel)
Simple Task (Venue)
Complex Task (Venue)
Open Condition
Closed Condition



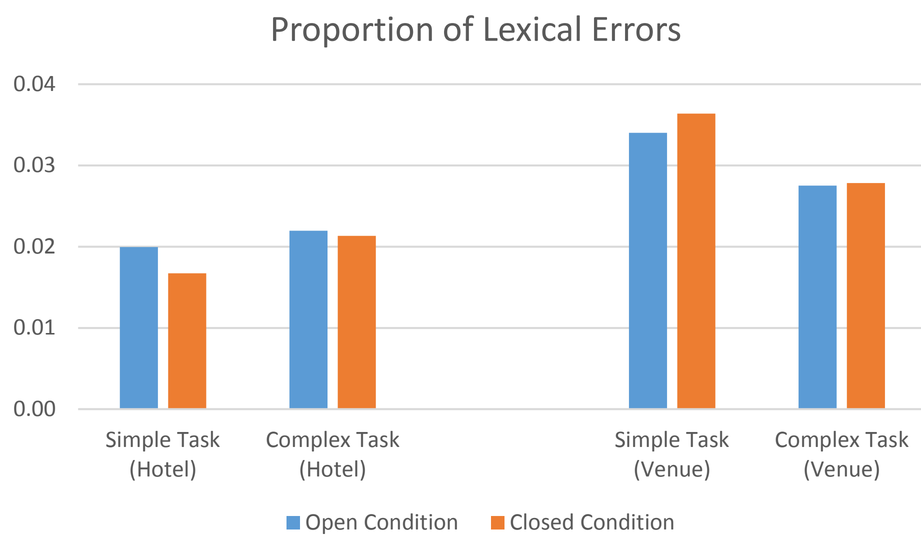


Figure 12. Accuracy measures

5.2 Mixed Effects Models

A series of mixed effects models (MEMs) was conducted, in order to investigate the effects of task complexity, manipulated in terms of number of elements, and task closure, whether a task was open or closed. In addition, several individual differences factors—working memory capacity, aptitudes for implicit and explicit learning, and English proficiency—were added to the models to see whether they affected L2 written performance, as well. Participants were determined as random effects for all models, and Bonferroni corrections were conducted to avoid an inflated Type I error due to the vast number of hypotheses and—statistical tests.

5.2.1 Task Effects on Cognitive Load

A series of MEMs was computed with a number of cognitive load measures serving as the dependent variable: 1) learner self-ratings on a nine-point Likert scale indicating overall perceived task difficulty, mental effort required for task performance, level of stress felt during task performance, time pressure felt during the planning stage, and time pressure felt during the writing stage, and 2) time spent on the planning stage, the writing stage, and the task as a whole. It was hypothesized that all measures of cognitive load would increase with increased task complexity. Moreover, closed tasks were hypothesized to place greater cognitive load on participants than open tasks.

On all measures, the Wald Test yielded significant results below the .01 significance level. Significant main effects of task complexity were found, such that complex tasks, when compared to simple tasks, led to significantly higher ratings of perceived difficulty (1.12 higher on 9-point Likert scale, $p < .001$, $f^2 = .574$), mental effort (1.24 higher, $p < .001$, $f^2 = .377$), stress (1.12 higher, $p < .001$, $f^2 = .241$), time pressure on planning (0.57 higher, $p < .001$, $f^2 = .176$) and writing (1.04 higher, $p < .001$, $f^2 = .224$), and significantly more time engaged in planning (59.329 more seconds, $p < .001$, $f^2 = .171$), writing (2.43 more minutes, $p < .001$, $f^2 = .313$), and task as a whole (3.42 more minutes, $p < .001$, $f^2 = .40$). A significant main effect of task closure was found on difficulty ratings, such that the ratings for the open task versions were higher than the closed versions (0.44 higher, $p = .05$, $f^2 = .010$). A significant interaction effect between task complexity and task closure was found on time-on-planning ($p = .03$, $f^2 = .017$). Post hoc Bonferroni pairwise comparisons revealed that there were significant differences in time-on-planning between the simple and complex versions for participants in both Open and Closed conditions ($p < .001$ and $p = .001$, respectively). Furthermore, significant differences were also found between the simple Open group and the complex Closed group, and the complex Open group and the simple Closed group ($p < .001$ for both), as shown in Figure 13.

Tables 23-30 summarizes the results obtained from the series of MEMs on cognitive load measures.

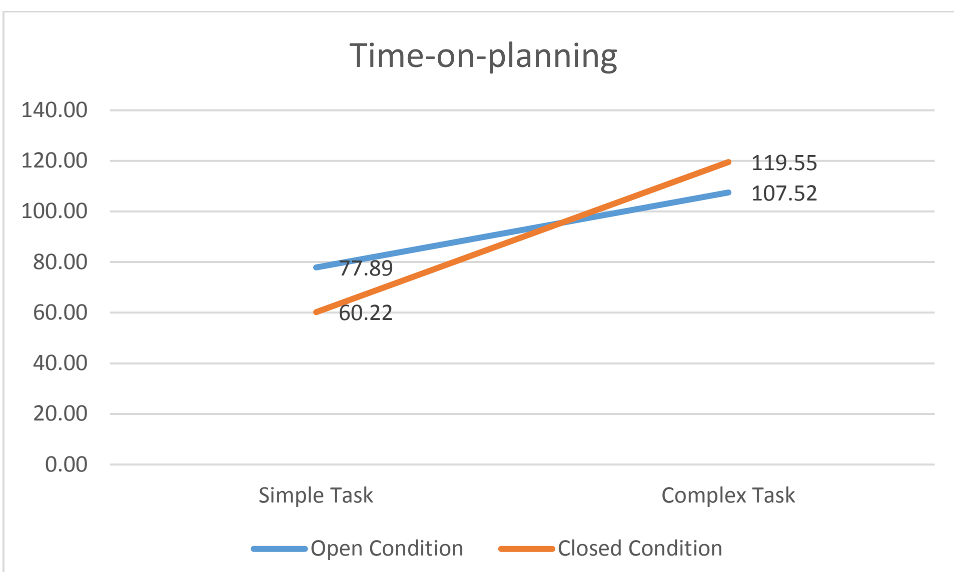


Figure 13. Significant interaction effect on time-on-planning (seconds)

Table 23. MEM results for difficulty ratings

Difficulty	Coef.	SE	z	P>|z|	[95% Conf. Interval]		f^2	Bonferroni p
Fixed-Effects								
Task Closure: Open	0.44	0.32	1.38	0.17	-0.19	1.07	0.010	0.05
Task Complexity: Complex	1.12	0.15	7.37	0.00	0.82	1.42	0.574	0.00
Task Closure × Task Complexity	0.31	0.21	1.43	0.15	-0.11	0.73	0.008	
constant	3.02	0.23	13.31	0.00	2.58	3.47		
Log Likelihood	-548.52							
Wald Chi-Square (*df*: 3)	148.36							
P-Value	0.00							
	Estimate	SE	[95% Conf. Interval]					
Random-Effects (Participant)								
SD (constant)	1.28	0.11	1.08	1.53				
SD (residual)	0.97	0.04	0.89	1.06				
LR Test vs. Linear Model	161.57							
P-Value	0.00							

Table 24. MEM results for mental effort ratings

Mental Effort	Coef.	SE	z	P>|z|	[95% Conf. Interval]		f^2	Bonferroni p
Fixed-Effects								
Task Closure: Open	0.56	0.34	1.63	0.10	-0.11	1.23	0.003	0.14
Task Complexity: Complex	1.24	0.17	7.30	0.00	0.91	1.58	0.377	0.00
Task Closure × Task Complexity	-0.16	0.24	-0.67	0.50	-0.63	0.31	0.002	
constant	3.57	0.24	14.66	0.00	3.10	4.05		
Log Likelihood	-581.70							
Wald Chi-Square (*df*: 3)	97.04							
P-Value	0.00							
	Estimate	SE	[95% Conf. Interval]					
Random-Effects (Participant)								
SD (constant)	1.36	0.12	1.14	1.62				
SD (residual)	1.09	0.05	1.00	1.19				
LR Test vs. Linear Model	146.84							
P-Value	0.00							

Table 25. MEM results for stress ratings

Stress	Coef.	SE	*z*	P>|*z*|	[95% Conf. Interval]		f^2	Bonferroni *p*
Fixed-Effects								
Task Closure: Open	0.63	0.34	1.86	0.06	-0.03	1.30	0.004	0.08
Task Complexity: Complex	1.12	0.19	5.92	0.00	0.75	1.49	0.241	0.00
Task Closure × Task Complexity	-0.17	0.27	-0.64	0.53	-0.69	0.35	0.002	
constant	2.94	0.24	12.15	0.00	2.46	2.46		
Log Likelihood	-606.24							
Wald Chi-Square (*df*: 3)	63.96							
P-Value	0.00							
	Estimate	SE	[95% Conf. Interval]					
Random-Effects (Participant)								
SD (constant)	1.29	0.12	1.07	1.55				
SD (residual)	1.21	0.05	1.11	1.33				
LR Test vs. Linear Model	109.08							
P-Value	0.00							

Table 26. MEM results for planning time pressure ratings

Planning time pressure	Coef.	SE	*z*	P>|*z*|	[95% Conf. Interval]		f^2	Bonferroni *p*
Fixed-Effects								
Task Closure: Open	0.36	0.31	1.19	0.23	-0.23	0.96	0.008	0.08
Task Complexity: Complex	0.57	0.15	3.74	0.00	0.27	0.87	0.176	0.00
Task Closure × Task Complexity	0.26	0.22	1.21	0.23	-0.16	0.68	0.006	
constant	2.23	0.22	10.27	0.00	1.81	2.66		
Log Likelihood	-545.95							
Wald Chi-Square (*df*: 3)	47.20							
P-Value	0.00							
	Estimate	SE	[95% Conf. Interval]					
Random-Effects (Participant)								
SD (constant)	1.21	0.11	1.01	1.44				
SD (residual)	0.98	0.04	0.90	1.07				
LR Test vs. Linear Model	143.21							
P-Value	0.00							

Table 27. MEM results for writing time pressure ratings

Writing time pressure	Coef.	SE	*z*	P>|*z*|	[95% Conf. Interval]		f^2	Bonferroni *p*
Fixed-Effects								
Task Closure: Open	0.18	0.38	0.48	0.64	-0.56	0.92	0.001	0.74
Task Complexity: Complex	1.04	0.19	5.59	0.00	0.67	1.40	0.224	0.00
Task Closure × Task Complexity	-0.12	0.26	-0.46	0.65	-0.63	0.39	0.001	
constant	2.93	0.27	10.84	0.00	2.40	3.46		
Log Likelihood	-611.80							
Wald Chi-Square (*df*: 3)	56.30							
P-Value	0.00							
	Estimate	SE	[95% Conf. Interval]					
Random-Effects (Participant)								
SD (constant)	1.51	0.14	1.27	1.80				
SD (residual)	1.19	0.05	1.09	1.30				
LR Test vs. Linear Model	152.41							
P-Value	0.00							

Table 28. MEM results for time-on-planning

Time-on-planning	Coef.	SE	*z*	P>|*z*|	[95% Conf. Interval]		f^2	Bonferroni *p*
Fixed-Effects								
Task Closure: Open	17.67	12.32	1.43	0.15	-6.47	41.82	0.018	0.78
Task Complexity: Complex	59.33	9.80	6.06	0.00	40.13	78.53	0.171	0.00
Task Closure × Task Complexity	-29.70	13.77	-2.16	0.03	-56.69	-2.71	0.017	
constant	60.22	8.76	6.87	0.00	43.04	77.40		
Log Likelihood	-1877.90							
Wald Chi-Square (*df*: 3)	46.14							
P-Value	0.00							
	Estimate	SE	[95% Conf. Interval]					
Random-Effects (Participant)								
SD (constant)	34.39	5.05	25.78	45.87				
SD (residual)	62.72	2.81	57.45	68.48				
LR Test vs. Linear Model	21.72							
P-Value	0.00							

Table 29. MEM results for time-on-writing

Time-on-writing	Coef.	SE	*z*	P>|*z*|	[95% Conf. Interval]		f^2	Bonferroni *p*
Fixed-Effects								
Task Closure: Open	40.57	49.99	0.81	0.42	-57.40	138.55	0.001	0.31
Task Complexity: Complex	146.05	24.56	5.95	0.00	97.92	194.18	0.313	0.00
Task Closure × Task Complexity	13.45	34.52	0.39	0.70	-54.21	81.11	0.001	
constant	402.93	35.56	11.33	0.00	333.23	472.62		
Log Likelihood	-2233.27							
Wald Chi-Square (*df*: 3)	79.60							
P-Value	0.00							
	Estimate	SE	[95% Conf. Interval]					
Random-Effects (Participant)								
SD (constant)	198.69	17.89	166.55	237.04				
SD (residual)	157.25	7.05	144.02	171.68				
LR Test vs. Linear Model	150.83							
P-Value	0.00							

Table 30. MEM results for time-on-whole task

Time-on-whole task	Coef.	SE	*z*	P>|*z*|	[95% Conf. Interval]		f^2	Bonferroni *p*
Fixed-Effects								
Task Closure: Open	58.25	54.11	1.08	0.28	-47.80	164.29	0.001	0.32
Task Complexity: Complex	205.38	27.94	7.35	0.00	150.62	260.14	0.400	0.00
Task Closure × Task Complexity	-16.25	39.28	-0.41	0.68	-93.23	60.74	0.001	
constant	463.15	38.49	12.03	0.00	387.71	538.58		
Log Likelihood	-2271.38							
Wald Chi-Square (*df*: 3)	101.95							
P-Value	0.00							
	Estimate	SE	[95% Conf. Interval]					
Random-Effects (Participant)								
SD (constant)	211.50	19.43	176.66	253.22				
SD (residual)	178.91	8.02	163.87	195.33				
LR Test vs. Linear Model	133.81							
P-Value	0.00							

5.2.2 Task Effects on L2 Writing

A series of MEMs were computed on participants' written output. Performance measures examined included syntactic complexity, linguistic diversity, and accuracy. Syntactic complexity was measured in terms of two aspects: 1) length of production—mean length of T-unit (MLT) and mean length of clause (MLC), and 2) subordination—number of subordinate clauses per T-unit. Guiraud's Index (GI) was used to measure lexical diversity. Accuracy was examined in terms of 1) the proportion of error-free structures—proportions of target-like use (TLU) of articles, error-free T-units, and error-free clauses, and 2) the proportion of lexical errors. It was hypothesized that all measures of syntactic complexity, the lexical diversity measure, and all measures of accuracy would increase with increased task complexity. Furthermore, closed tasks were hypothesized to elicit better performance in the three aspects than open tasks.

Results of the Wald Test on the models with MLT, MLC, and subordinate clause per T-unit as the dependent variable were non-significant, indicating that the explanatory variables were not significant, $\chi^2 = 8.00$, $p = .33$, $\chi^2 = 11.00$, $p = .14$, and $\chi^2 = 6.99$, $p = .43$, respectively. Because of the non-significant results, further analyses were not conducted on syntactic complexity measures.

When Guiraud's Index was included in the MEM as the dependent variable, results of the Wald Test were found to be significant, $\chi^2 = 120.85$, $p < .001$. The MEM yielded significant main effects of task complexity and task closure, such that complex tasks elicited 58.5% more diverse vocabulary than simple tasks ($p < .001$, $f^2 = .384$), lending support for the CH, and participants in the Open condition produced 49.0% more diverse vocabulary than those in the Closed condition ($p < .001$, $f^2 =$

.020). Although the task closure results conflict with the predictions of the study, they are in line with the cognitive load ratings and expert judgments, in that the open tasks placed a heavier cognitive load on learners than the closed tasks, resulting in greater linguistic complexity. Another interesting finding involves the significant positive relationship between WMC and lexical diversity. It was found that, as participants' WMC increased by 0.1 on the Ospan task (the score on which could range from zero to one), their GI increased by 0.056 (p = .05, f^2 = .004).

When a series of MEMs was computed for accuracy measures, the results of the Wald Test were significant for all measures at or below the .01 level: proportion of TLU articles (χ^2 = 41.32), proportion of error-free T-units, (χ^2 = 29.34), proportion of error-free clauses (χ^2 = 23.85), and proportion of lexical errors (χ^2 = 19.17). Although task closure was not found to have a significant effect on any accuracy measure, a significant main effect of task complexity on the proportion of TLU articles was observed in that complex tasks elicited a 3.7% lower ratio than simple ones (p = .01, f^2 = .049).

The MEMs conducted on accuracy measures also yielded interesting findings regarding aptitude for implicit learning, measured by LLAMA D, and English proficiency, measured by Brown's (1980) cloze test. Although it was not surprising that greater L2 proficiency would be associated with a higher proportion of TLU articles and a lower proportion of errors, only aptitude for implicit learning was found to have a positive relationship with accuracy measures. In the case of L2 proficiency, an increase in one point on the cloze test was related to positive changes in the proportion of TLU articles (.007 higher, $p < .001$, f^2 = .029), proportion of error-free

T-units (.011 higher, $p < .001$, $f^2 = .016$), error-free clauses (.010 higher, $p < .001$, $f^2 = .016$), and proportion of lexical errors (.001 lower, $p < .001$, $f^2 = .009$). An increase in one point on the LLAMA D subtest lead to changes in proportion of TLU articles (.006 higher, $p = .03$, $f^2 = .009$), proportion of error-free T-units (.016 higher, $p < .001$, $f^2 = .015$), proportion of error-free clauses (.015 higher, $p < .001$, $f^2 = .017$), and proportion of lexical errors (.002 lower, $p < .001$, $f^2 = .018$). On the other hand, LLAMA F scores were found to have a negative relationship with the proportion of error-free T-units: an increase in 1 point on the LLAMA subtest led to a decrease in the proportion of error-free T-units (.009 lower, $p = .03$, $f^2 = .007$). However, the magnitude of this effect was so small that it was considered negligible.

A summary of the results obtained from the series of MEMs on performance measures are shown in Tables 31-38.

Table 31. MEM results for MLT

MLT	Coef.	SE	*z*	P>|*z*|	[95% Conf.	Interval]	f^2	Bonferroni *p*
Fixed-Effects								
Task Closure: Open	-0.94	0.55	-1.71	0.09	-2.02	0.14		0.24
Task Complexity: Complex	-0.71	0.39	-1.85	0.07	-1.47	0.04		0.21
Task Closure × Task Complexity	0.75	0.54	1.38	0.17	-0.32	1.81		
LLAMA D	-0.02	0.08	-0.26	0.79	-0.17	0.13		
LLAMA F	0.00	0.07	-0.07	0.94	-0.14	0.13		
Ospan Task	1.12	0.96	1.16	0.25	-0.77	3.01		
Cloze test	0.07	0.06	1.27	0.21	-0.04	0.18		
constant	10.38	2.37	4.38	0.00	5.73	15.02		
Log Likelihood	-817.20							
Wald Chi-Square (*df*: 3)	8.00							
P-Value	0.33							
	Estimate	SE	[95% Conf. Interval]					

Random-Effects (Participant)								
SD (constant)	1.76	0.21	1.39	2.21				
SD (residual)	2.47	0.11	2.26	2.70				
LR Test vs. Linear Model	133.81							
P-Value	0.00							

Table 32. MEM results for MLC

MLC	Coef.	SE	*z*	P>|*z*|	[95% Conf. Interval]		f^2	Bonferroni *p*
Fixed-Effects								
Task Closure: Open	-0.57	0.24	-2.42	0.02	-1.03	-0.11		0.02
Task Complexity: Complex	-0.08	0.13	-0.61	0.54	-0.35	0.18		0.87
Task Closure × Task Complexity	0.13	0.19	0.70	0.48	-0.24	0.50		
LLAMA D	0.01	0.04	0.33	0.74	-0.06	0.08		
LLAMA F	0.04	0.03	1.47	0.14	-0.01	0.10		
Ospan Task	-0.14	0.43	-0.31	0.75	-0.99	0.71		
Cloze test	0.03	0.02	1.13	0.26	-0.02	0.08		
constant	5.45	1.06	5.13	0.00	3.37	7.54		
Log Likelihood	-488.95							
Wald Chi-Square (*df*: 3)	11.00							
P-Value	0.14							
	Estimate	SE	[95% Conf. Interval]					
Random-Effects (Participant)								
SD (constant)	0.86	0.08	0.71	1.05				
SD (residual)	0.86	0.04	0.79	0.94				
LR Test vs. Linear Model	96.71							
P-Value	0.00							

Table 33. MEM results for number of subordinate clauses per T-unit

Subordinate clauses per T-unit	Coef.	SE	*z*	P>|*z*|	[95% Conf. Interval]		f^2	Bonferroni *p*
Fixed-Effects								
Task Closure: Open	-0.01	0.09	-0.13	0.90	-0.18	0.16		0.63
Task Complexity: Complex	-0.10	0.06	-1.49	0.14	-0.22	0.03		0.29
Task Closure × Task Complexity	0.10	0.09	1.06	0.29	-0.08	0.27		

LLAMA D	0.00	0.01	-0.39	0.69	-0.03	0.02
LLAMA F	-0.01	0.01	-1.28	0.20	-0.03	0.01
Ospan Task	0.21	0.15	1.38	0.17	-0.09	0.51
Cloze test	0.01	0.01	0.83	0.41	-0.01	0.02
constant	0.85	0.37	2.29	0.02	0.12	1.58
Log Likelihood	-216.51					
Wald Chi-Square (*df*: 3)	6.99					
P-Value	0.43					
	Estimate	SE	[95% Conf. Interval]			
Random-Effects (Participant)						
SD (constant)	0.27	0.03	0.21			
SD (residual)	0.41	0.02	0.38			
LR Test vs. Linear Model	35.19					
P-Value	0.00					

Table 34. MEM results for GI

GI	Coef.	SE	*z*	P>|*z*|	[95% Conf. Interval]		f^2	Bonferroni *p*
Fixed-Effects								
Task Closure: Open	0.49	0.16	3.14	0.00	0.18	0.80	0.020	0.00
Task Complexity: Complex	0.59	0.10	6.15	0.00	0.40	0.77	0.384	0.00
Task Closure × Task Complexity	0.15	0.13	1.09	0.28	-0.12	0.41	0.005	
LLAMA D	-0.02	0.02	-0.90	0.37	-0.07	0.02		
LLAMA F	0.01	0.02	0.39	0.69	-0.03	0.05		
Ospan Task	0.56	0.28	1.97	0.05	0.00	1.12	0.004	
Cloze test	0.01	0.02	0.84	0.40	-0.02	0.05		
constant	5.23	0.70	7.51	0.00	3.87	6.60		
Log Likelihood	-367.24							
Wald Chi-Square (*df*: 3)	120.85							
P-Value	0.00							
	Estimate	SE	[95% Conf. Interval]					
Random-Effects (Participant)								
SD (constant)	0.55	0.06	0.45	0.68				
SD (residual)	0.61	0.03	0.56	0.67				
LR Test vs. Linear Model	78.96							
P-Value	0.00							

Table 35. MEM results for proportion of TLU articles

Proportion of TLU articles	Coef.	SE	*z*	P>|*z*|	[95% Conf. Interval]		f^2	Bonferroni *p*
Fixed-Effects								
Task Closure: Open	-0.03	0.02	-1.39	0.17	-0.06	0.01	0.005	0.08
Task Complexity: Complex	-0.04	0.02	-2.49	0.01	-0.07	-0.01	0.049	0.00
Task Closure × Task Complexity	0.00	0.02	-0.07	0.95	-0.04	0.04	0.000	
LLAMA D	0.01	0.00	2.21	0.03	0.00	0.01	0.009	
LLAMA F	0.00	0.00	-1.14	0.25	-0.01	0.00		
Ospan Task	0.02	0.03	0.68	0.50	-0.04	0.08		
Cloze test	0.01	0.00	3.99	0.00	0.00	0.01	0.029	
constant	0.68	0.08	8.83	0.00	0.53	0.83		
Log Likelihood	275.42							
Wald Chi-Square (*df*: 3)	41.32							
P-Value	0.00							
	Estimate	SE	[95% Conf. Interval]					
Random-Effects (Participant)								
SD (constant)	0.05	0.01	0.04	0.07				
SD (residual)	0.10	0.00	0.09	0.11				
LR Test vs. Linear Model	18.88							
P-Value	0.00							

Table 36. MEM results for proportion of error-free T-units

Proportion of error-free T-units	Coef.	SE	*z*	P>|*z*|	[95% Conf. Interval]		f^2	Bonferroni *p*
Fixed-Effects								
Task Closure: Open	0.06	0.04	1.75	0.08	-0.01	0.13	0.015	0.38
Task Complexity: Complex	0.05	0.03	1.90	0.06	0.00	0.10	0.016	0.43
Task Closure × Task Complexity	-0.07	0.04	-1.92	0.06	-0.14	0.00	0.014	
LLAMA D	0.02	0.00	3.23	0.00	0.01	0.03	0.015	
LLAMA F	-0.01	0.00	-2.16	0.03	-0.02	0.00	0.007	
Ospan Task	0.03	0.06	0.43	0.67	-0.09	0.14		
Cloze test	0.01	0.00	3.33	0.00	0.00	0.02	0.016	
constant	0.08	0.15	0.56	0.58	-0.21	0.37		
Log Likelihood	86.32							
Wald Chi-Square (*df*: 3)	29.34							
P-Value	0.00							

	Estimate	SE	[95% Conf. Interval]	
Random-Effects (Participant)				
SD (constant)	0.11	0.01	0.08	0.14
SD (residual)	0.17	0.01	0.15	0.18
LR Test vs. Linear Model	33.82			
P-Value	0.00			

Table 37. MEM results for proportion of error-free clauses

Proportion of error-free clauses	Coef.	SE	*z*	P>|*z*|	[95% Conf. Interval]		f^2	Bonferroni *p*
Fixed-Effects								
Task Closure: Open	0.05	0.04	1.49	0.14	-0.02	0.12	0.006	0.28
Task Complexity: Complex	0.04	0.03	1.45	0.15	-0.01	0.09	0.008	0.06
Task Closure × Task Complexity	-0.04	0.04	-1.08	0.28	-0.12	0.03	0.004	
LLAMA D	0.02	0.00	3.17	0.00	0.01	0.02	0.017	
LLAMA F	-0.01	0.00	-1.61	0.11	-0.01	0.00		
Ospan Task	-0.01	0.06	-0.12	0.91	-0.12	0.11		
Cloze test	0.01	0.00	3.08	0.00	0.00	0.02	0.016	
constant	0.03	0.15	0.21	0.83	-0.26	0.32		
Log Likelihood	72.25							
Wald Chi-Square (*df*: 3)	23.85							
P-Value	0.00							
	Estimate	SE	[95% Conf. Interval]					
Random-Effects (Participant)								
SD (constant)	0.10	0.01	0.08	0.13				
SD (residual)	0.18	0.01	0.16	0.19				
LR Test vs. Linear Model	23.89							
P-Value	0.00							

Table 38. MEM results for proportion of lexical errors

Proportion of lexical errors	Coef.	SE	*z*	P>|*z*|	[95% Conf. Interval]		f^2	Bonferroni *p*
Fixed-Effects								
Task Closure: Open	0.00	0.00	-0.48	0.63	-0.01	0.01	0.001	0.54
Task Complexity: Complex	0.00	0.00	-0.72	0.48	-0.01	0.00	0.005	0.28

Task Closure × Task Complexity	0.00	0.00	-0.07	0.94	-0.01	0.01	0.000
LLAMA D	0.00	0.00	-3.31	0.00	0.00	0.00	0.018
LLAMA F	0.00	0.00	0.56	0.58	0.00	0.00	
Ospan Task	0.00	0.01	0.50	0.62	-0.01	0.02	
Cloze test	0.00	0.00	-2.39	0.02	0.00	0.00	0.009
constant	0.07	0.02	4.38	0.00	0.04	0.10	
Log Likelihood	832.84						
Wald Chi-Square (*df*: 3)	19.17						
P-Value	0.00						
	Estimate	SE	[95% Conf. Interval]				
Random-Effects (Participant)							
SD (constant)	0.01	0.00	0.01	0.01			
SD (residual)	0.02	0.00	0.02	0.02			
LR Test vs. Linear Model	26.70						
P-Value	0.00						

5.2.3 Summary of MEM Results

Results obtained from of a series of MEMs conducted on cognitive load measures revealed positive task complexity effects, such that greater task complexity led to higher learner self-ratings and more time-on-task. A significant interaction effect between task complexity and task closure was also found on time-on-planning. In comparison to the participants in the Open condition, those in the Closed condition spent *less* time on planning for the *simple* task versions, and spent *more* time on the *complex* versions. In other words, task complexity had a greater impact on those in the Closed condition than on those in the Open condition.

When a series of MEMs was conducted on performance measures, task complexity and task closure were not found to have significant effects on syntactic complexity in terms of MLT, MLC, and the number of subordinate clauses per T-unit.

On the other hand, they were found to have a significant positive impact on lexical diversity. WMC was also found to have a positive relationship with lexical diversity. In the case of accuracy, measured by proportions of TLU articles, error-free T-units, error-free clauses, and lexical errors, while task closure did not have a significant effect on any of the accuracy measures, task complexity was found to have a significant negative effect on the proportion of TLU articles. That is, complex task versions elicited a lower proportion of TLU articles. It was found that that ILA and L2 proficiency had a significant positive relationship with accuracy, and ELA had a significant negative, yet negligible, relationship with the proportion of error-free T-units.

Chapter 6. Discussion and Conclusions

The present study examined the interactive effects of task complexity and task closure on the written performance of L2 learners. Another goal was to obtain independent evidence that the task versions intended to be more complex led to the expected changes in cognitive load. The effects of cognitive individual differences—working memory capacity (WMC) and aptitudes for explicit and implicit learning—were also investigated.

6.1 Task Complexity Effects on Cognitive Load

In the present study, task complexity was manipulated in terms of the number of elements, a resource-directing dimension in Robinson's CH. Two tasks were employed, in which participants were required to choose the best location (hotel or venue) for a certain event. The number of locations to choose from, the number of categories, and the number of requirements in the closed tasks were manipulated. It was hypothesized that the task versions involving more elements would place greater cognitive load on the learners.

Cognitive load was assessed via three measures: learner self-ratings, expert judgments, and time-on-task. Hypotheses of the study were sustained, as results obtained from all measures showed that there were significant differences between the simple and complex versions. In the case of learner self-ratings, participants indicated that they considered the complex tasks to be more difficult and stressful and to require greater mental effort. They also felt more pressed for time during the planning and writing stages of complex tasks. Overall, learner self-ratings were able to capture task complexity effects on cognitive load. Expert judgments regarding ratings of overall difficulty and mental effort required yielded similar results, although inferential statistics were not computed due to the small *N*-size. That is, experts also believed that the difference between the two mainly lies in the number of elements involved. They also mentioned that, due to the differences in number of elements, the complex versions require more prioritization of categories/information, such that they need to focus more on which categories they place greater importance on when choosing a location.

The present study is the first to employ time-on-task as a measure of cognitive load in task-based research. Time engaged in the planning and writing stages and task as a whole were recorded, and it was hypothesized that they would increase along with increases in task complexity. Such predictions were borne out, as task complexity was found to affect all three measures of time-on-task positively. More interestingly, a significant interaction between task complexity and task closure was found on time-on-planning. A closer inspection revealed that while planning times for participants in both Open and Closed conditions were affected when they were carrying out complex tasks, task complexity had a greater impact in the Closed condition. The complex closed task required more time for participants to figure out the solution than the complex open task, most likely because those in the Closed condition had to compare and contrast the information provided to find the solution before being able to write. Once participants in both conditions found the/a solution, differences between the two conditions disappeared, and only task complexity effects remained. This finding runs counter to one of the expert's claim that the existence of requirements in closed tasks would obviate the need for a planning stage because participants do not need to think too hard. On the contrary, time-on-planning findings show that a planning stage is necessary, when performing a complex closed task.

To summarize, the findings obtained from learner self-ratings, expert judgments, and time-on-task all point to the effects of task complexity, such that a greater number of elements involved in a task places a greater cognitive processing load. Furthermore, time-on-task was found to be a valid measure of cognitive load.

6.2 Task Complexity Effects on L2 Written Performance

The major goal of TBLT and task-based L2 research is for learners to perform tasks successfully while simultaneously developing their interlanguage. As standard measures in this field, syntactic complexity, grammatical accuracy, and fluency (CAF) have been used to assess task performance. In the present study, L2 written performance was examined in terms of syntactic complexity, lexical diversity, and accuracy. It was hypothesized that increases in task complexity would lead to greater syntactic complexity, lexical diversity, and accuracy. However, only the hypothesis regarding lexical diversity was confirmed.

Syntactic complexity was examined in two ways, length of production and subordination. Length of production was measured via mean length of T-unit (MLT), mean length of clause (MLC). The number of subordinate clauses per T-unit was used to measure subordination. A series of mixed effects models (MEMs) was conducted, but no significant effects for task complexity were found. On closer inspection of the written data, participants did not seem to produce longer sentences or more subordinate clauses when performing complex tasks. Table 39 provides writing samples from two participants, one in the Open condition, the other in the Closed condition. Based on the samples, the non-significant effects of task complexity on linguistic complexity could be attributed to the following: 1) failure to give *detailed* explanations, 2) failure to provide reasons why they chose a location or did *not* choose the other locations, 3) use of similar structures for both the simple and complex versions, and 4) less use of subordinate clauses. For instance, when the participant in the Open condition carried out the complex Hotel task, it was only briefly mentioned that the hotel they chose "has two single-sized beds and a long check out time." This sentence alone contains two elements, each of which could have been discussed in more detail. Furthermore, the same participant claimed that "The Lighthouse and The Lunchroom should be not included" when performing the complex Venue task, but did not fully explain why neither of them would work. In case of the participant in the Closed condition, more attention was paid to detail, and all elements were mentioned. However, the paragraph and sentence structures were similar on the simple and complex tasks, resulting in a very small difference in syntactic complexity between the simple and complex versions. Furthermore, many participants used constructions without the need for subordinate clauses, or they used coordinating conjunctions, such as *and*, *but*, and *so*, e.g., "Second, they want calm atmosphere *but* 'The Spring' have live band session, *so* its atmosphere is deleted." Even though this example provides a reason for not choosing "The Springs," it does not include a subordinate clause. Note that the present study used proportions and not sums of units to measure syntactic complexity. Therefore, participants would have had to use more subordinated structures for each element in order for task complexity effects to be significant.

Table 39. Writing samples to compare syntactic complexity

Condition	Task	Simple	Complex
Open	Hotel	Considering the daily rate, The Utopia might be the best place to stay. However, Internet access is really bad in Utopia. These days, people use smartphone a lot. Especially when they travel, they take lots of picture or selfie and arrange them at night in their hotel. And before they go to sleep they use a google map for searching a street or mall they will visit the next day. For all of these convenience, people need high quality of Internet access. And If they stay in Utopia they might need to pay a lot of money to use the Internet. And in case of The Sunset, even if the internet access is cheaper the Utopia, the total daily rate is too high. Therefore, It is not best choice. Therefore, I recommend The Moonriver, because it has proper daily rate and the internet access is free. Finally I recommend The Moonriver.	I think The Utopia is the best place to stay in New York. It has several proper options unlike others. First of all, it has pretty good daily rate. But Imperial has too expensive rate. So it cannot be a good choice. Second, Utopia has free wifi which is the most important thing in modern society. While traveling, we use lots of internet to search a map or get information about tourists. But if the internet is impossible or the fee is expensive, there will be a lot of restrictions to travel. Therefore, The Urban and Echo is not a good choice. Third, The Utopia has proper daily rate and free wifi. Also, it has two single sized bed and long check out time. So, I highly suggest the Utopia is the best option.
	Venue	I think The castle could be the best place to give a party for their mother. Regardless of age, sex and race, every person likes to have music on their birthday. And music which is the most important part cannot be used in The square. Therefore, The square is not a best choice. As their mother's age is 70, she might have a lot of friends or families to invite. But If they choose The Royal to throw the party, there will be a lack of chairs or tables that guests can use. For these reasons, The Royal is also a bad option. Finally, In case of castle, the price is	I strongly recommend that The Tower is the best place to throw their father's party. Firstly, as the party is for their father's birthday, several music and videos are requisite for rising the mood. So The Lighthouse and The Lunchroom should be not included. Secondly, even if The Tower is more expensive than The Springs, the parking pay is free for The tower. Considering the guests' number, Choosing The Tower for the party is much more efficient. Therefore, for these reasons, I highly recommend The Tower.

		reasonable and the accommodation number is appropriate. Besides, it has proper music to use at the party. So, I think The Castle is the best option.	
Closed	Hotel	In my opinion, Carry and Amanda should reserve the Moonriver hotel for 3 reasons below. First, The Moonriver hotel has a free wi-fi. However, if they use other hotels they should pay for it to use. Second, The Moonriver hotel has affordable price. However, other hotels are more expensive than the norm they choose. Third, The Moonriver hotel is close enough to the beach to use bus. As they wanted the hotel where can find public bus or subway nearby, this hotel fits with the conditions they wanted. For these reasons, I recommend the Moonriver hotel to Carry and Amanda.	I recommend John and Peter to make a reservation at The utopia hotel for the reasons below. First, The Urban hotel is cheap, but internet is not available. Also this hotel wants customers to check out by 11 o'clock, so it doesn't fit with the condition John and Peter picked. Secondly, The Eco hotel is affordable to pay, but it's too far from the public stations. If they choose this hotel, they should take a public transportation even they don't want. Also, this hotel wants customers to check out by 11 o'clock, too. Lastly, The imperial hotel is too expensive than the budget they selected. Also they should pay for using the internet. However, The utopia hotel fits in every condition which John and Peter wanted. For these reasons, I recommend the utopia hotel to them.
	Venue	In my opinion, I recommend Brian and Kate to reserve the Castle restaurant for the reasons below. First, the square restaurant is more expensive than the budget they expected. Also it cannot afford more than 30 people in a room. Secondly, The Royal restaurant is also more expensive than the budget they expected, which is $50. This restaurant can only afford 20 people in a room, so it doesn't fit with the condition they were looking for. Also, they were finding the quiet place, but this restaurant plays the live session music which could be noisy. However, the Castle perfectly fits with all the condition they wanted. For these reasons, I recommend the Castle to them.	In my opinion, I would recommend Liam and Kate to reserve the Lunchroom for the reasons below. All four restaurants are affordable and fit with the budget they expected. Also they all have parking places nearby and give at least 3 hours time to the customers. But the background music of the tower and the springs restaurant is not quiet, so they don't fit with the condition which Liam and Kate wanted. Moreover, the Lighthouse cannot afford more than 20 people in a room even Liam and Kate want to invite more than 30 people. Also, the Lighthouse and the springs cannot use the projector and screen, so they are not the place where Liam and Kate are looking for. For these reasons, the Lunchroom is the best place for Liam and Kate to open a celebration.

Note. Underlined words indicate subordinate clauses.

Task complexity was manipulated in terms of number of elements. In addition to the elements included in the simple versions, the complex versions added one more location option (4 vs. 3), three more categories that contained more information about location characteristics (6 vs. 3), and three more requirements in the closed tasks that needed to be considered in the Closed condition (6 vs. 3). If participants carried out the tasks successfully, by covering every category and providing reasons for why they chose a certain location and why they did *not* choose the others, they would naturally use a greater variety of words in their writing. Unlike the findings regarding syntactic complexity, significant effects of task complexity were found on lexical diversity, measured via Guiraud's Index (GI). That is, the complex task versions, compared with the simple ones, elicited a wider range of vocabulary, which was in line with predictions.

On the other hand, almost null findings were obtained for accuracy measures. While there was a significant main effect of task complexity on the proportion of TLU articles, the direction ran counter to the CH, in that participants made more errors in terms of TLU articles on the complex tasks. In order to see what caused such results, a closer examination of the written data was required. Table 40 provides examples of participants who showed less accurate use of articles in complex tasks. Some participants correctly used an article in front of a particular noun in the simple version, but incorrectly used it in the complex version, e.g., "the internet," "live music," and "public transportation." In many cases, the decision to use "a," "the," or no article was determined by context, and the inconsistent/erroneous use of articles clearly indicates that participants did not have full knowledge of articles. However, articles are one of the most difficult grammatical structures to master fully, and it was obvious that participants' proficiency level was not advanced enough. Because complex tasks required participants to discuss more elements in the form of noun phrases, they were bound to make more errors in article use.

Table 40. Writing samples to compare accuracy

Condition	Task	Simple	Complex
Open	Hotel	I suggest The Sunset hotel to them because of a few reasons. In my experience it is so bothering and costs some money to take public transportation abroad. In this way, I excluded The Moonriver hotel which can reach a beach by bus, but the Sunset hotel needs only 5 minutes to the beach on foot. Secondly, the internet which is necessary costs a lot in The Utopia hotel even if there are <u>25$ gap</u> between them. I guess the internet fee in The Utopia hotel which is 5$ per magabyte would overpass the gap easily."	I suggest The Urban hotel. In my opinion, I value <u>the access</u> to <u>the public transportation</u> the most during <u>trip</u>. In this way, The Urban and The Imperial seems to have <u>nice location</u>. However, the daily rate in The Imperial is double of The Urban's which is critical gap. Moreover, in The Urban you can save lots of money because it includes breakfast which is not included in The Utopia and in The Imperial. It also provides you with two double size beds which are appropriate for you. However, I feel sorry that it does not have any internet access and the check out time is a bit earlier than the other hotels which is 11 a.m, but you can use <u>internet</u> in <u>the cafe</u> and I don't think it is important to have late check-out time, also it is not that late time.
	Venue	I recommend to go 'The Castle' for your mother's 70th anniversary. Because The price is cheaper than <u>other's two restaurants</u>, and also it has calm music. So it will be great for your mother and relatives enjoy the party. Also, the accommodation of 'The Castle' is 40, which will be enough to accommodate your all relatives. However, only 20 people can enter to 'The Royal' and also it is more expensive than 'The Castle'. And 'The Square' is the highest price among these restaurants and there is no music. Therefore, 'The Castle' is appropriate for your mother's anniversary.	I recommend you to choose 'The Springs' restaurant for three reasons. First, the price is very reasonable. Compared to the other restaurants, 'The Springs' is much cheaper than <u>others</u>. Second, there is <u>the performance of live band</u>, so all relatives and your father can enjoy the party. Third, there is no limit of time. If you hold a party, unexpected accidents can happen. So the party can be delayed than you expected. Therefore, I think it is better to choose <u>the restaurant</u> [that you can use freely. And also, 'The Springs' has <u>the reasonable price</u> for <u>parking lot</u>. Other restaurants like 'The

			Lunchroom' and 'The Lighthouse' have the parking lot but it should be paid 3 dollars for one hour. So, I think one dollar for a hour is quite cheap price for parking lot. 'The Springs' has no equipment like beam projector and screen but I think you don't need any equipment for your father's 80th anniversary. Because the most important thing in that kinds of anniversary is communication between relatives and your family.
Closed	Hotel	At the point of daily rate, Carrey and Amanda want under $100. In this condition, 'The Moonriver' and 'The Utopia' fit it. Next, they want to get free Internet access at the hotel. Among the two hotel, 'The Moonriver' hotel provides free wifi. The hotel, 'The Utopia', give guests Internet access for a megabyte per five dollars. Therefore, Carrey and Amanda would choose 'The Moonriver' hotel to fit their conditions.	John and Peter would like [to go 'The Utopia' hotel for their fantastic trip. The first condition is their budget. they can spend 250 dollars per one day, so they cannot go 'The imperial' hotel. Next, their use of the amount of Internet is large, so 'The Urban', which do not give the access of Internet, would be kicked out. Next, they want to check out in afternoon, but 'The Echo' hotel makes guest check out until 11 o'clock. Considering these things, the best hotel for them would be 'The Utopia'.
	Venue	Are you looking for the restaurant suitable to your purpose? I think The Castle would be the best choice. Its accommodation is okay until 40 people, while the others, The Square and The Royal can hold 30 and 20 people respectively. Also, silent music will not interfere with the mood of your party. All you need is just pay 45 bills for the meal, which is cheaper than your expected budget.	_Hello, Liam, and Kate. I know you have a problem with choosing restaurant for the party. Let me [help you. I think The Lunchroom would be the best choice. Its atmosphere is silent enough while the other restaurant have kind of music. There is parking space, too. Well, however it takes two buck every hour. But, its fee is just 55 dollars, which is much cheaper than expected budget for coming people. You can bring more than 30 guys in it. And It has equipment for your need such as beam projector and screen. It has 3 hours limitation, but It doesn't matter, does it? So, are you still agonizing? just choose The Lunchroom!

Note. Underlined words indicate errors in TLU articles.

6.3 Task Condition Effects on Cognitive Load

In the present study, task condition was manipulated in terms of task closure, i.e., whether a task had a single or finite set of solutions (closed) or allowed a wide set of possible solutions (open). Unlike open tasks, closed tasks included a number of requirements for participants to consider when making a decision and writing about it. Because restrictions on the number of acceptable solutions force learners to pursue the task, claims have been made that closed tasks are more beneficial than open tasks for the promotion of negotiation of meaning, and greater provision and incorporation of feedback. Although such claims were made with regard to interactive tasks, the present study assumed that the beneficial effects of interactive closed tasks would also apply to monologic tasks. Accordingly, it was hypothesized that the active search for the correct solution would result in a greater cognitive load placed on the learner.

Results obtained from a series of MEMs did not reveal any significant main effects of task closure on cognitive load. However, task closure was found to have a significant interaction effect with task complexity on time-on-planning. In other words, hypotheses were mostly rejected because there were no significant differences between the Open and Closed conditions in terms of 1) learner self-ratings of overall difficulty, mental effort required, stress, and time pressure felt during the planning and writing stages, and 2) time engaged in the writing stage and the task as a whole. Contrary to predictions, the descriptive statistics for these measures showed the opposite pattern, pointing towards the notion that open tasks actually require more mental processing than closed tasks. Even ratings from experts on the overall difficulty and mental effort required supported this trend. One unexpected finding was the claim that the existence of requirements makes the planning stage obsolete, because it eliminates the process of having to work out a solution. Furthermore, it was also mentioned that unlike closed tasks, open tasks require the ability to adapt, synthesize, and infer information. Because participants are aware that there is only one solution, they merely have to compare and contrast the information provided minimally to satisfy the task requirements. Furthermore, they can even directly copy/translate the requirements into the L2 when giving reasons for why they choose a certain location. On the other hand, in the Open condition, participants must compare and contrast information thoroughly, in order to work out the best solution and be persuasive enough for the imaginary reader, whose preferences are unknown. In other words, closed tasks require convergent thinking, a straightforward process that involves solving a problem with a single, correct answer. One expert also stated that learners are more familiar with the format of closed tasks, because they are exposed to these kinds of tasks since high school. In contrast, open tasks involve convergent thinking, which involves exploring multiple unique and creative solutions to problems or ideas. Therefore, open tasks are claimed to require a higher degree of mental processing.

As mentioned earlier in Section 6.1, the expert's claim that requirements in closed tasks obviated the need for a planning stage was challenged by the significant interaction between task complexity and task closure on time-on-planning. Increases in task complexity affected both Open and Closed conditions, but had a greater impact on the latter. Although participants in the Closed condition spent approximately one minute on planning for simple tasks, those in the Open condition they spent 72-84 seconds. On the other hand, those in the Closed condition spent a longer time on planning for complex task versions than those in the Open condition (105-134 seconds vs. 104-111 seconds, respectively). Put simply, in the Closed condition, the task required more thought in the planning stage when it involved consideration of more elements. However, this was the only finding that was in line with the predictions of the study; all other trends in ratings pointed to the opposing argument, that open tasks may be more cognitively challenging than closed tasks.

Because a trend was found in learner self-ratings and expert judgments for open tasks to be more cognitively challenging than closed tasks, it was necessary to make a comparison between the differences in the task performance processes between the two conditions (see Figure 14). Open and closed tasks share similar processes, but the biggest difference lies in the basis for a decision: do participants have to meet other people's requirements, or can they make a decision based on their own preferences and/or personal experiences? For closed tasks, all that participants were required to do was find a hotel that minimally met all listed criteria. For instance, when one of the requirements was that they must spend under $100/night for a hotel room and the three hotel options are priced at $100/night, $120/night, and $95/night, the participant only needed to eliminate the second hotel from the list. However, those in the Open condition did not have any basis on which to make the decision, so participants had to examine and prioritize all categories to make their final decision.

Open condition

1. Read and understand hotel/venue characteristics
2. Compare/contrast location by category
3. Prioritize category
4. Narrow down choices by personal preference
5. Choose best hotel/venue
6. Access L2 lemmas
7. Formulate L2 structures
8. State opinion on best location
9. Provide supporting argument for choice based on personal preference

Closed condition

1. Read and understand requirements
2. Read and understand hotel/venue characteristics
3. Compare/contrast location by category
4. Narrow down choices by minimum requirement fulfillment
5. Choose best hotel/venue
6. Access L2 lemmas
7. Formulate L2 structures
8. State opinion on best location
9. Provide supporting argument for choice based on requirement fulfillment

Figure 14. Different task performance processes between Open and Closed conditions

6.4 Task Condition Effects on L2 Written Performance

The present study hypothesized that closed task versions would elicit greater syntactic complexity, lexical diversity, and accuracy than open task versions. The Wald Test for the MEMs conducted on syntactic complexity measures did not yield significant results, indicating that the explanatory variables in the models were not significant. Thus, the hypotheses regarding syntactic complexity were rejected, as no significant effects for task complexity were found for task closure. Hypotheses regarding task closure effects on accuracy were also rejected, as no significant effects were found.

Despite the null findings regarding task closure effects on syntactic complexity and accuracy, the descriptive statistics for MLC showed a trend for closed task versions to elicit greater MLC than open versions. Therefore, a closer look at participants MLC was warranted. When averaged across tasks and task complexity versions, participants in the Closed condition used 7.19 words per clause, and those in the Open condition, 6.55 words per clause. In other words, those in the Closed condition had a tendency to produce slightly longer clauses.

Table 41 provides writing samples from two participants in each condition. For comparison purposes, the participant in the Open condition was chosen for their overall low MLC across tasks, and the one in the Closed condition for their overall high MLC across tasks. In general, the main clauses and subordinate clauses in the Open condition examples consist of fewer words, e.g., “The Urban is the cheapest,” “That is so unacceptable,” “The price is reasonable” “Our luggage is ENORMOUS,” and “The price is fine.” Even though the participant gave detailed reasons for his choice of hotel, the clauses and subordinate clauses are relatively short. On the other hand, the participant in the Closed condition produced longer sentences, especially in terms of coordinated verbal phrases to provide reasons for his decision, e.g., “The reason is that it can fit with the budget within 250 dollars per day, provide free-Wifi, have 2 single beds, check out over 12 p.m,” It is worth mentioning that although this participant did not use many subordinate clauses in their writing, they were very thorough in providing detailed reasons for why they did *not* choose other locations. Two possibilities behind the differences in the MLC between the two groups can be suggested: 1) because participants in the Closed condition only needed to meet the minimum requirements to complete the task successfully, their attentional resources were freed up once they found the solution, and it was possible for them to use those resources on other aspects of language, such as sentence structure, or 2) when performing tasks, the Closed group could make use of the requirements listed in front of them and translate them into sentences during the writing stage. This simple translation would have allowed them to write sentences easily, which could have resulted in longer clauses.

Table 41. Writing samples to compare MLC

Condition	Task	Complex Task Example
Open	Hotel	Hey guys, still struggling with hotel selection right? I've searched for specific details of those hotels. The Urban is the cheapest but we cannot use the internet. That is so unacceptable... The next cheapest is The Utopia. The price is reasonable but it takes 30 minutes walk to get there. I don't think we can go there for a walk. Our luggage is ENORMOUS. The Imperial is the most reachable. It takes only 5 minutes walk. But it is too expensive for us and it also don't serve breakfast. UNACCEPTABLE!!! So, why don't we go to The Echo. The price is fine, internet charging is affordable, includes breakfast, bed is fine. It has one problem though, it takes 50 minutes walk. I'll pay this when we go, and why don't you pay when we go to the airport? Deal?
Closed	Hotel	I would recommend ""The Utopia"" for them. The reason is that it can fit with the budget within 250 dollars per day, provide free-Wifi, have 2 single beds, check out over 12 p.m. On the other hand, I would not recommend the other three hotels for them. First, as for The Urban, it cannot use Internet and check out before 11 a.m. Next, for The Echo, it satisfies all conditions, but the budget is overused because of the Internet fee. They tend to use Internet a lot. Finally, in terms of The Imperial, it cannot fit with the budget.

Note. Underlined words indicate subordinate clauses.

In terms of lexical diversity measured by GI, a significant main effect of task closure was found, such that participants in the Open condition produced significantly more diverse vocabulary than those in the Closed condition. Running counter to the predictions of the study, the hypothesis regarding lexical diversity was rejected. While those in the Closed condition only needed to discuss how the requirements were met or not met, most likely because they were aware that the tasks had only one answer, those in the Open condition had no restrictions on the contents of the writing. Therefore, participants discussed their reasons in much more detail, in order for the supporting arguments to be more convincing, even though the gist of the arguments in the two conditions were similar. For instance, in the examples in Table 42, the participant in the Open condition uses different words to express the same meaning, or uses more examples to support their decision, e.g., “smartphone necessary generation,” “internet connection problem,” “emergent connection,” “tired and annoying,” “Starbucks or any other public place,” etc. On the other hand, while the instructions explicitly stated that they should provide detailed explanations, the participant in the Closed condition repeatedly used the same words and structures in their writing. As mentioned by one of the experts, closed tasks involved convergent thinking, and open tasks divergent thinking, which may have resulted in a significant difference in lexical diversity between the two conditions.

Table 42. Writing samples to compare GI

Condition	Task	Complex Task Example

Open	Hotel	John and Peter. I considered a lot and I want to recommend The Urban Hotel for your trip. Before I tell you why, I'll show you the advantage and disadvantage of each hotel and evaluate it. First, The Urban, which I recommend, has no internet available. Of course it's very weak point in this smartphone necessary generation. However, it is located quite near [to get bus or subway. Also the daily rate is the cheapest even it supplies two double size bed and breakfast. And check-out time is not so important because there is only 3 hours gap and you'll not be in the hotel again after you have lunch. Because there are no hotel which give lunch too, that 3 hours have no meaningful value. You'll might spend just for walking between the restaurant and hotel again. Then isn't it reasonable that just check out and take a lunch and keep travel or come back home? I think the only disadvantage of The Urban is the internet connection problem. But you can use freely on Starbucks or any other public place and then come back to the hotel. If you need emergent connection, then you can just move only a little more time than fifteen minutes to access the internet. Then let's examine others. The Utopia has free internet, but it has the most narrow size of beds and they don't give you breakfast. Also it spends thirty minutes to transportation so you might totally think that 'Is this hotel only for sleep and use Internet?' and feel not so good about it. The Echo seems not so different. Because it has the internet but you have to pay $2 for every single megabyte you use. Also it is the most far hotel from the transportation so it might be very tired and annoying for moving. So the size of bed and breakfast might not be an advantage anymore. How about The Imperial? Wow, It costs the most. Also it requires the most fee for the internet. Even it doesn't offer a meal. Although it has two queen size of bed, I think this is too much spending for a travel. Therefore finally I concluded that you had better choose The Urban!
Closed	Hotel	I recommend the Utopia by the following reasons. It needs 200 dollars per day and it supplies free wifi. Also, it contains two single sized beds. You can check-out by 2 p.m. I do not recommend the rest of t and three hotels, the Urban, the Echo, the Imperial. In the case of the Urban, you cannot use the internet and you have to check-out by 11 a.m In the Echo's case, you need to pay two dollars per megabyte when you use the internet and check-out by 11 a.m. The Imperial requires 300 dollars per day, and you should pay 5 dollars per megabyte. Therefore, I recommend the Utopia.

6.5 Working Memory Capacity and L2 Written Performance

Borrowing from the work of Kellogg (1999), the present study assumed that writing processes involved active working memory (WM) components, especially during the formulation and monitoring stages, in which sentences are planned, generated, and edited. Because increases in task complexity involves a greater cognitive load, a positive relationship between working memory capacity (WMC) and task performance was hypothesized, such that those with higher WMC would have sufficient attentional resources to focus on linguistic complexity and accuracy simultaneously. In the present study, WMC was measured via an adaptation of an Ospan task, and results obtained from a series of MEMs with WMC added as a control variable revealed that it was a significant predictor of task performance only in terms of lexical diversity. In other words, significant effects for WMC were not found on syntactic complexity or accuracy measures, and hypotheses regarding this research question were only partially sustained.

Two possibilities can be suggested to account for these results. The Ospan task used in the study included two to five blocks of letters. There were ceiling effects for some participants, so the test may not have been difficult enough to differentiate high-WMC learners. Had the Ospan used English words instead of letters, or incorporated blocks with more elements, there may have been more significant findings with regard to WMC and other performance measures. The other possibility is that the relationship between WMC and task performance is captured best by lexical measures. WM involves the temporary storage and management of relevant information in the midst of some kind of activity. When L2 learners produce written or spoken language, they must initially access L2 lemmas before syntactic building procedures are activated. The more capacity they have to hold information, the more lemmas they can retain and process for production. Therefore, those with higher WMC would be able to use more diverse vocabulary in their writing, resulting in a greater GI than those with lower WMC. On the other hand, WMC would be less associated with learners' syntactic complexity and/or accuracy because words can come in various syntactic structures, and it is up to the learner to choose the structure in which a word appears. If the learner does not have advanced English proficiency, like the participants in the present study, the syntactic information that is available would still be unstable. In other words, although some learners may have high WMC, they would not necessarily produce more complex structures and/or possess incorrect syntactic knowledge. Table 43 provides samples to compare the writing among one high-WMC learner (Ospan task score 1) with a high GI (7.77) and high subordinate clause per T-unit (0.8), one high-WMC learner (Ospan task score 1) with a high GI (6.57) and low subordinate clause per T-unit (0.33), and a low-WMC learner (Ospan task score 0.05) with a low GI (4.85) and low subordinate clause per T-unit (0.44). Repeated uses of words and structures can be found in the low-WMC learner's writing, such as "There has," "Check out time is," and "you can/cannot use breakfast." On the other hand, the high-WMC learner with a high syntactic complexity score uses a variety of words in a variety of structures. However, the high-WMC learner with a low syntactic complexity score uses a variety of words in simple structures with fewer subordinate clauses. It is highly likely that L2 proficiency played a role here, contributing to the significant positive relationship between WMC and lexical diversity only.

Table 43. Writing samples to compare WMC, lexical diversity, and syntactic complexity

WMC /GI	Syntactic complexity	Complex Hotel Task
High	High	Among the four, I recommend John and Peter 1 to choose ‘The Imperial.’ The reason why I choose ‘The Imperial’ is related to the distance from the transportation. New York is very is big city. If you want to enjoy this city as much as possible, distance is very important. It takes only 5 minutes on foot to go to the transportation compared to the others. Although it does not include breakfast, I recommend you to enjoy local food instead of the hotel one. Furthermore, check-out time is up to 2:00 p.m. When people travel, they usually sleep late. So, enough check-out time will make you to prepare for the next schedule cheerfully. Two queen size beds will also guarantee of good sleep. The internet fee is expensive compared to the others but I am sure that there is little time to use internet with your phone. Just put you in New York city and enjoy!
High	Low	I recommend The Utopia to John and Peter. There are several reasons 2 that support my suggestion. First, The Utopia is the second most inexpensive hotel of all choices. It also provides free Wi-Fi. Two single-size beds are enough for two men. Although it doesn't provide breakfasts, checking out is possible until 2 pm. However, The Urban doesn't provide any Internet services. The Echo does provide Wi-Fi but it costs $2 per megabyte. The Imperial costs the most and it also costs $5 per megabyte in order to use Wi-Fi. Therefore, I recommend The Utopia for their final choice.
Low	Low	I know you are looking for hotel during New York travelling. I will explain four place to help you. Firstly, The Urban takes 150 dollars in daily rate and 15 minutes to walk to take transportation. But there are no wifi. There has double size two beds and you can use breakfast in free. Check out time is until 11AM. Secondly, The Utopia takes 200 dollars in daily rate and 30 minutes [to walk 12] [to take transportation. There has free wifi. There has single size two beds and you cannot use breakfast. Check out time is until 2PM. Thirdly, The Echo takes 250 dollars in daily rate and 50 minutes to walk to take transportation. There has wifi but you have to pay 2 dollars. There has king size bed and single size bed and you can use breakfast in free. Check out time is until 11AM. Lastly, The Imperial takes 300 dollars in daily rate and 5 minutes to walk to take transportation. There has wifi but you have to pay 5dollars. There has queen size beds and you cannot use breakfast in

free. Check out time is until 2PM. So except check out time, I think The Urban is most appropriate place to you.

6.6 Aptitudes for Language Learning and L2 Written Performance

Regarding the relationship between L2 aptitude and task performance, it was hypothesized that participants with high aptitudes for language learning would show greater syntactic complexity, lexical diversity, and accuracy in their writing. An interesting finding in the present study was that scores on the LLAMA D subtest, which is claimed to represent aptitude for implicit learning (ILA), was positively associated with learners' accuracy. Put simply, those with higher aptitude for implicit learning showed a tendency to produce more accurate structures, in terms of all measures of accuracy: proportions of TLU articles, error-free T-units, error-free clauses, and lexical errors. Therefore, hypotheses regarding ILA and L2 writing were sustained. On the other hand, LLAMA F scores, which represent aptitude for explicit learning (ELA), had a negligible effect on any of the accuracy measures. In fact, LLAMA F was actually found to have a negative effect on the proportion of error-free T-units. This result is not in line with the predictions of the study, as it was predicted that writing, especially untimed or unspeeded, would involve reliance on explicit knowledge and explicit learning mechanisms.

When comparing speech and written production, writing is generally viewed as a more planned activity that involves a higher degree of conscious effort as writers plan, monitor, or review their production. However, it could be argued that the tasks used in the present study can also be considered more 'spontaneous' than others, in that they did not include any sort of pre-task activity, there were no guidelines on how to write their letters, and participants were not allowed to rely on other kinds of supporting materials. Although they were given as much time as needed to plan their writing, they spent around one minute to plan for the simple tasks, and less than two minutes to plan for the complex tasks. It is assumed that during this short planning stage, participants conceptualized their ideas in their L1 when trying to find the solution to the task. Accordingly, L2 lemmas are accessed and syntactic information is activated. Given the short period of time in which planning took place, it is assumed that formulation processes regarding the translation of linguistic expressions occurred almost at the same time as the execution process in which the expressions were typed on the computer/laptop during the writing stage. In this regard, the 'spontaneity' of the writing process drove participants to rely on easy and rapid access to implicit knowledge without their awareness, especially if their primary focus was on content than on form. If those who had higher ILA were able to retrieve more implicit L2 knowledge that underlies accurate spontaneous writing, then they would be able to produce more accurate writing as a result. On the other hand, if a low-ILA learner had more difficulty in employing implicit knowledge of the L2, they would produce less accurate structures. Table 44 provides samples from a participant with high ILA (1) and overall high accuracy scores, and a participant with low ILA and overall low accuracy scores. As can be seen, the high-ILA learner was able to produce accurate sentences, but the low-ILA learner failed to do so in many respects. In fact, they did not even capitalize each sentence.

Table 44. Writing samples to compare ILA and accuracy

ILA	Accuracy	Complex Venue Task
High	High	I recommend The Lunchroom to Liam and Kate. There are several reasons for this idea. First, The Lunchroom is the second most inexpensive restaurant of all. It also has the biggest accommodation, a projector and a screen. Therefore, it can be available for as much people as possible and if there is a video to play for the event, the projector and the screen can help. In addition, The Tower costs the most and it doesn't have available space for more than thirty people. The Springs and The Lighthouse don't have a projector nor a screen. To add, The Lighthouse has the smallest accommodation and the parking lot requests for the most expensive price per hour for parking. Therefore, I recommend The Lunchroom for Liam and Kate's father's birthday party.
Low	Low	i recommend the springs. actually it is very cheap price. if i don't care the price. i will position this place 2nd. they have not free parking space. but i think it is very near the restaurant. and fee is very cheap. this place has a live band and even they don't have the time limit. also they can accept the 40 people. i know they don't have the projector and screen. but it doesn't matter at party. the lunchroom don't have the music. and parking fee is expensive than springs. and it have 3 hours time limit. i really don't like limit the lighthouse is expensive meal price. and parking fee is expensive also accommodation is so small so we should select the people who do not come. the tower is very expensive. even that price they have the time limit.

6.7 Conclusion and Directions for Future Research

The current study investigated the extent to which task complexity, task closure, and cognitive individual differences affected the cognitive load of the task and the linguistic complexity and accuracy of L2 writing. There are many dimensions to tasks, and in order to avoid any confounding variables, the present study set out to disambiguate the effects of task complexity and task condition by manipulating factors that were relatively clear-cut: number of elements for task complexity, and task closure for task condition. Because task complexity was manipulated along a resource-directing dimension of Robinson's CH, hypotheses were formulated such that increases in task complexity would lead to greater cognitive load, and in turn, greater syntactic complexity, lexical diversity, and accuracy. While results obtained from statistical analyses supported the CH in terms of task complexity effects on cognitive load, mixed or null findings were reported for its effects on performance measures, unable to provide unambiguous support for the CH or the TOH.

As evidenced by results of learner self-ratings, expert judgments, and time-on-task measures, the study found that increasing the number of elements involved in a task actually led to the expected changes in cognitive load. In other words, tasks that were intended to be more complex were perceived as such. However, this did not necessarily lead to the desired changes in task performance. Although lexical diversity increased with increased task complexity, there were no significant changes in syntactic complexity measures. In fact, learners were even found to produce a lower proportion of TLU articles when performing complex tasks.

Several studies have found that what researchers intended is not always how participants interpret tasks. For instance, in Lee's (2018) validation study that employed three types of tasks, each with three levels of complexity, it was found that participants ignored the added complexity in the most complex versions of certain tasks. This was also found in Sasayama's (2016) study, where participants carried out a story-telling task with four levels of task complexity. Although the simple version was interpreted as such, the third most complex version elicited the best performance from the L2 learners. Likewise, the closed versions of the tasks employed in the present study proved to be less complex than the open versions, because the ‘openness’ of the tasks enabled participants to be creative, resulting in more complex written output.

From a methodological standpoint, time-on-task was found to be a valid measure of cognitive load. Previous studies that attempted to validate task complexity manipulations used a combination of self-ratings, expert judgments, time estimations, dual task methodology, and eye-tracking, to measure cognitive load. This study is the first in task-based research to use time-on-task as a measure of cognitive load. Due to its relative ease of measuring the time of task onset to completion with the appropriate tools, further research on cognitive load would benefit from using this measure.

The major contribution that the present study makes to the literature is the empirical evidence it obtained regarding task closure effects. There has been very little empirical research on the differential effects of open and closed tasks, and there are even fewer studies that correctly manipulate task closure. During task design, extra care was taken, so that all other variables besides task complexity and task closure were kept constant. Except for the instructions explicitly stating whether there was only one answer (closed) or no answer (open), and the requirements added to the closed tasks, all other task conditions were the same. Therefore, when task closure effects were found, which the present study did, it is highly likely that the results could be attributed solely to the restriction placed on the number of acceptable solutions. Contrary to predictions, participants in the Open condition outperformed those in the Closed condition in terms lexical diversity. Because they could write freely, they used more diverse vocabulary in the form of synonyms, or provided various examples to support their arguments. However, there was a trend for those in the Closed condition to produce longer clauses than those in the Open condition when performing complex tasks. It was speculated that the requirements provided in order to close the task could have freed up attentional capacities or allowed for easy translations of requirements into sentences, resulting in longer clauses.

Significant interaction effects between task complexity and task closure were also found on time-on-planning. It is worth noting that when the effects of task complexity interacted with those of task closure, they had a greater impact in the Closed than in the Open condition. When examining the time engaged in planning, those Closed condition took significantly longer to plan for the complex tasks than the simple ones. However, this effect was not found in the time-on-writing measure, indicating that a greater cognitive load was not placed on those in the Closed condition during the actual writing stages.

Cognitive individual differences were also found to play a role in task performance. Higher working memory capacity was found to be associated with greater lexical diversity, most likely because it is easier to access a greater number of L2 lemmas to express ideas for high-WMC learners. In the case of language aptitude, implicit learning aptitude, and not explicit learning aptitude, was found to be positively related to all measures of accuracy. In order to account for this finding, a possible explanation was suggested, in that the writing tasks employed in the study required relatively more spontaneous language use, which drew upon implicit knowledge that enabled participants to produce more accurate T-units.

Limitations of the present study should be noted, so that more in-depth investigations can be conducted in the future. Task complexity was manipulated in terms of the number of elements involved, but an even greater number of elements may result in more significant task complexity effects on syntactic complexity. For instance, more significant task complexity effects might have been found if the number of location options had doubled in the complex versions. There is also the possibility that number of elements is not a good reflection of task complexity, and other variables such as ± Here-and-Now or ± planning time would work better. Learner self-ratings and expert ratings of difficulty and mental effort were not very high for the complex tasks, and the complex tasks may not have been cognitively challenging enough to push learners to produce more syntactically complex structures. Written monologic tasks were employed in the present study, and different findings might be obtained for task closure effects if spoken interactive tasks were used. In terms of the Ospan task, there were ceiling effects for some participants, and it would have been more beneficial to include at least three more tests to measure WMC, ILA, and ELA separately. Furthermore, participants in the study possessed lower- to upper-intermediate English proficiency. More significant results could have been obtained if more advanced learners were involved in the study. Further research could also test learners at other levels in order to see if L2 proficiency interacts with task closure effects.

Appendix A: Language Background Questionnaire (English)

1. Gender: M ☐ / F ☐

2. Age: ____________

3. Scores on recent standardized English test (Date: Month ________ Year ________)

	Reading	Listening	Speaking	Writing	Grammar	Vocabulary	Total
TOEFL							
TOEIC							
TEPS							

4. At what age did you first learn English? ____________

5. List foreign language(s) you are familiar with, in order of proficiency (You may include English)

1: __________________ 2: __________________ 3: __________________

5. Experience living in an English-speaking country? Yes ☐ / No ☐

Year	Country	Duration	Purpose

7. English courses at the university level? Yes ☐ / No ☐

Year	Duration and frequency	Course name

8. Degree earned: High school ☐ Bachelor's ☐ Master's ☐ Ph.D. ☐

9. Major: ________________________________

10. Amount of English use in a week (in <u>hours</u>)

Reading	0	1-2	3-4	5-6	7-8	At least 9h
Listening	0	1-2	3-4	5-6	7-8	At least 9h
Speaking	0	1-2	3-4	5-6	7-8	At least 9h
Writing	0	1-2	3-4	5-6	7-8	At least 9h

Appendix B: Simple Open Hotel Task (English)

Finding Hotel in Hawaii

Carrey and Amanda are trying to find a hotel to stay at during their trip in Hawaii. However, they are having a hard time coming to an agreement on which hotel to choose among the following three: (1) The Moonriver, (2) The Sunset, and (3) The Utopia. Based on the information about the hotels, write a letter about which hotel you would recommend.

* **Explain why you choose one location and why you do not choose the others.**

** **Include your personal opinion in the letter, as there is no right answer.**

	The Moonriver	**The Sunset**	**The Utopia**
Daily rate	$100	$120	$95
Internet access	Free Wi-fi	$2 per megabyte	$5 per megabyte
Distance to the beach	10 minutes by bus	5 minutes on foot	10 minutes on foot

Appendix C: Complex Open Hotel Task (English)

Finding Hotel in New York

John and Peter are trying to find a hotel to stay at during their trip in New York. However, they are having a hard time coming to an agreement on which hotel to choose among the following four: (1) The Urban, (2) The Utopia, (3) The Echo, and (4) The Imperial. Based on the information about the hotels, write a letter about which hotel you would recommend.

* **Explain why you choose one location and why you do not choose the others.**

**** Include your personal opinion in the letter, as there is no right answer.**

	The Urban	**The Utopia**	**The Echo**	**The Imperial**
Daily rate	$150	$200	$250	$300
Access to public transportation	15 min on foot	30 min on foot	50 min on foot	5 min on foot
Internet	No internet	Free Wi-fi	$2 per megabyte	$5 per megabyte
Bed	2 double sized	2 single sized	1 king sized, 1 single sized	2 queen sized
Breakfast	Included	Not included	Included	Not included
Check-out time	Until 11am	Until 2pm	Until 11am	Until 2pm

Appendix D: Simple Closed Hotel Task (English)

Finding Hotel in Hawaii

Carrey and Amanda are trying to find a hotel to stay at during their trip in Hawaii. However, they are having a hard time coming to an agreement on which hotel to choose among the following three: (1) The Moonriver, (2) The Sunset, and (3) The Utopia. Based on Carrey and Amanda's requirements and the information about the hotels, write a letter about which hotel you would recommend.

* **Explain why you choose one location and why you do not choose the others.**

**** There is only one hotel that meets all requirements.**

Carrey and Amanda's requirements:

- **Budget: $100/night**
- **Internet: Free Wi-fi**
- **Transportation: Public transportation available**

	The Moonriver	**The Sunset**	**The Utopia**
Daily rate	$100	$120	$95
Internet access	Free Wi-fi	$2 per megabyte	$5 per megabyte
Distance to the beach	10 minutes by bus	5 minutes on foot	10 minutes on foot

Appendix E: Complex Closed Hotel Task (English)

Finding Hotel in New York

John and Peter are trying to find a hotel to stay at during their trip in New York. However, they are having a hard time coming to an agreement on which hotel to choose among the following four: (1) The Urban, (2) The Utopia, (3) The Echo, and (4) The Imperial. Based on John and Peter's requirements and the information about the hotels, write a letter about which hotel you would recommend.

* **Explain why you choose one location and why you do not choose the others.**

**** There is only one hotel that meets all requirements.**

John and Peter's requirements:

- **Budget: $250/night**
- **Transportation: Prefers walking to taking public transportation**
- **Internet: Uses a lot**
- **Bed: Must have 2 (size does not matter)**
- **Breakfast: Prefers local food to hotel food**
- **Check-out: Prefers afternoon check-out**

	The Urban	**The Utopia**	**The Echo**	**The Imperial**
Daily rate	$150	$200	$250	$300
Access to public transportation	15 min on foot	30 min on foot	50 min on foot	5 min on foot
Internet	No internet	Free Wi-fi	$2 per megabyte	$5 per megabyte
Bed	2 double sized	2 single sized	1 king sized, 1 single sized	2 queen sized
Breakfast	Included	Not included	Included	Not included
Check-out time	Until 11am	Until 2pm	Until 11am	Until 2pm

Appendix F: Simple Open Venue Task (English)

Finding Venue for 70th birthday party

Brian and Alice are trying to find a venue to hold their mother's 70th birthday party. However, they are having a hard time coming to an agreement on which restaurant to choose among the following three: (1) The Square, (2) The Castle, and (3) The Royal. Based on the information about the restaurants, write a letter about which restaurant you would recommend for holding the birthday party.

* **Explain why you choose one location and why you do not choose the others.**

**** Include your personal opinion in the letter, as there is no right answer.**

	The Square	**The Castle**	**The Royal**
Average meal cost per person	$65	$45	$55
Room accommodation	30	40	20
Atmosphere	No music	Quiet music	Live music

Appendix G: Complex Open Venue Task (English)

Finding Venue for 80th birthday party

Liam and Kate are trying to find a venue to hold their father's 80th birthday party. However, they are having a hard time coming to an agreement on which restaurant to choose among the following four: (1) The Lunchroom, (2) The Lighthouse, (3) The Tower, and (4) The Springs. Based on the information about the restaurants, write a letter about which restaurant you would recommend for holding the birthday party.

**** Include your personal opinion in the letter, as there is no right answer.**

	The Lunchroom	**The Lighthouse**	**The Tower**	**The Springs**
Average meal cost per person	$55	$63	$70	$45
Atmosphere	No music	Quiet music	Live piano music	Live band performance
Parking space	Public parking 5 minutes on foot $2 per hour	Public parking 10 minutes on foot $3 per hour	Free private parking	Private parking $1 per hour
Room accommodation	50	20	30	40
Room time limit	3 hours	No limit	3 hours	No limit
Video display equipment	Projector (O), Screen (O)	Projector (X), Screen (X)	Projector (O), Screen (O)	Projector (X), Screen (X)

Appendix H: Simple Closed Venue Task (English)

Finding Venue for 70th birthday party

Brian and Alice are trying to find a venue to hold their mother's 70th birthday party. However, they are having a hard time coming to an agreement on which restaurant to choose among the following three: (1) The Square, (2) The Castle, and (3) The Royal. Based on Brian and Alice's requirements and the information about the restaurants, write a letter about which restaurant you would recommend for holding the birthday party.

* **Explain why you choose one location and why you do not choose the others.**

**** There is only one restaurant that meets all requirements.**

Brian and Alice's requirements:

- **Budget: Maximum $50 per person**
- **Room accommodation: At least 40 people**
- **Atmosphere: Quiet**

	The Square	**The Castle**	**The Royal**
Average meal cost per person	$65	$45	$55
Room accommodation	30	40	20
Atmosphere	No music	Quiet music	Live music

Appendix I: Complex Closed Venue Task (English)

Finding venue for 80th birthday party

Liam and Kate are trying to find a venue to hold their father's 80th birthday party. However, they are having a hard time coming to an agreement on which restaurant to choose among the following four: (1) The Lunchroom, (2) The Lighthouse, (3) The Tower, and (4) The Springs. Based on Liam and Kate's requirements and the information about the restaurants, write a letter about which restaurant you would recommend for holding the birthday party.

* **Explain why you choose one location and why you do not choose the others.**

**** There is only one restaurant that meets all requirements.**

Liam and Kate's requirements:

- **Budget: Maximum $70 per person**
- **Atmosphere: Quiet**
- **Parking space: A must**
- **Room accommodation: At least 30 people**
- **Room use: At most 3 hours**
- **Video display equipment: A must**

	The Lunchroom	**The Lighthouse**	**The Tower**	**The Springs**
Average meal cost per person	$55	$63	$70	$45
Atmosphere	No music	Quiet music	Live piano music	Live band performance
Parking space	Public parking 5 minutes on foot $2 per hour	Public parking 10 minutes on foot $3 per hour	Free private parking	Private parking $1 per hour
Room accommodation	50	20	30	40
Room time limit	3 hours	No limit	3 hours	No limit
Video display equipment	Projector (O), Screen (O)	Projector (X), Screen (X)	Projector (O), Screen (O)	Projector (X), Screen (X)

Appendix J: Learner Self-rating Questionnaire (English)

Please answer the following questions after completing each task version.

1. What was the level of overall difficulty of the task? (Please circle)

1 2 3 4 5 6 7 8 9

Very Easy Very Difficult

2. How much mental effort was required to perform the task? (Please circle)

1 2 3 4 5 6 7 8 9

Very Little Effort Very Much Effort

3. What was the level of stress you felt while performing the task? (Please circle)

1 2 3 4 5 6 7 8 9

Very Relaxed Very Frustrated

4. How interesting was the task?

1 2 3 4 5 6 7 8 9

Not interesting Very Interesting

5. How much time pressure did you feel during the planning stage?

1 2 3 4 5 6 7 8 9

Very little pressure Very much pressure

6. How much time pressure did you feel during the writing stage?

1 2 3 4 5 6 7 8 9

Very little pressure Very much pressure

7. How well do you believe you carried out the task?

1 2 3 4 5 6 7 8 9

Very poorly Very well

Bibliography

Baddeley, A. (2003). Working memory: looking back and looking forward. *Nature Reviews Neuroscience, 4*(10), 829-839.

Baddeley, A., & Hitch, G. (1974). Working memory. *The psychology of learning and motivation, 8*, 47-89.

Baddeley, A., & Logie, R. (1992). Working memory: An overview. *Science*, 255, 556-559.

Baralt, M. (2013). The impact of cognitive complexity on feedback efficacy during online versus face-to-face interactive tasks. *Studies in Second Language Acquisition*, *35*(04), 689-725.

Baralt, M., Gilabert, R., & Robinson, P. (2014). An introduction to theory and research in task sequencing and instructed second language learning. *Task sequencing and instructed second language learning*, 1-50.

Block, R. A. (1992). Prospective and retrospective duration judgment: The role of information processing and memory. In E Macar, V.Pouthas, & W.1.Friedman (Eds.), *Time, action and cognition: Towards bridging the gap* (pp. 109- 118). Dordrecht: Kluwer.

Block, R. A. (2003) Psychological timing without a timer: The roles of attention and memory. In H. Helfrich (Ed.), *Time and mind II* (pp. 41- 60). Göttingen: Hogrefe & Huber.

Block, R. A., Hancock, P. A., & Zakay, D. (2010). How cognitive load affects duration judgments: A meta-analytic review. *Acta Psychologica, 134*(3), 330-343.

Block, R. A., & Zakay, D. (2008) Timing and remembering the past, the present, and the future. In S. Grondin (Ed.), *Psychology of time* (pp. 367-394). Emerald Group Publishing Ltd.

Brown, J. D. (1980). Relative merits of four methods for scoring cloze tests. *The Modern Language Journal*, *64*(3), 311-317.

Brunken, R., Plass, J. L., & Leutner, D. (2003). Direct Measurement of Cognitive Load in Multimedia Learning. *Educational Psychologist, 38*(1), 53-61.

Carroll, J. B. (1989). The Carroll model a 25-year retrospective and prospective view. *Educational Researcher, 18*(1), 26-31.

Carroll, J. B. (1990). Cognitive abilities in foreign language aptitude: Then and now. *Language aptitude reconsidered*, 11-29.

Carrol, J. B., & Sapon, S. M. (1959). Modern language aptitude test: Mlat. *New York: Psychological Corporation*, *68*, n1.

Chrabaszcz, A., & Jiang, N. (2014). The role of the native language in the use of the English nongeneric definite article by L2 learners: A cross-linguistic comparison. *Second Language Research*, *30*(3), 351-379.

Cobb,T. *Web Vocabprofile* [accessed August 2018 from http://www.lextutor.ca/vp/], an adaptation of Heatley, Nation & Coxhead's (2002) *Range*.

Conway, A. R., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic bulletin & review, 12*(5), 769-786.

DeKeyser, R. M., & Koeth, J. (2011). Cognitive aptitudes for second language learning. *Handbook of research in second language teaching and learning*, *2*, 395-406.

Dörnyei, Z. (2005). *The psychology of the language learner : Individual differences in second language acquisition* (Second language acquisition research). Mahwah, N.J.: L. Erlbaum.

Ellis, D. (2011). The role of task complexity in the linguistic complexity of native speaker output. Qualifying paper, PhD in Second Language Acquisition Program. College Park, MD: University of Maryland.

Ellis, R. (2017). Position paper: Moving task-based language teaching forward. *Language Teaching*, *50*(4), 507-526.

Engle, R. W., Cantor, J., & Carullo, J. J. (1992). Individual differences in working memory and comprehension: a test of four hypotheses. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*(5), 972.

Flower, L., & Hayes, J. R. (1980). The cognition of discovery: Defining a rhetorical problem. *College composition and communication*, *31*(1), 21-32.

Forster, K. I., & Forster, J. C. (2003). DMDX: A Windows display program with millisecond accuracy. *Behavior research methods, instruments, & computers*, *35*(1), 116-124.

Foster, P., & Skehan, P. (1996). The Influence of Planning and Task Type on Second Language Performance. *Studies in Second Language Acquisition, 18*(03), 299-323.

Foster, P., & Tavakoli, P. (2009). Native speakers and task performance: Comparing effects on complexity, fluency, and lexical diversity. *Language learning*, *59*(4), 866-896.

Gathercole, S. E. (1999). Cognitive approaches to the development of short-term memory. *Trends in cognitive sciences*, *3*(11), 410-419.

Gilabert, R. (2007). The Simultaneous Manipulation of Task Complexity Along Planning Time and (+/- Here-and-Now): Effects on L2 Oral Production. In M. d. P. G. Mayo (Ed.), *Investigating Tasks in Formal Language Learning* (pp. 44-68). Tonawanda, NY: Multilingual Matters Ltd.

Gilabert, R., Barón, J., & Llanes, À. (2009). Manipulating cognitive complexity across task types and its impact on learners' interaction during oral performance. *International Review of Applied Linguistics in Language Teaching, 47*(3-4), 367-395.

Grañena, G. (2013). Cognitive aptitudes for second language learning and the LLAMA Language Aptitude Test. *Sensitive periods, language aptitude, and ultimate L2 attainment*, *35*, 105.

Goldhammer, F., & Entink, R. H. K. (2011). Speed of reasoning and its relation to reasoning ability. *Intelligence*, *39*(2-3), 108-119.

Goldhammer, F., Naumann, J., Stelter, A., Tóth, K., Rölke, H., & Klieme, E. (2014). The time on task effect in reading and problem solving is moderated by task difficulty and skill: Insights from a computer-based large-scale assessment. *Journal of Educational Psychology*, *106*(3), 608.

Guiraud, P. (1954). *Les Charactères Statistiques du Vocabulaire. Essai de méthodologie*. Paris: Presses Universitaires de France.

Heatley, A., Nation, I.S.P. & Coxhead, A. (2002). RANGE and FREQUENCY programs. Available at http://www.victoria.ac.nz/lals/staff/paul-nation.aspx .

Hunt, K. (1964*). Differences in grammatical structures written at three grade levels* (Cooperative Research Project No. 1998). Tallahassee: Florida State University.

Ishikawa, T. (2007). The effect of manipulating task complexity along the (+Here-and-Now) dimension on L2 written narrative discourse. In M. d. P. G. Mayo (Ed.), *Investigating Tasks in Formal Language Learning* (pp. 136-156). Tonawanda, NY: Multilingual Matters Ltd.

Ishikawa, T. (2011). Examining the influence of intentional reasoning demands on learner perceptions of task difficulty and L2 monologic speech. In P. Robinson (Ed.), *Second Language Task Complexity: Researching the Cognition Hypothesis of Language Learning and Performance* (pp. 307-330). Philadelphia, PA: John Benjamins Publishing Company.

Jackson, D. O., & Suethanapornkul, S. (2013). The cognition hypothesis: A synthesis and meta-analysis of research on second language task complexity. *Language Learning, 63*(2), 330-367.

Juffs, A., & Harrington, M. (2011). Aspects of working memory in L2 learning. *Language Teaching*, *44*(02), 137-166.

Kellogg, R. T. (1996). A model of working memory in writing. In C. M. Levy & S. Ransdell (Eds.), *The science of writing: Theories, methods, individual* *differences, and applications* (pp. 57–71). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Kellogg, R. T. (1999). *The psychology of writing*. Oxford University Press.

Kurilesz, M. (1969). *Man and his world: A structured reader*. New York: Crowell.

Kim, Y. (2003). *Effects of input elaboration and enhancement on second language vocabulary acquisition through reading by Korean learners of English. Unpublished doctoral dissertation*. University of Hawai'i. Hawai'i.

Kim, Y., Payant, C., & Pearson, P. (2015). The intersection of task-based interaction, task complexity, and working memory. *Studies in Second Language Acquisition*, *37*(03), 549-581.

Kormos, J. (2013). New conceptualizations of language aptitude in second language attainment. *Sensitive periods, language aptitude and ultimate attainment*, 131-152.

Kormos, J., & Trebits, A. (2011). Working memory capacity and narrative task performance. In P. Robinson (Ed.), *Second language task complexity: Researching the Cognition Hypothesis of language learning and performance* (pp. 267-285). Philadelphia, PA: John Benjamins Publishing Company.

Kuiken, F., & Vedder, I. (2007). Task complexity and measures of linguistic performance in L2 writing. *International Review of Applied Linguistics in Language Teaching, 45*(3), 261-284.

Kuiken, F., & Vedder, I. (2008). Cognitive task complexity and written output in Italian and French as a foreign language. *Journal of Second Language Writing, 17*(1), 48-60.

Lee, J. (2018). Task complexity, cognitive load, and l1 speech. *Applied Linguistics, 39*(3), 444-444. doi:10.1093/applin/amy011

Linck, J. A., Osthus, P., Koeth, J. T., & Bunting, M. F. (2014). Working memory and second language comprehension and production: A meta-analysis. Psychonomic Bulletin & Review, 21(4), 861-883.

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

Loschky, L., & Bley-Vroman, R. (1993). Grammar and task-based methodology. *Multilingual Matters*, 123-123.

Long, M. H. (1985). A role for instruction in second language acquisition: task-based language teaching. In K. Hyltenstam & M. Pienemann (Eds.), *Modelling and Assessing Second Language Acquisition* (pp. 77-99). Clevedon: Multilingual Matters Ltd.

Long, M. H. (1989). Task, group, and task-group interactions. *University of Hawai'i Working Papers in ESL, 8*(2), 1-26. (Reprinted in S. Anivan (Ed.), *Language teaching methodology for the nineties*, pp. 31-50, 1990, Singapore: SEAMEO Regional Language Center.)

Long, M. H. (2015). Task-Based Syllabus Design *Second Language Acquisition and Task-Based Language Teaching* (pp. 205-247). Malden, MA: Wiley-Blackwell.

Long, M. H., & Crookes, G. (1992). Three Approaches to Task-Based Syllabus Design. *TESOL Quarterly, 26*(1), 27-56.

Malicka, A., & Levkina, M. (2012). Measuring task complexity: does L2 proficiency matter. *Task-based Language Teaching in Foreign Language Contexts: Research and Implementation. John Benjamins*.

Malone, J. (2018). Incidental vocabulary learning in sla. *Studies in Second Language Acquisition, 40*(03), 651-675. doi:10.1017/S0272263117000341

Meara, P. (2005). LLAMA language aptitude tests: The manual. *Swansea: Lognostics*.

Michel, M. C. (2011). Effects of task complexity and interaction on L2 performance. In P. Robinson (Ed.), *Second Language Task Complexity: Researching the Cognition Hypothesis of Language Learning and Performance* (pp. 141-173). Philadelphia, PA: John Benjamins Publishing Company.

Miyake, A., & Friedman, N. P. (2012). The nature and organization of individual differences in executive functions four general conclusions. *Current directions in psychological science*, *21*(1), 8-14.

Montero, F. (2018). Effects of task complexity and task closure on the speech of L2 learners of Spanish. Qualifying paper, PhD in Second Language Acquisition Program. College Park, MD: University of Maryland.

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

Nunan, D. (1991). Communicative tasks and the language curriculum. *TESOL quarterly*, *25*(2), 279-295.

Petersen, C. R., & Al-Haik, A. R. (1976). The Development of the Defense Language Aptitude Battery (Dlab. *Educational and Psychological Measurement*, *36*(2), 369-380.

Pimsleur, P. (1966). *Pimsleur Language Aptitude Battery (form S)*. Harcourt, Brace and world, Incorporated.

Rahimpour, M. (2009). Impact of task condition on L2 learners' oral performance. *Journal of Faculty of Letters & Humanities*, No.208, 13-23.

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.

Révész, A., Sachs, R., & Hama, M. (2014). The effects of task complexity and input frequency on the acquisition of the past counterfactual construction through recasts. *Language Learning*, *64*(3), 615-650.

Révész, A., Michel, M., & Gilabert, R. (2015). Measuring cognitive task demands using dual task methodology, subjective self-ratings, and expert judgments: a validation study. *Studies in Second Language Acquisition, 28*(4), 1-35.

Robinson, P. (1995). Task Complexity and Second Language Narrative Discourse. *Language Learning, 45*(1), 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). New York, NY: Cambridge University Press.

Robinson, P. (2001b). Task complexity, task difficulty, and task production: Exploring interactions in a componential framework. *Applied Linguistics, 22*(1), 27-57.

Robinson, P. (2001c). Individual differences, cognitive abilities, aptitude complexes and learning conditions in second language acquisition. *Second language research*, *17*(4), 368-392.

Robinson, P. (2003). The Cognitive Hypothesis, Task Design, and Adult Task-Based Language Learning. *Second Language Studies, 21*(2), 45-105.

Robinson, P. (2005a). Aptitude and second language acquisition. *Annual Review of Applied Linguistics*, *25*, 46-73.

Robinson, P. (2005b). Cognitive Complexity and Task Sequencing: Studies in a Componential Framework for Second Language Task Design. *International Review of Applied Linguistics in Language Teaching, 43*(1), 1-32.

Robinson, P. (2007). Criteria for Classifying and Sequencing Pedagogic Tasks. In M. d. P. G. Mayo (Ed.), *Investigating Tasks in Formal Language Learning* (pp. 7-26). Tonawanda, NY: Multilingual Matters Ltd.

Robinson, P. (2010) Situating and distributing cognition across task demands: The SSARC model of pedagogic task sequencing. In M. Putz and L. Sicola (eds), *Cognitive processing in second language acquisition: Inside the learner's mind.* (pp. 243-68). Amsterdam/Philadelphia: John Benjamins.

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). Philadelphia, PA: John Benjamins Publishing Company.

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

Rogers, V., Meara, P., Barnett-Legh, T., Curry, C., & Davie, E. (2017). Examining the LLAMA aptitude tests. *Journal of the European Second Language Association*, *1*(1).

Sasayama, S. (2016). Is a 'complex' task really complex? Validating the assumption of cognitive task complexity. *The Modern Language Journal, 100*(1), 231-254.

Skehan, P. (1996). A Framework for the Implementation of Task-based Instruction. *Applied Linguistics, 17*(1), 38-62.

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

Skehan, P. (2001). Tasks and language performance. *Researching pedagogic tasks: Second language learning, teaching, and testing*, 167-185.

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

Skehan, P. (2014). *Processing perspectives on task performance*. Philadelphia, PA: John Benjamins Publishing Company.

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., & Foster, P. (1997). Task type and task processing conditions as influences on foreign language performance. *Language Teaching Research,* *1*(3), 185-211.

Sullivan, P., Warren, E., & White, P. (2000). Students' responses to content specific open-ended mathematical tasks. *Mathematics education research journal*, *12*(1), 2-17.

Suzuki, Y., & DeKeyser, R. (2017). The interface of explicit and implicit knowledge in a second language: Insights from individual differences in cognitive aptitudes. *Language Learning*, *67*(4), 747-790.

Turner, J., & Paris, S. G. (1995). How literacy tasks influence children's motivation for literacy. *The reading teacher*, *48*(8), 662-673.

VanPatten, B. (1990). Attending to Form and Content in the Input. *Studies in Second Language Acquisition, 12*(03), 287-301.

Vasylets, O. (2016). *Task-modality effects: A study of task complexity effects in speech and writing*. Unpublished doctoral dissertation. University of Barcelona, Barcelona, Spain.

Yalçın, Ş., Çeçen, S., & Erçetin, G. (2016). The relationship between aptitude and working memory: an instructed SLA context. *Language Awareness*, *25*(1-2), 144-158.

Yilmaz, Y. (2013). The role of working memory capacity and language analytic ability in the effectiveness of explicit correction and recasts. *Applied Linguistics*, *34*(3), 344-368.

Yilmaz, Y., & Grañena, G. (2016). The role of cognitive aptitudes for explicit language learning in the relative effects of explicit and implicit feedback. *Bilingualism: Language and Cognition*, *19*(01), 147-161.

Zakay, D. (1992). On prospective time estimation, temporal relevance and temporal uncertainty. In E Macar, V.Pouthas, & W.1.Friedman (Eds.), *Time, action and cognition: Towards bridging the gap* (pp. 109- 118). Dordrecht: Kluwer.